A new approach to El Niño Southern Oscillation origin and forecasting: implications for predictability

Un nuevo enfoque para el origen y la predicción de El Niño Oscilación Sur: implicaciones para la predictibilidad

Desislava Bozhidarova Petrova

A new approach to El Niño Southern Oscillation origin and forecasting: implications for predictability

Un nuevo enfoque para el origen y la predicción de El Niño Oscilación Sur: implicaciones para la predictibilidad

Author
Desislava Bozhidarova Petrova

Director
Xavier Rodó López

Barcelona
April 2017

University of Barcelona

Doctoral Program in Physics, Department of Physics

Doctoral thesis developed at the
Catalan Institute for Climate Sciences

Barcelona,
April 2017

A new approach to El Niño Southern Oscillation origin and forecasting: implications for predictability

Un nuevo enfoque para el origen y la predicción de El Niño Oscilación Sur: implicaciones para la predictibilidad

Author
Desislava Bozhidarova Petrova
Catalan Institute for Climate Sciences

Director
Xavier Rodó López
Catalan Institute for Climate Sciences

Tutor
Ileana Blade
University of Barcelona

Abstract

El Niño Southern Oscillation (ENSO) is an interannual climate phenomenon that arises in the tropical Pacific as a result of coupled interactions between the atmosphere and the ocean. It is the most prominent modulator of climate variability on this timescale, and in this respect it plays a fundamental role as a precursor in seasonal climate forecasting worldwide. The nature of the dynamical system of the atmosphere is chaotic and in that sense its predictability is sensitive to initial conditions, which constraints our ability to foresee the evolution of ENSO for an unlimited period of time in advance. This dissertation is dedicated to exploring the possibility for extending the state-of-the-art prediction of the phenomenon. In particular, it focuses on the identification of precursory signals in the ocean and in the atmosphere that improve the understanding and long-lead forecasts of the events. In addition, a new statistical modelling technique based on dynamic components and state-space methods is developed and applied to the problem of ENSO prediction, and the selected precursor covariates are incorporated in its design. This exceedingly flexible methodology has been tested and verified in other areas of science such as engineering and econometrics, but is only beginning to enter the field of climate science. In this thesis it is shown that to some extent this methodological strategy can bridge the purely dynamical and purely statistical concepts of a forecasting exercise.

Very early premonitory signals that are a result of an in-depth and step-by-step analysis of the processes accompanying the origin and evolution of El Niño, and especially those happening in the subsurface ocean, which is less impacted by initial conditions, are established. These tracers are defined in the far western and central tropical Pacific region and are shown to anticipate El Niño at least two and a half years before its boreal winter peak in the eastern equatorial Pacific. Initial intensification of the easterly trade winds at this time is associated with convergence of mass, downwelling and anomalous warming of the subsurface tropical ocean layers in the far west. In this way, both the South Equatorial Current and the Equatorial Undercurrent are strengthened, which leads to the propagation of warm subsurface anomalies eastward along and below the thermocline. These anomalous patterns lead to changes in the circulation and anomalous warming of the surface of the ocean in the central tropical Pacific at a later stage, which then leads to the suppression of the

easterly trade winds. The main area of tropical convection shifts to the east, which weakens the Walker circulation and triggers the Bjerknes feedback. This allows the further propagation of the subsurface warm ocean anomalies, which eventually reach the eastern Pacific and are upwelled to the surface, which marks the onset of an El Niño event.

Warm anomalies in the subsurface equatorial ocean have been previously used as precursors in statistical ENSO forecasting models via the integration of the upper ocean heat content or through the incorporation of anomalies of the 20°C isotherm. In this way, however, the propagation feature of the anomalies is not taken into account, and no direct connection is made between the first anomalous patterns and the occurrence of a warm event. Instead, only the peak in upper ocean heat content is used to signal a forthcoming El Niño, and it occurs much later in the overall evolution of the event, i.e. about 9 months before its mature phase. Hence, the predictive potential of the incipient warming that starts much earlier in the western tropical Pacific, and that could substantially extend the forecast lead time, is not harnessed in these prediction schemes. Furthermore, through experimentation with a complex physical model it is also shown here that a significant association between the increase in heat content in the western and central Pacific and the occurrence of El Niño exists at the longer lead time of 21 months, and not just during the peak of equatorial heat content 9 months before. Therefore, the work presented in the dissertation provides strong implications for the possibility of improvement of the capacity and long-lead capabilities of other models.

The definition of ENSO predictors at specific depths and regions in the ocean and atmosphere also requires the availability of reliable surface and subsurface measurements of various climate variables. Although sporadic measurements have been taken in the tropical Pacific for a long time as a result of shipping expeditions and scientific missions, regular and high-resolution measurements have only begun with the satellite era in the 1980s, and especially with the placement of an observation system in the tropics as a result of the Tropical Ocean Global Atmosphere Program (1985-1994). This Program has been launched primarily to deepen the theoretical understanding of ENSO and to significantly advance its forecasting. In the dissertation it is verified that the provision of more comprehensive and consistent data sets as a result of the Program also significantly enhances the forecasting capabilities of the designed statistical model. Forecasts evidently and substantially improve after 1994, and the change is distinctly pronounced for the long-lead forecasts that rely on good-quality subsurface information about the ocean thermal structure in the western and central tropical Pacific. It is thus established that the higher temporal and spatial resolution data sets of key variables for El Niño dynamics are now long

enough for statistical forecasting models to make better use of.

The practical utility of multi-year forecasts of El Niño is also explored in the thesis. Major teleconnections are driven by ENSO, and this makes it a good predictor for the climate patterns worldwide, but especially more so in the vicinity of the tropical Pacific. A well established link in the literature exists between ENSO and local climatic changes in the coastal areas of Ecuador. Following El Niño is a generally warmer surface temperature and significantly enhanced precipitation in that region. These two variables, on the other hand, control the dynamics of mosquito population and the number of mosquito breeding sites, and in this way affect the incidence of dengue and other infectious diseases spread by these vectors. Therefore, an experiment is performed where long-lead forecasts of El Niño are used within a dengue prediction model for the province of El Oro in southern coastal Ecuador. The potential for very long-lead anticipation of dengue outbreaks in the region is explored, as well as the prospects for developing an early warning system for dengue epidemics. The analysis serves as a demonstration of the possibility for an operational climate service in support of the public health community in Ecuador.

Resumen

El Niño Oscilación Sur (ENSO) es un fenómeno climático interanual que surge en el Pacífico tropical como resultado de las interacciones acopladas entre la atmósfera y el océano. Es el modulador más prominente de la variabilidad climática en esta escala del tiempo y, a este respecto, desempeña un papel fundamental como precursor en el pronóstico climático estacional en todo el mundo. La naturaleza del sistema dinámico de la atmósfera es caótica y en ese sentido su predictibilidad es sensible a las condiciones iniciales, lo que limita nuestra capacidad de prever la evolución del ENSO por un período de tiempo ilimitado de antemano. Esta tesis se dedica a explorar la posibilidad de extender y mejorar la predicción de última generación. En particular, se centra en la identificación de señales precursoras en el océano y en la atmósfera que mejoran la comprensión y los pronósticos a largos plazos de los eventos. Además, se desarrolla una nueva técnica de modelación estadística basada en componentes dinámicos y métodos de modelos de espacio de estado y se aplica al problema de la predicción de ENSO, y las covariables precursoras seleccionadas se incorporan en su diseño. Esta metodología extremadamente flexible ha sido probada y verificada en otras áreas de la ciencia como la ingeniería y la econometría, pero está comenzando a entrar en el campo de la ciencia del clima. En esta tesis se demuestra que en cierta medida esta estrategia metodológica puede unir los conceptos puramente dinámicos y puramente estadísticos de un ejercicio de pronóstico.

Las señales premonitorias muy tempranas que son el resultado de un análisis en profundidad y paso a paso de los procesos que acompañan al origen y evolución de El Niño, y especialmente aquellos que ocurren en el océano subsuperficial, menos impactado por las condiciones iniciales, son establecidas. Estos trazadores se definen en el extremo occidental y central de la región del Pacífico tropical y se demuestra que anticipan a El Niño por lo menos por dos años y medio antes de su pico boreal de invierno en el Pacífico ecuatorial oriental. La intensificación inicial de los vientos alisios del este en este momento se asocia con la convergencia de masa, el proceso de *downwelling* y el calentamiento anómalo de las capas de profundidad del océano tropical subsuperficiales en el lejano oeste. De esta manera, tanto la Corriente Ecuatorial Sur como la Corriente Subterránea Ecuatorial se fortalecen, lo que conduce a la propagación de anomalías cálidas subsuperficiales hacia el este a lo largo y por

debajo de la termoclina. Estos procedimientos anómalos conducen a cambios en la circulación y calentamiento anómalo de la superficie del océano en el Pacífico tropical central en una etapa posterior, que luego conduce a la supresión de los vientos alisios de este. La principal área de convección tropical se desplaza hacia el este, lo que debilita la circulación de Walker y provoca la retroalimentación de Bjerknes. Esto permite la propagación ulterior de las anomalías oceánicas cálidas del subsuelo, que eventualmente alcanzan el Pacífico oriental y se elevan a la superficie por las reacciones advectivas y termoclinas, que marca el inicio de un evento El Niño.

Las anomalías cálidas en el océano ecuatorial subsuperficial se han utilizado como precursores previamente en modelos estadísticos de predicción de ENSO mediante la integración del contenido de calor oceánico superior o mediante la incorporación de anomalías de la isoterma de 20 °C. De esta manera, sin embargo, no se tiene en cuenta la característica de propagación de las anomalías y no se hace ninguna conexión directa entre los primeros patrones anómalos y la ocurrencia de un evento cálido. En cambio, sólo el pico en el contenido de calor del océano superior se utiliza para señalar un próximo El Niño, y este pico se produce mucho más tarde en la evolución general de un evento, es decir, alrededor de 9 meses antes de su fase madura. Por lo tanto, el potencial predictivo del incipiente calentamiento que ocurre mucho antes en el Pacífico tropical occidental, y que podría prolongar sustancialmente el plazo de pronóstico, no se utiliza en estos esquemas de predicción. Además mediante la experimentación con un modelo físico complejo también se demuestra aquí que existe una asociación significativa entre el aumento del contenido de calor en el Pacífico occidental y central y la ocurrencia de El Niño en el plazo más largo de 21 meses y no sólo durante el pico del contenido de calor ecuatorial 9 meses antes. Por lo tanto, el trabajo presentado en la tesis ofrece fuertes implicaciones para la posibilidad de mejorar las capacidades de largo plazo de otros modelos.

La definición de los predictores de ENSO a profundidades y regiones específicas en el océano y la atmósfera también requiere la disponibilidad de fiables mediciones superficiales y subsuperficiales de diversas variables climáticas. Aunque en el Pacífico tropical se han realizado mediciones esporádicas durante mucho tiempo como resultado de expediciones marítimas y misiones científicas, las mediciones regulares y de alta resolución sólo han comenzado con la era de los satélites en los años ochenta y, especialmente, con la colocación de un sistema de observación en los trópicos como resultado del Programa del Océano Tropical Atmósfera Global (1985-1994). Este Programa ha sido lanzado principalmente para profundizar la comprensión teórica de ENSO y para avanzar significativamente en su pronóstico. En esta disertación se verifica que la provisión de conjuntos de datos más completos y coherentes como resultado del Programa también mejora significativamente las capacidades de pronóstico

del modelo estadístico diseñado. Los pronósticos mejoran evidentemente y sustancialmente después de 1994, y el cambio es claramente pronunciado para los pronósticos a largo plazo que se basan en información subsuperficial de buena calidad sobre la estructura térmica oceánica en el Pacífico tropical occidental y central. Se establece así que los conjuntos de datos de las variables de mayor resolución temporal y espacial, y claves para la dinámica de El Niño son ahora lo suficientemente largos como para que los modelos de predicción estadística hagan un mejor uso e ellos.

En esta tesis también se explora la utilidad práctica de los pronósticos multi-anuales de El Niño. Teleconecciones principales son impulsadas por ENSO, lo que lo convierte en un buen predictor de los patrones climáticos en todo el mundo, pero sobre todo en las cercanías del Pacífico tropical. Un vínculo bien establecido en la literatura es el existente entre ENSO y los cambios climáticos locales en las zonas costeras de Ecuador. Después de un evento de El Niño hay una temperatura superficial generalmente más cálida y una precipitación significativamente intensificada en esa región. Estas dos variables, por otro lado, controlan la dinámica de la población de mosquitos y el número de criaderos de mosquitos, y de esta manera afectan la incidencia de dengue y otras enfermedades infecciosas propagadas por estos vectores. Por lo tanto, se realiza un experimento en el que las predicciones de largo plazo de El Niño se utilizan en un modelo de predicción del dengue para la provincia de El Oro en el sur de la costa de Ecuador. Se analiza el potencial de anticipación de las epidemias de dengue en la región, así como las perspectivas de desarrollar un sistema de alerta temprana para las epidemias de dengue. El análisis sirve como una demostración de la posibilidad para un servicio climático operacional en apoyo de la comunidad de salud pública en Ecuador.

To my grandmother who lives in my heart.
Parting with her has brought forward the realization
that life is also a phase of an oscillation.

Acknowledgements

I would like to express my sincere gratitude to those who have supported me throughout my PhD studies, and who have turned this period into and exciting part of my lifelong learning experience and self-knowledge.

Foremost, I would like to thank my supervisor Xavier Rodó for his encouragement, reassurance and confidence in me. He has provided excellent advice, expertise and professional guidance whenever I needed it. He has also helped me to improve and consolidate my scientific research skills, and has deepened my knowledge in the area of climate science. Needless to say, this thesis would not have been possible without him.

I would also like to thank my other collaborators in this research. First, I express my thankfulness to Joan Ballester for his informal supervisory role. I have gained a tremendous amount of practical and theoretical knowledge from my interaction with him, and I am privileged to have had the opportunity to discuss science with him on a regular basis. I further pay my respects and appreciation to Siem Jan Koopman who has introduced me to the mathematics of structural time series models and the state-space approach. He has also worked patiently and sometimes tirelessly with me on the coding issues that I encountered throughout my modelling work. Special recognition also goes to Axel Timmermann for sharing his in-depth theoretical understanding of the physics of the climate system, and ENSO in particular, and to Rachel Lowe for our excellent collaboration on climate and health research.

The Catalan Institute for Climate Sciences (IC3) provided a very good environment for my research, so I would like to also appreciate all my colleagues there who assisted me in the course of developing the thesis. I especially thank Carlos Dommar for his friendly support in countless occasions, and for our interesting scientific, but also ordinary-life conversations.

Finally, I would like to thank the people that reside in my heart, and whose love and support provide the context for all my accomplishments, including the work on the thesis: Vassil Tzanov for being the man I love, for his care and knowledge, and for helping me with technical problems of all sorts whenever I encounter them; my parents for loving me and always putting me first; my brother for being the person that I can eternally count on; and the rest of my family for their constant moral support and friendship.

List of Figures

List of Abbreviations

CPAC Central Equatorial Pacific

EN El Niño

ENSO El Niño Southern Oscillation

EPAC Eastern Equatorial Pacific

EUC Equatorial Undercurrent

HC Heat Content

LN La Niña

N34 Niño3.4

NECC North Equatorial Countercurrent

RB RossBell

SEC South Equatorial Current

SLP Sea Level Pressure

SOI Southern Oscillation Index

SST Sea Surface Temperature

TAO-TRITON Tropical Atmosphere-Ocean-Triangle Trans-Ocean Buoy Network

TOGA Tropical Ocean Global Atmosphere

WPAC Western Equatorial Pacific

WWV Warm Water Volume

Contents

Chapter 1

State of the art in El Niño Southern Oscillation Science (Introduction and Motivation)

1.1   Physical Processes in the Equatorial Pacific

The equatorial Pacific is characterized by a well-defined contrast in the oceanic and atmospheric conditions between the western Pacific (WPAC) and the eastern Pacific (EPAC). In the west the upper ocean is warm, the sea level pressure (SLP) is low, and in the atmosphere above there are cumulonimbus clouds, which determine an area of deep convection and heavy precipitation (Figure 1.1, Sarachik and Cane (2010)). The WPAC is also referred to as the "warm pool" due to these general characteristics. The ocean in the eastern part, on the other hand, is cold, with high pressure at the surface, and with scarce precipitation, and it is often referred to as the "cold tongue". Thus, the tropical winds, also called trade winds, typically blow from east to west, i.e. from the area of high pressure towards the area of low pressure. The warm moist air in the WPAC rises due to convection and reaches the upper-troposphere westerly branch of the Walker Circulation, through which it is carried to the east, where it descents and sustains the area of high pressure (Walker, 1924). The Southern Oscillation is defined through this contrast of high pressure in the east and low pressure in the west, and is measured by the Southern Oscillation Index (SOI, NOAA/CPC (2017)).

The easterly winds near the surface also drive a westward ocean current, and thus affect the ocean dynamics by deepening the thermocline - the border between the upper warm layer of the ocean and the deeper cold layers - in the west, and bringing it up in the east (Jin and An, 1999). This tilt of the thermocline is another measure of the strength of the Walker Circulation apart from the SOI (Jin and Neelin, 1993; Jin, 1997a). The westward surface winds are associated
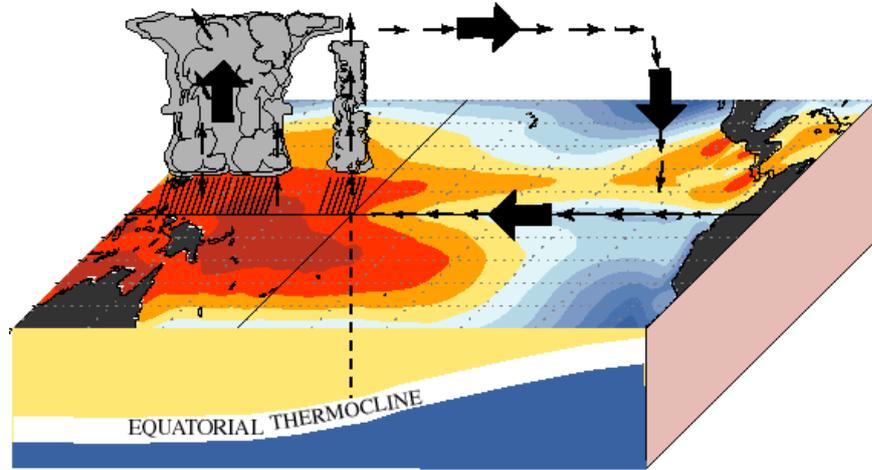
Figure 1.1: Neutral state of the ocean-atmosphere system in the equatorial Pacific in boreal winter (December-January). Illustrated is the contrast between warm conditions in the west and cold in the east (shading). Depicted are also the Walker Circulation, the area of convection in the WPAC and the ocean thermocline. Courtesy to the NOAA Climate Prediction Centre.

with poleward ocean motion in both hemispheres within the upper 50 metres of the equatorial ocean, and this water displacement is compensated by upwelling processes at the equator itself (Sarachik and Cane, 2010). However, since the thermocline is shallow in the east, the upwelled water there is colder than that in the west.

In the sense described above, the atmosphere and the ocean in the tropical Pacific act together in a coupled system. Bjerknes (1969) was the first to theorize the nature of this coupling framework, within which a positive feedback amplifies initial conditions in the ocean or the atmosphere. For example, an increase in the east-west zonal gradient of sea surface temperature (SST) intensifies the convection and the upper-tropospheric divergence, thus strengthening the Walaker Circulation (Bjerknes, 1966). Superimposed over the normal state of the system described so far is a cycle of warm and cold phases that is known as El Niño Southern Oscillation (ENSO), and the coupling between the tropical ocean and atmosphere is a fundamental mechanism for the occurrence of this phenomenon (Fedorov and Philander, 2001; Fedorov et al., 2003).

1.2 Description of the ENSO Cycle: Phases and Oscillatory Nature

ENSO is an irregular interannual oscillation and its return period is typically between 2 and 7 years (Philander, 1989). Its warm and cold phases are referred to as El Niño (EN) and La Niña (LN), respectively. Associated with EN is a relaxation of the surface easterly trade winds and the occurrence of westerly wind bursts, deepening of the thermocline in the east and shoaling in the west, increase of SST in the EPAC, and slowing down of the Walker Circulation (i.e. SOI is more negative, Figure 1.2a,c,e). Conversely, during LN there is an intensification of the trade winds, deepening of the thermocline in the west and shoaling in the east, and decrease in the SST in the EPAC, as well as strengthening of the Walker Circulation (Figure 1.2b,d,f). As a result of the strong coupling between the ocean and the atmosphere, as well as the Bjerknes feedback, an initial perturbation in the ocean is quickly transmitted to the atmosphere or vice versa. Thus, an initial warming of SST during EN acts to suppress the easterly trade winds and amplify the warm anomalies in the ocean or alternatively, initially suppressed easterlies result in warmer SST. The opposite happens during LN when enhanced easterlies act to decrease the SST or a drop in SST tends to reinforce them. The exact location of such initial disturbances, as well as of the peak anomalies may vary during the different events, which accounts for their general classification into two types - Eastern-Pacific or Central-Pacific ENSO - based on their properties and especially on their SST structure (Kao and Yu, 2009). However, the ocean-atmosphere coupling and the Bjerknes feedback only explain why such a perturbation would be sustained long enough and amplified for EN or LN to occur, but do not unravel the reasons for the transition between one phase to the other or back to neutral conditions corresponding to the normal state of the tropical ocean-atmosphere system. A number of theories have been proposed to provide a mechanistic explanation for these swings between warm and cold phases.
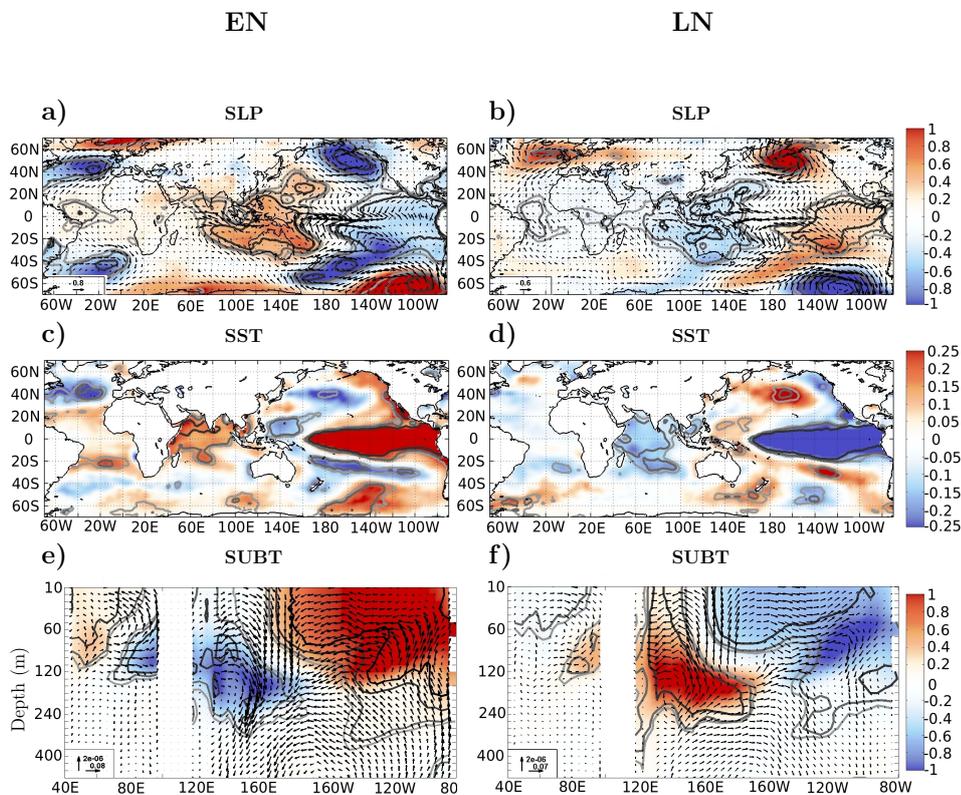
Figure 1.2: Composites of interannual monthly anomalies of **a)-b)** SLP (in $[Nm^-2]$, shading) and wind stress curl (in $[Nm^-3]$, arrows) from the NCEP/NCAR reanalysis (Kalnay et al., 1996), **c)-d)** SST (in $[°C]$, shading) - the NOAA-ERSST-V3 from NOAA/OAR/ESRL PSD, **e)-f)** subsurface potential temperature (SUBT; in $[°C]$, shading) and zonal and vertical currents (arrows) - the ECDA3.1 from the Geophysical Fluid Dynamics Laboratory (GFDL). Contours represent significance at the 90%, 95%, 99% and 99.99% levels. Composites are calculated for the EN and LN events in the period 1955-2010.

1.2.1   Oscillatory Theories

The Delayed Oscillator theory of Suarez and Schopf (1988) considers the effect of equatorially-confined waves propagating in the ocean subsurface. During ENSO SST anomalies are mainly in the EPAC, while wind anomalies are predominantly in the central Pacific (CPAC; Figure 1.2a,b). During EN the westerly wind anomalies in the CPAC trigger upwelling Rossby and downwelling Kelvin waves in the ocean below the anomalous circulation. Rossby waves travel westward, shallow the warm upper layer of the ocean and reflect off of the western boundary as downwelling Kelvin waves. Kelvin waves travel eastward to the EPAC, where they deepen the thermocline. The delay effect, which also renders the memory for the oscillation, comes from the fact that the propagation speed of Rossby waves is three times slower than that of Kelvin waves (Fedorov and Brown, 2009). Rossby waves then need much longer to traverse the Pacific basin ($\sim$ 9 months, Troccoli (2010)) and reverse the initial SST warm anomalies through re-adjustments of the thermocline. Within this theory the most important mechanism is the western boundary reflection of equatorial internal waves.

The Western Pacific Oscillator theory of Weisberg and Wang (1997) stresses on the importance of the generation of easterly wind anomalies in the equatorial western Pacific during the mature phase of EN. At this peak stage of El Niño anomalous off-equatorial anticyclonic wind causes the formation of easterly winds in the WPAC, which trigger Kelvin waves in the ocean, as well as upwelling processes, so that cooling proceeds eastward. Thus a negative feedback ensues and the system is transitioned to LN conditions.

The Advective-Reflective Oscillator theory of Picaut et al. (1997) focuses on zonal ocean currents in the equatorial Pacific. During EN the warm water that normally resides in the WPAC is advected eastward and anomalous westerly winds are present in the CPAC. They cause Kelvin waves to propagate eastward and initiate downwelling processes in the EPAC. Then, at the eastern ocean boundary they are reflected as Rossby waves that propagate westward and initiate upwelling processes in the WPAC. Since both types of waves are associated with a westward zonal current, the mean zonal current converges to the eastern edge of the warm pool, which is gradually pushed back to the WPAC.

The Recharge-Discharge Oscillator theory of Jin (1997a) and Jin (1997b) regards the dynamic feedback between the thermocline in the ocean and the wind stress at the surface of the ocean, and proposes that the tilt of the thermocline reacts immediately to changes in the wind stress, while wind stress reacts immediately to changes in SST (Figure 1.3). During EN the divergence of Sverdrup transport due to westerly wind anomalies in the CPAC and warm SST anomalies in the EPAC results into the discharge of warm water from the equatorial Pacific to higher latitudes. This causes the thermocline to be shallow over the whole domain, and trigger processes that bring up cold subsurface water to the surface. This is how cooling of the SST in the EPAC proceeds, and the oscillation gradually reverses its sign to LN. Then the opposite process starts with the buildup of warm water in the WPAC as a result of intensified trade winds, i.e. the recharge phase. It should be noted that the onset of the tropical Pacific discharge/recharge

always leads the SST anomalies in the EPAC by about 8-9 months and is seasonally dependent (Jin, 1997a; Meinen and McPhaden, 2000).

The Unified Oscillator theory of Wang (2001) attempts to reconcile all of the above theories by considering the SST anomalies in the EPAC, zonal wind stress anomalies in the CPAC, thermocline anomalies in the equatorial and off-equatorial region in the WPAC, and zonal wind stress anomalies in the equatorial WPAC. It includes the physics of all the other oscillators in a unified system of equations, and proposes that all these feedbacks are important for the alternation between the warm and cold phases of ENSO at different times of its development.
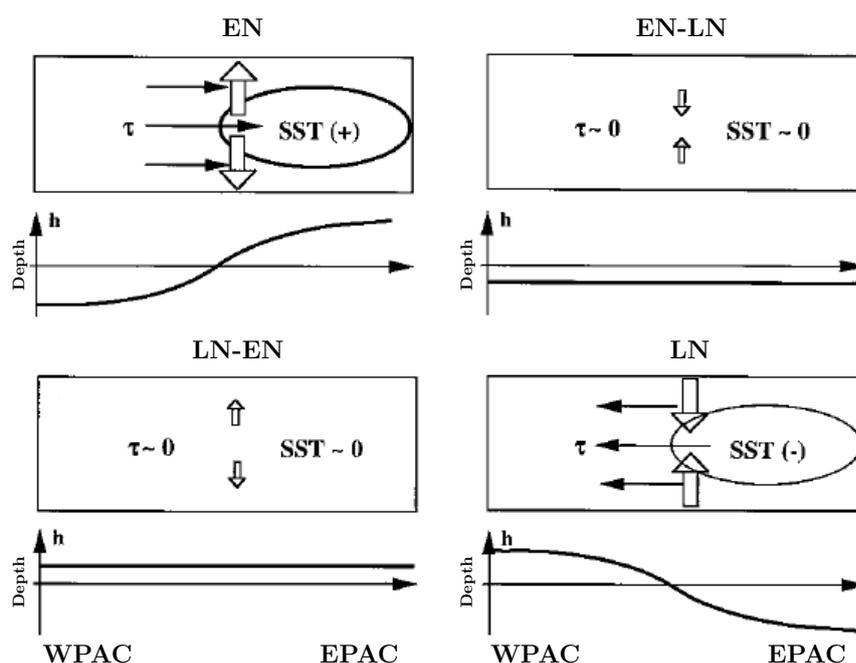
Figure 1.3: Schematic of the Recharge-Discharge Oscillator. Depicted from the top left panel in a clockwise direction are the: EN phase (Meridional Discharge), transition between EN and LN, LN phase (Meridional Recharge), transition between LN and EN. The boxes represent the domain of the equatorial Pacific, the ellipses are the areas of SST anomalies, the thick arrows indicate the discharge and recharge of mass from the equator, and the thin filled arrows represent wind stress anomalies. Below each panel is given the anomalous increase/decrease of the thermocline depth during the respective phase of the oscillation. Adapted from (Jin, 1997a).

1.2.2   Theory for the Evolution of El Niño

The recharge-discharge paradigm of Jin (1997a) is in essence the mathematical formulation of a theory first proposed by Wyrtki (1975). He suggested that on average about two years before the peak of an El Niño event in the equatorial Pacific, intensified easterly trade winds blow in the CPAC (Figure 1.4a,b), which lead to the accumulation of heat in the subsurface ocean in the WPAC. This causes the warm pool to expand and extend towards the CPAC, and positive SST anomalies to appear there as a result. Due to the strong coupling between the ocean and the atmosphere, these SST anomalies then feedback to the winds, so that the trades relax and westerly anomalies occur (Figure 1.4c,d), which further push the warm pool in an eastward direction during the growing phase of EN. The theory, however, did not discuss the exact physical mechanisms in detail, and failed to convey that the subsurface propagation of the anomalies towards the eastern portion of the basin is key. Ramesh and Murtugudde (2013) demonstrated that the buildup of subsurface heat in the WPAC/CPAC always accompanies the initial stage of EN development, and the start of its eastward displacement along the equatorial thermocline occurs at a particular step of this evolution about 18-20 months before the peak of an EN event, when the warm anomalies also start to appear on the surface in the WPAC and later in the CPAC (Figure 1.5) and start to weaken the trades. Moreover, their analysis suggests that this process has remained unchanged regardless of the major regime shift in the ocean in 1977 (Hare and Mantua, 2000).

The first two studies in this dissertation, Ballester et al. (2015) and Petrova et al. (2016), studied the dynamical mechanisms of the subsurface buildup of heat in the WPAC at the early stage of El Niño origination. Ballester et al. (2015) defined a salinity front at the eastern warm pool edge (between 160°E and 180°), where low-salinity waters from the WPAC, due to the normally intense rainfall there, meet high-salinity waters from the CPAC, due to the normally strong trade winds and enhanced evaporation in the area. It is further illustrated in this work that the position of this salinity front is different during EN and LN events.

On average between 14-30 months before the peak of EN intensified trade winds drive anomalous enhanced clockwise (anti-clockwise) wind stress curl in the central north (south) off-equatorial regions (Figure 1.4a,b), which drive an anomalous Sverdrup transport and meridional convergence of mass towards the equator, especially in the CPAC and WPAC. About 21-28 months before the mature EN La Niña-like conditions are found in the equatorial Pacific, with colder than normal surface temperatures in the EPAC and CPAC and warmer in the WPAC (Figure 1.5a,b), where the thermocline is at its deepest and the subsurface temperature anomalies are the most intense (Figure 1.6). The convection and precipitation typical for this region are shifted further west, which also means that the salinity front at the eastern edge of the warm pool is shifted westwards. This increases the zonal salinity contrast and the difference in the upper ocean density. At the same time, the intensified trades in the CPAC also strengthen the surface westward South Equatorial Current (SEC; Yu and McPhaden (1999)), especially in the vicinity of the salinity front. This contributes for the occurrence of horizontal convergence in
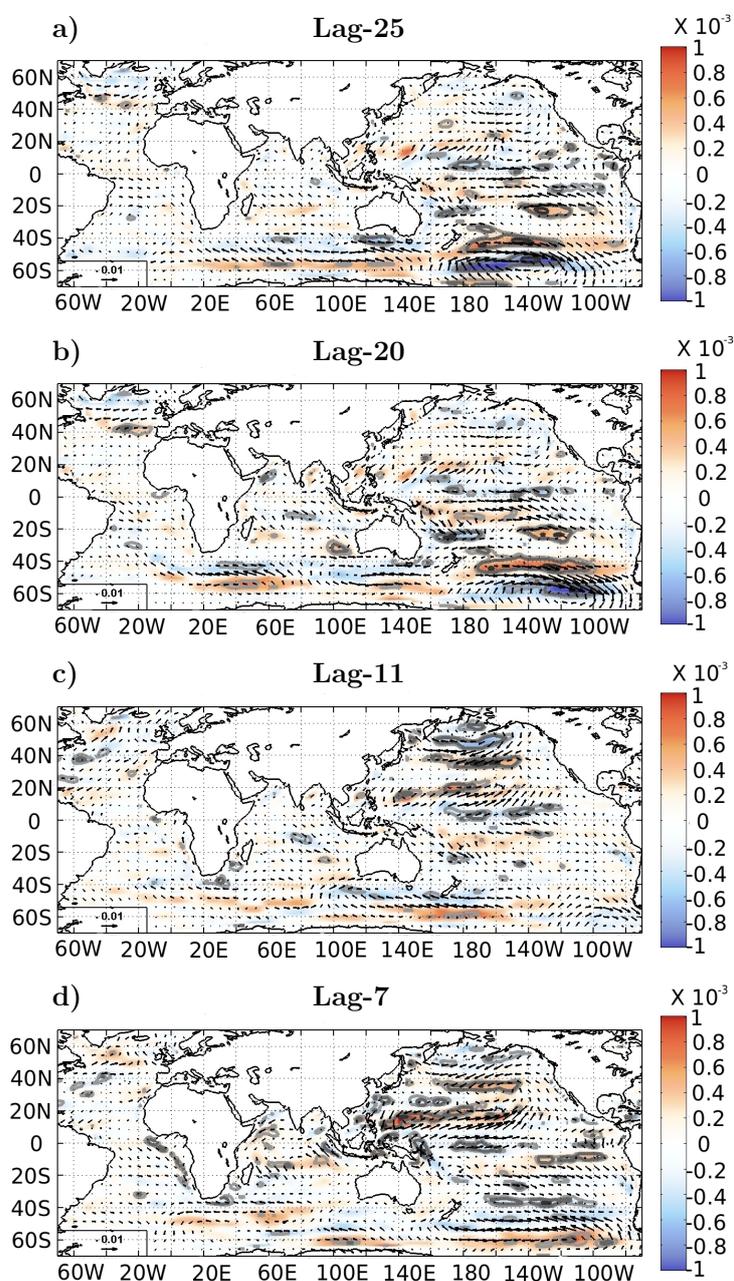
Figure 1.4: Composites of interannual monthly surface zonal and meridional wind stress anomalies (in $[Nm^{-}2]$, arrows) and wind stress curl (in $[Nm^{-}3]$, shading) from the NCEP/NCAR reanalysis (Kalnay et al., 1996) for **a)** 25, **b)** 20, **c)** 11 and **d)** 7 months before the winter peak of EN. Composites are calculated for the EN events in the period 1978-2010.
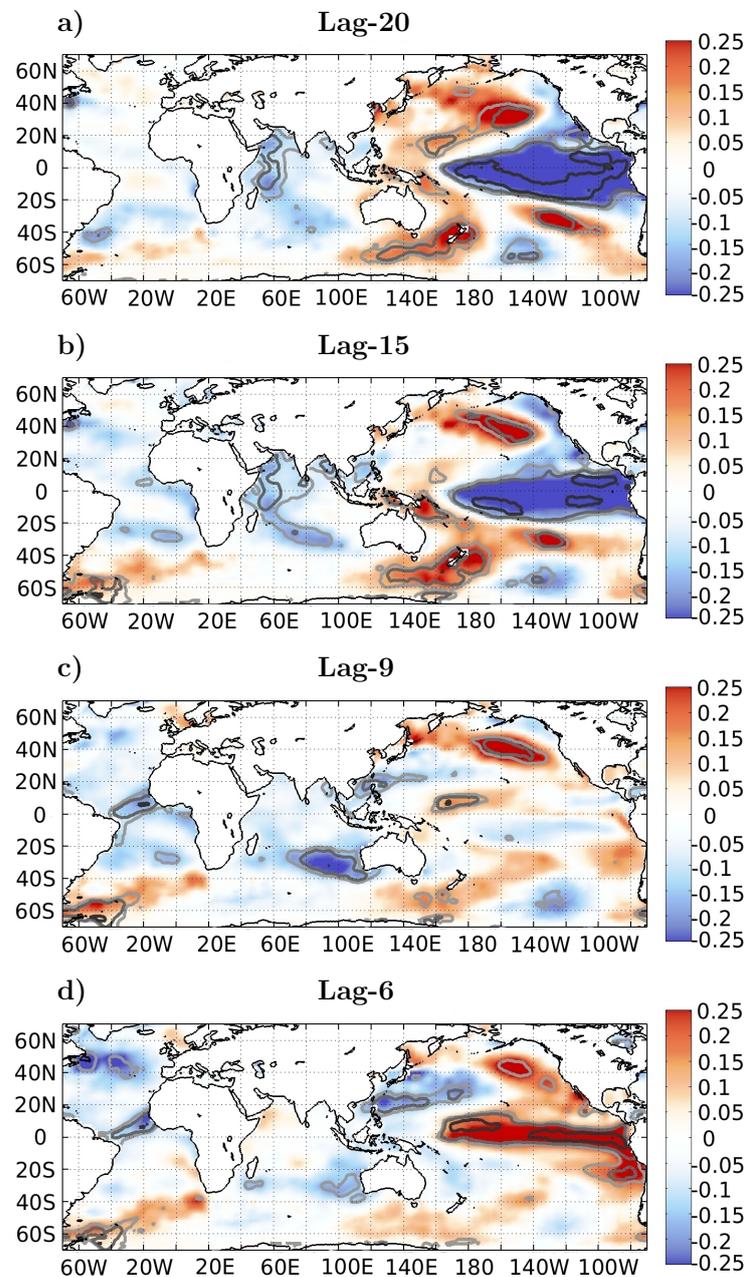
Figure 1.5: Composites of interannual monthly SST anomalies (in $[^\circ C]$, shading) from the NOAA-ERSST-V3, NOAA/OAR/ESRL PSD for **a)** 20, **b)** 15, **c)** 9 and **d)** 6 months before the winter peak of EN. Composites are calculated for the EN events in the period 1978-2010. Contours represent significance at the 90%, 95%, 99% and 99.99% levels.

the upper layer of the ocean (up to 75 metres depth), and divergence below this depth explained by the zonal current component and resulting from the strengthened eastward Equatorial Undercurrent (EUC) in the CPAC, which is driven by the east-west pressure gradient in the equatorial plane and directly affected by the strength of the easterly zonal wind stress (Ballester et al., 2015). Collectively these processes contribute for an overall downwelling motion in the warm pool and near the salinity front, which further increases the subsurface warm buildup there at these early stages of EN generation (Petrova et al. (2016); Figure 1.6).

The above description summarizes the interplay between zonal wind stress forcing in the CPAC and oceanic circulation and subsurface temperature responses in the WPAC at very long lead times with respect to the mature phase of EN, on the order of 2.5 years. Petrova et al. (2016) and the third study in the dissertation, Ballester et al. (2016a), further describe the mechanisms of subsurface warming and eastward propagation at the later stages of EN development. The piling up of warm water masses in the WPAC causes the warm pool to expand and to spread towards the CPAC in the subsurface through the EUC and the positive zonal advection along and below the thermocline. As a result of the upwelling in the CPAC due to the meridional overturning cells (Ballester et al., 2016a), this warming then also occurs in the upper 50 metres of the ocean about 15-17 months prior to the EN peak (Figure 1.5; Petrova et al. (2016)), where it induces an area of low pressure and, hence, the occurrence of westerly wind anomalies at the equator near the dateline some few months later (Figure 1.4c; McPhaden (2004)).

Such westerly wind anomalies play an important role for the coupling between the ocean and the atmosphere (Eisenman et al., 2005), and accompany the majority of warm ENSO events (McPhaden, 2004). Importantly, there is a deterministic component governing this anomalous zonal atmospheric circulation (Tziperman and Yu, 2007; Gebbie and Tziperman, 2009). It leads to an overall relaxation of the trade winds in the CPAC, to a decrease in the tilt of the thermocline (Jin, 1997a), and consequently, to the further propagation of the accumulated warm water masses towards the east through the geostrophic EUC and the eastward North Equatorial Countercurrent (NECC), which is driven by Sverdrup dynamics (Wyrtki, 1985). This is the onset of the recharge ENSO phase discussed earlier (Jin, 1997a), when the maximum subsurface warming occurs in the CPAC, unlike the previous stage of highest warming in the WPAC. This phase starts about 16 months before the El Niño (Jin, 1997a) and its peak conditions occur at about 9 months before. After this point cold anomalies start to appear in the subsurface in the WPAC, and an EN event grows in the EPAC. There is a weakening of the EUC in the WPAC (Figure 1.7), but enhancement in the EPAC, where warm subsurface anomalies have propagated along the thermocline via the zonal advective feedback (Jin and An, 1999), and start to be upwelled to the surface (Figure 1.7) by vertical currents and the Ekman pumping feedback (Jin et al., 2006). This is also a time when the SEC is weakened along with the eastern Pacific coastal upwelling and the western pacific downwelling (Ballester et al., 2016a). It is interesting to note that during the mature EN stage (4-6 months before the peak) there is already a negative meridional heat advection in the CPAC and EPAC in the proximity of the thermocline, which starts to weaken the heat anomalies in the subsurface and to prepare the system for the reversal of the sign of

the oscillation.

As seen above, the buildup and eastward propagation of heat in the equatorial Pacific necessarily precede EN events, and the time between warm events is roughly determined by the time needed to accumulate enough heat. Moreover, the amplitude of an event is in proportion to the magnitude of the anomalous heat content prior to the event (McPhaden, 2004; Ballester et al., 2016b).
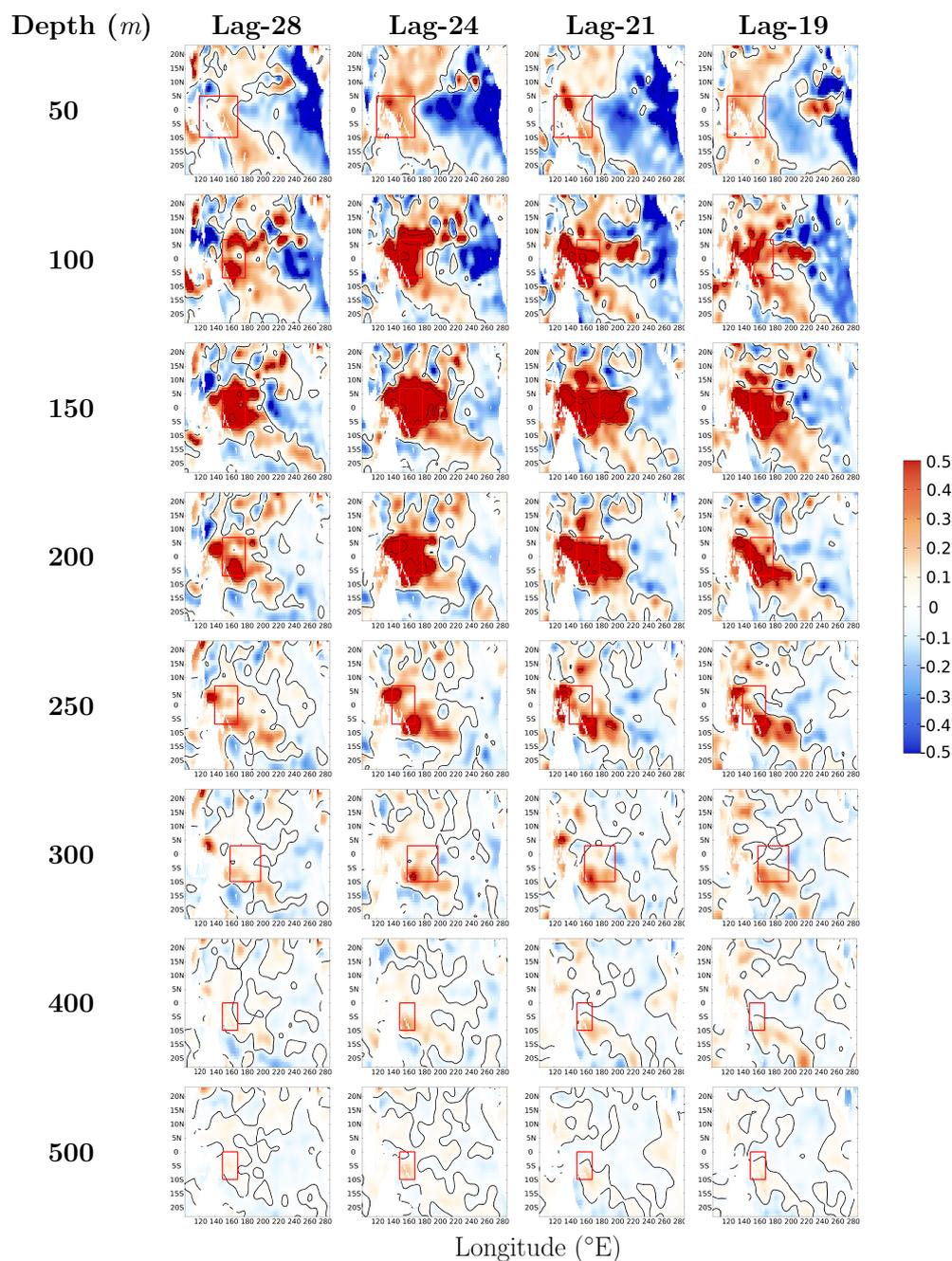
Figure 1.6: Composites of interannual monthly subsurface temperature anomalies between 50-500 metres from the Subsurface Temperature and Salinity Analyses by Ishii et al. (2005) depth at lead times of 19, 21, 24 and 28 months ahead of the EN peak. Red boxes indicate regions for derivation of precursors. Composites are with respect to all EN events in the period 1978-2012.
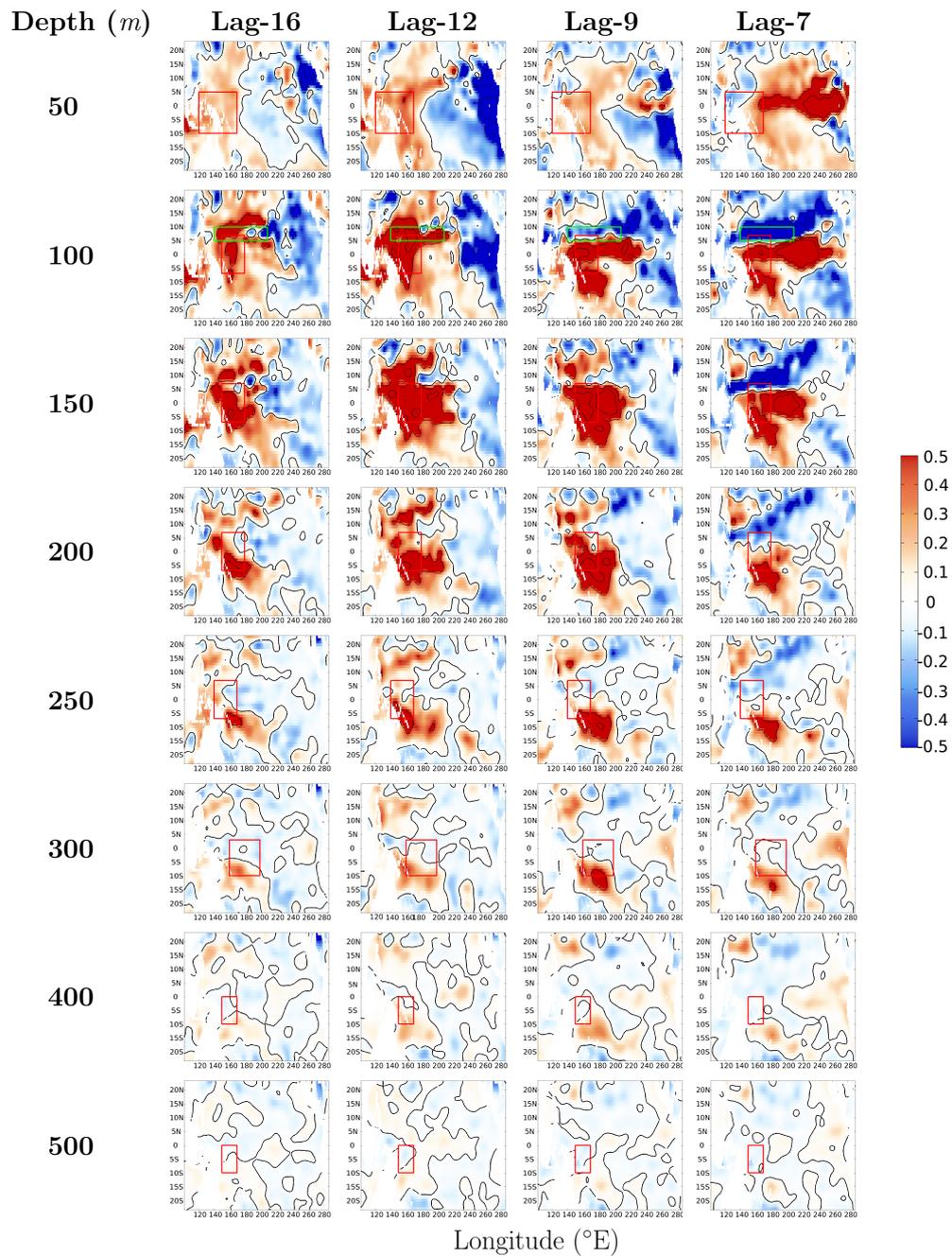
Figure 1.7: As in Figure 1.6, but at lead times of 7, 9, 12 and 16 months ahead of the EN peak. Red and green boxes indicate regions for derivation of precursors. Composites are with respect to all EN events in the period 1978-2012.

1.3   ENSO Predictability

ENSO has been recognized as being the leading source of interannual atmospheric variability, and it is a main triggering mechanism for atmospheric teleconnections worldwide (Rasmusson and Carpenter, 1982; Kiladis and Diaz, 1989). Therefore, forecasting warm and cold events long before their occurrence is not only highly desirable because of the large-scale anomalous weather patterns that they cause in different parts of the world, but also because of the vital role of such forecasts for the overall seasonal forecasting endeavours globally (Goddard et al., 2001).

The evolution of ENSO events is predictable many months in advance mainly as a result of the nature of the atmosphere-ocean coupling, which is strong in the tropical region (Sarachik and Cane, 2010). There SST directly influence the atmospheric circulation, as the ocean supplies the air above it with moisture through evaporation processes. The trades converge to the warmer area in the WPAC, where moist air condenses and latent heat is released. These processes of evaporation, condensation and release of latent heat control the Walker Circulation, and hence the stronger diabatic heating in the tropics (Fedorov, 2008). Since the distribution of SST in the tropical ocean there defines the statistical distribution of cloud heating on a monthly timescale, in this way the slow dynamics of SST in the ocean can determine the evolution of the statistics of the atmosphere (Shukla, 1981; Barnett et al., 1993). Moreover, as a result of the coupling, the atmospheric heat fluxes also directly impact the evolution of the ocean surface temperature (Gebbie and Tziperman, 2009). ENSO forecasts are in effect forecasts of the SST in the EPAC or the equilibrium between the SST and the statistics of the atmosphere (Jin et al., 2008). Thus, the slow variations of the tropical atmosphere are determined by changes in the boundary conditions of the ocean and vice versa (Latif et al., 1999).

Apart from monitoring the SOI, the most common way for identifying ENSO events is by measuring the monthly SST anomalies averaged in an area extending over the EPAC and the CPAC, i.e. the Niño3.4 region (N34; [5°N-5°S, 170°W-120°W]). This is also the region of strongest SST anomalies associated with EN and LN events (Figure 1.2c,d). There are several other temperature indices specified in different regions over the equatorial Pacific, which are used to distinguish different types of events. The Niño1+2 index is defined in the immediate proximity of the South American coast ([0°-10°S, 90°W-80°W]), where *El Niño* was first termed by the local Peruvian population to refer to the ocean warming that occurs near Christmas time, as El Niño actually means the Christ child. The Niño1+2 nowadays signals the so-called *coastal El Niño* events, when SST warming occurs exclusively near the coast of South America. There are also the Niño3 region ([5°N-5°S, 150°W-90°W]), which was first used to define ENSO events, but was later replaced by N34 as a more appropriate one (Trenberth, 1997), as well as the Niño 4 region ([5°N-5°S, 160°E-150°W]) used to characterize better EN with peak SST anomalies occurring in the CPAC (Kao and Yu, 2009). The Oceanic Niño Index (ONI) is also used. It is defined over the same region as N34, and represents a three-month running mean of SST anomalies there. In the thesis N34 is used throughout for analysis, modelling and simulation purposes.
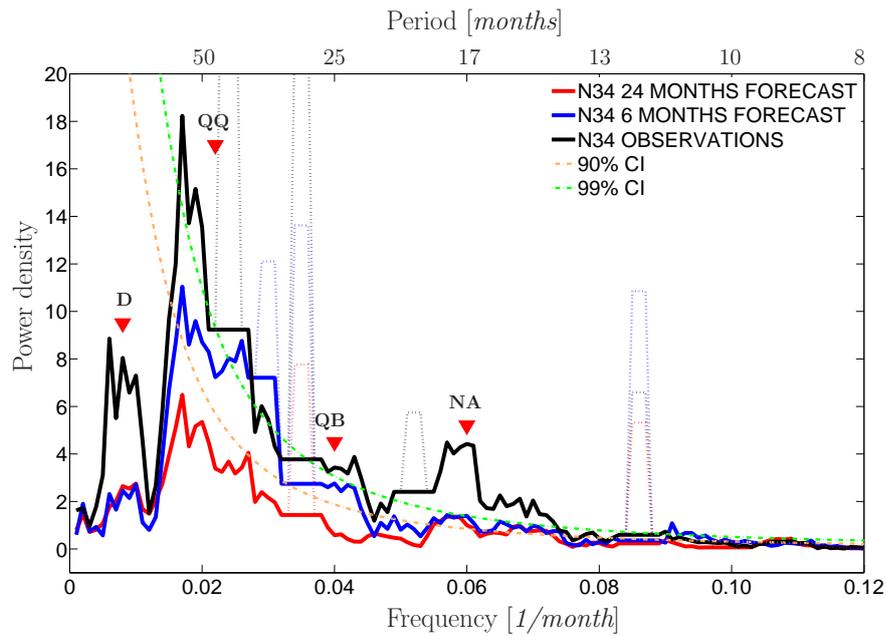
Figure 1.8: Multi Taper Method (MTM) power spectra for the observed N34 time series (black), and predictions with the dynamic components model (Petrova et al., 2017) at 6 month lead time (blue) and 24 month lead time (red). The solid lines indicate the power density, dotted lines harmonic peaks and dashed lines confidence levels based on a red noise null hypothesis. The red markers indicate the near-annual (NA), quasi-biannual (QB), quasi-quadrennial (QQ) and decadal (D) modes of variability.

The spectrum of the N34 reveals variability on interannual timescales with several dominant peaks (Figure 1.8). A distinguished feature is its quasi-periodicity with a main quasi-quadrennial and a quasi-biannual peak (Rasmusson and Carpenter, 1982; Kirtman and Zebiak, 1997). The theoretical predictability of ENSO more than one year in advance is actually implied by these two spectral peaks (Latif et al., 1999).

The theoretical predictability of the phenomenon depends on its mechanistic explanation. One of the two leading paradigms for its physics suggests that it is due to a stable interaction between the ocean and the atmosphere, representing a damped oscillation (Lengaigne et al., 2004). If this theory is correct, ENSO predictability could not be extended much more than a year in advance, which is the time needed for an initial disturbance of the coupled system to grow. According to the alternative paradigm, however, ENSO arises due to unstable interactions between the ocean and the atmosphere, and is rather a self-sustained oscillation (Cane et al., 1986; Chen et al., 2004). Within this framework its predictability depends on the duration of time for which the equatorial ocean thermocline retains a stable position, which represents to so-called memory of the ocean (Neelin et al., 1998). In this case the evolution of the system could be predicted much longer than a year in advance as is achieved in a retrospective mode by Barnett et al. (1993), Chen et al. (2004), Petrova et al. (2016) and Gonzalez and Goddard (2016), and this is true especially for the larger events, which are normally better predicted than the weaker ones (Jin et al., 2008; Barnett et al., 1993; Chen et al., 2004). The theoretical understanding of ENSO and its predictability have been considerably advanced as a result of the equatorial Pacific monitoring and observing system deployed by the Tropical Ocean Global Atmosphere (TOGA) Program (McPhaden and Yu, 1999; Latif et al., 1999). The studies included in this dissertation and especially Petrova et al. (2017) add to the evidence by previous work that if comprehensive subsurface tropical ocean information is utilized, ENSO is predictable at least a cycle length in advance (Figure 1.8), i.e. years ahead rather than months (Cane et al., 1986; Chen et al., 2004; Gonzalez and Goddard, 2016). It should be highlighted, however, that even in this case there are factors that can still limit the predictability of individual events such as random noise (Latif et al., 1999; McPhaden and Yu, 1999; Fedorov et al., 2003), which could modulate the strength of the coupling between the tropical ocean and atmosphere (Kirtman and Schopf, 1998); nonlinear interactions with the annual cycle (Jin et al., 1994; Tziperman et al., 1994); the exponential growth of initial conditions errors (Chen et al., 1995); as well as fluctuations of the mean state of the ocean on decadal timescales (Latif et al., 1999).

There are statistical and dynamical models for ENSO forecasting, as well as hybrid systems that are typically composed of a dynamical ocean and a statistical atmosphere (Cane et al., 1986; Jin et al., 2008; Barnston et al., 2012). The first dynamical model to correctly predict EN in real time was the Zebiak-Cane intermediate complexity model (Zebiak and Cane, 1987), which forecast the onset of the 1986/87 EN and set the beginning of operational ENSO predictions. Forecast skill has not improved dramatically since then, and no forecasting strategy is proven to clearly outperform the others, although General Circulation Models (GCMs) appear to have an edge when predicting the boreal spring season and across this season. Still, they have no clear

superiority when predicting the boreal winter season, which is when ENSO events peak (Barnston et al., 2012). A number of models now produce real-time operational forecasts for the SST in the N34, and these predictions are summarized monthly in the *ENSO prediction plume* (IRI, 2017). Dynamical models normally release ensemble predictions and the ensemble mean is utilized as a deterministic forecast, while statistical models typically produce one single forecast (Barnston et al., 2012). The most standard measures of skill for such predictions are the correlation (*corr*) between the predicted and the observed values for a long period of time, as well as the root mean square error (*rmse*). Since *corr* measures the correspondence between the phase of the forecast and the observations, but it does not account for the amplitude, it is not a sufficient measure of skill, especially considering that all current models have problems with simulating correctly the amplitude of the events (Sarachik and Cane, 2010). In this sense, the *rmse* is an important additional metric for assessing model performance. In order to validate a given forecast system, a number of retrospective forecasts are performed of past observations, and the *corr* and *rmse* between the two are obtained and compared to those of other validated schemes. This is also a way to identify persistent biases of the model and possibly correct the forecasts in a post-processing procedure. A comparison of operational ENSO models has indicated that no model is demonstrably better than the rest (Sarachik and Cane, 2010; Barnston et al., 2012), and the best probability guess of future outcomes is obtained by combining the forecasts from all individual models into multi-model ensembles (Jin et al., 2008), which reduces systematic errors (Palmer et al., 2000). Therefore, the more model simulations are available, the better is the overall forecast and its usefulness to potential end users.

As briefly mentioned above, there are a number of challenges for the ENSO prediction systems. For example, the mean background state and the mean annual cycle represent important portions of the interannual variability in the equatorial Pacific (Zebiak and Cane, 1987; Suarez and Schopf, 1988), which are often not well simulated by the existing models. There are errors in the basic state, as well as the phase and amplitude of the annual cycle, especially with increasing lead time, which also affect the forecast of the interannual variability (Jin et al., 2008). Moreover, a clear relationship has been identified between the correct simulation of the annual cycle and the phase-locking of ENSO to the annual cycle, and this especially concerns the amplitude variation of the events, as there is a linear relationship between the intensity of ENSO events and the intensity of the annual cycle in dynamical model simulations (Jin et al., 2008). There is a pronounced annual periodicity in the seasonal cycle of the EPAC propagating westwards, and a semi-annual periodicity in the WPAC. Another issue for instance is that EN and LN events are typically better predicted by models than neutral conditions, and EN also tends to be more predictable than LN, given that the amplitude of the cold phase is often underestimated in simulations (Kirtman and Zebiak, 1997). The growing phase of the events is generally better captured by models than the decaying phase (Jin et al., 2008), with operational models predicting a continuance of SST anomalies after the end of the events, especially at the longer lead times (Barnston et al., 2012). Such a feature is also evident in the forecasts shown in Petrova et al. (2016) and Petrova et al. (2017). Overall, models tend to capture ENSO transition stages with

delay, with a tendency toward persistence (Barnston et al., 2012). Since this appears to be a systematic error for the majority of the schemes, it represents a problem that could potentially be fixed by post-processing and bias corrections of the forecasts. The most plausible explanation for this in the case of the warm events is the fact that the EPAC is warmest during spring due to the annual cycle variability (Sarachik and Cane, 2010). Models, however, are often unable to simulate the suppressed amplitude of the annual cycle during the EN phase, and thus tend to extend the event until later in the calendar year instead of terminating it in spring.

The predictive capacity of models also varies from one decade to another (Balmaseda et al., 1995; Kirtman and Schopf, 1998; Chen et al., 2004), and the reasons for this are as yet not well-understood (Sarachik and Cane, 2010). The modulation of SST on decadal timescales represents part of the natural variability in the tropical Pacific (see Figure 1.8), with some studies explaining it through the interaction of the tropical and extratropical ocean, which produce variations in the tropical thermocline (Gu and Philander, 1997). Others suggest that uncoupled atmospheric noise could be the cause for decadal ENSO variability (Kirtman and Schopf, 1998).There is also global warming, whose timescale is mixed with that of decadal variability. Usually, decades with high interannual variability are also decades that are associated with relatively high predictability when the system is controlled by the oscillatory behaviour described earlier, and when the signal-to-noise ratio is high, which maximizes the model correlation skill (Barnston et al., 2012). Conversely, decades with low interannual variability are associated with less predictability (Chen et al., 2004), when the atmospheric noise is a dominating component (Kirtman and Schopf, 1998). In this second case, it has been shown that the different level of predictability is independent of the type of noise, and the noise in fact produces changes in the mean state of the tropics. This also implies that the prescription of stationary climatology in models is not always appropriate, because the mean background slowly changes (Barnett et al., 1993). However, this uncoupled noise is not necessarily unpredictable itself, but represents variations that are different from the main ENSO oscillator. In this sense, at least part of it could be predictable by improving the prediction of external noise. During decades with high predictability some forecasts with a coupled model have been shown to be skilful for lead times of up to 15-18 months, while during decades with low predictability the same forecasts are only skilful for up to 5-6 months lead (Balmaseda et al., 1995; Kirtman and Schopf, 1998). Petrova et al. (2017) has demonstrated that the inclusion of a decadal cycle in the statistical prediction system described therein, and thus accounting for some of the decadal variability (Figure 1.8), has led to better retrospective forecasts of at least some of the historical ENSO events.

Forecasts issued during and before the Northern Hemisphere spring season are problematic for all models (Barnston et al., 2012) - this represents the so-called "spring predictability barrier" (Webster, 1995). Reasons for this barrier are the low signal-to-noise ratio in spring (Goswami and Shukla, 1991), and the fact that during this season the system loses part of its memory (Zebiak and Cane, 1987). Moreover, SST anomalies are less pronounced in spring as a result of the seasonal locking of ENSO (and the quasi-biannual oscillation in particular) to the annual cycle (Clarke and van Gorder, 1999). This makes forecasts more difficult due to the higher

proportion of noise in the coupled system. In the observations there is a marked seasonality of SST in the EPAC, with a peak in the variance in winter and a through in spring, which is a well-established connection between the annual cycle and the interannual variability in the region (Jin et al., 2008). This could further explain the drop in forecast skill in this season, as well as the fact that it normally increases in the following season, especially in the case of dynamical model predictions (Xue et al., 1994; Balmaseda et al., 1995). Results from some earlier studies with dynamical models, however, indicate that the decrease in forecast skill in spring is only modest as compared to the drop in persistence skill (Chen et al., 1995, 2004), and that it is consistent with the seasonal minimum in signal variance at this time. Petrova et al. (2017) reports similar results with a statistical dynamic components model, which support the conclusion that the spring predictability barrier is not inherent to the climate system, but is more likely model-dependent (Chen et al., 1995). Another study, Balmaseda et al. (1995), found that this barrier is clearly present in some decades and much less prominent in others, claiming that the major reason for this is a change in the seasonal dependence of ENSO from one period to another. According to this view and the results in the study, it is more strongly locked to the seasonal cycle in those decades when the spring barrier is present, with the transition between cold and warm phases occurring mostly in spring.

As discussed in the previous sections, all ENSO events are preceded by anomalies in the equatorial Pacific Ocean heat content (HC), and similarly to SST, the predictability of HC is also seasonally dependent, with a winter barrier, rather than a spring one, weaker than for SST and in agreement with their general lag relationship (Zebiak and Cane, 1987; Zebiak, 1989; Balmaseda et al., 1995; McPhaden, 2003). As with SST the drop in HC forecast skill is much more pronounced in some decades than in others, and is associated with lower subsurface variance in boreal winter and thus increased sensitivity to noise (Xue et al., 1994). The fact that the predictive skill drops in different times for the HC and SST means that the important forecast information is not lost simultaneously on the ocean surface and in the subsurface, which suggests that skill does not have to be totally lost as a result of seasonal dependence, especially considering that correlations for warm water volume (WWV, a proxy for HC) are highest at 7-10 months lead time starting in spring between February-May, and coinciding with the onset of ENSO events, but when SST anomalies are less pronounced and ambiguous (McPhaden, 2003). This has very good implications for the overall predictability of the phenomenon at medium to long lead times. At the same time, some models demonstrate a drop in skill at a lead time of about 12 months for the December, January, February (DJF) target season (Barnett et al., 1993), including the model described in Petrova et al. (2017) (see Figure 3 therein), which indicates that the drop in HC predictability in winter also affects the predictability of SST at this lead time. However, the overall skill of models is shown to increase for lead times between 6-12 months when they are fed with accurate HC variation information (Latif et al., 1999), which holds promise for the theoretical long-lead predictability.

1.4   ENSO Modelling

The components of a forecasting system are: observations, assimilation, analysis, initialization, forecast, and validation (Sarachik and Cane, 2010). Dynamical and statistical models differ in the type of input and output information that they require and provide. A typical physical ENSO model is built in terms of the partial differential equations representing the discretization of the ocean-atmosphere system, uses observations of oceanic and atmospheric variables, and initializing techniques in order to render precipitation, winds, SST and thermocline changes. More comprehensive complex models also simulate the mean annual cycle, while some simpler models prescribe it and calculate anomalies with respect to the observed climatology (Zebiak and Cane, 1987). In such a basic ENSO model the dynamical quantities in the ocean are calculated using the linear shallow water equations and then they are applied into an advective temperature equation (Zebiak, 1985). The shallow water approximation and equations also govern the state of the atmosphere, so that they can be used to build a full ocean-atmosphere coupled model (Zebiak, 1985; Sarachik and Cane, 2010). The most general form of the momentum and continuity equations that represent the fluid motion are:

$$\frac{du}{dt} - fv = -g\frac{\partial h}{\partial x} + F^u + D^u \tag{1.1}$$

$$\frac{dv}{dt} + fu = -g\frac{\partial h}{\partial y} + F^v + D^v \tag{1.2}$$

$$\frac{\partial h}{\partial t} + \frac{\partial(\eta u)}{\partial x} + \frac{\partial(\eta v)}{\partial y} = 0 \tag{1.3}$$

and the total derivative for the respective quantities is defined as follows:

$$\frac{d\alpha}{dt} = \frac{\partial(\alpha)}{\partial t} + \frac{\partial(u\alpha)}{\partial x} + \frac{\partial(v\alpha)}{\partial y} \tag{1.4}$$

In the above equations u and v are the horizontal velocity components, f is the Coriolis force term, h is the anomalous deviation of a pressure surface from its mean height position H in the fluid, $\eta = H + h$ and F is a forcing term that in the case of the ocean represents the component of surface wind stress, while for the atmosphere it is an external momentum forcing. D is a linear dissipation term. Conservation of mass could be used for the calculation of the vertical velocity component at the base of the surface ocean layer, while in the case of the atmosphere a vertical structure is assumed that excludes vertical energy propagation (Zebiak, 1985). Extensive information about the main assumptions and detailed derivation of the equations can be found in Zebiak (1985); Neelin et al. (1998); Sarachik and Cane (2010) and others.

Statistical models, on the other hand, typically rely on some form of regression (Penland and

Magorian, 1993; Xue et al., 1994; Clarke and van Gorder, 2003), neural networks (Tangang et al., 1997) or analogues (Van den Dool, 2007), use area-averaged SST, SLP, wind stress, and inte-grated HC as input to predict the SST in one of the equatorial ENSO regions, most commonly the N34 (Barnston et al., 2012). As discussed earlier, both types of modelling strategies are not devoid of problems. In the case of dynamical models they are caused by high-frequency atmospheric noise, initialization errors due to data analysis and assimilation that grow with simulation time, as well as biases as a result of incorrect numerical parametrizations. In the case of statistical models atmospheric noise is also a problem, along with the need for long data records in order to infer stable relationships between the predictors and the predictand, and the coarser spatial and temporal resolution of the variables used, i.e. area-averaged monthly data. In general terms, both types of models manage to reproduce the overall ENSO variability reasonably well at short and medium lead times, with some problems and discrepancies that arise due to reasons described above. However, long-lead forecasts on the order of more than one year of EN and LN events are still scarce, and there are none at the operational level. At the same time the climatic impact of ENSO is far-reaching and large-scale teleconnections (Kiladis and Diaz, 1989; Rodó et al., 2006), and the affected communities could benefit immensely from forecasts of the events longer time in advance.

In recent years the efforts of the ENSO modelling community have been concentrated on the development and improvements of more complex dynamical models. Statistical models at the same time have been left behind (Barnston et al., 2012), and systematically underfunded due to the assumption that they have reached their optimal skill level. In this dissertation it is demon-strated that it is just the opposite, as statistical forecasting has yet vast unexplored potential also at the very long lead times of several years. In a statistical prediction framework the main objective is to identify a set of reliable precursors that would optimally forecast the evolution of the predictand variable of interest. As discussed in the previous sections, it has been suggested by earlier work that the evolution of ENSO cannot be forecast without having information about the upper ocean HC, and the long-lead experimental forecasts of the phenomenon suggest the important role of the subsurface ocean, which has a substantial thermal inertia (Cane et al., 1986).

As mentioned earlier, SST in the N34 tend to lag equatorial HC by some months, and thus fluctuations in HC are generally regarded as a good precursor of the ENSO events (McPhaden, 2003). Similarly to Chen et al. (2004); Petrova et al. (2016); ? and Petrova et al. (2017) have reached the conclusion that atmospheric noise is responsible for irregularities in the exact onset, termination and amplitude of EN and LN, but the overall dynamical features of the events are governed by the state of the subsurface ocean. In addition, wind bursts occurring before ENSO are to a large extent controlled by the oceanic dynamics and are predictable in this sense (Chen et al., 2004; Gebbie and Tziperman, 2009). Thus, a good ENSO forecasting system needs to necessarily incorporate properly the relevant information from the ocean internal dynamics, as well as the surface atmospheric dynamics. In addition, quieter periods when the interannual variability is low could also be predicted, if a forecasting system is fed with the most relevant

data information (Chen et al., 2004), which has to be carefully selected since the background state is more dominant in such periods.

As described in Section 1.1.2, an initial increase of the south-easterly trade winds more than two years ahead of an El Niño drives the accumulation of warm water mass in the far western Pacific Ocean. Then months later, as a result of the coupling mechanism and the positive feedback responses of the tropical ocean and atmosphere, a decrease in the equatorial trade winds leads to a surge eastward of the accumulated warm water. This process warms the normally cooler subsurface water in the EPAC, and upwelling and zonal advection in the region cause the SST there to warm. The first simple linear ocean models from the 1970s and 1980s were generally able to reproduce these fluctuations in the equatorial thermocline only with forcing from the observed equatorial zonal wind anomalies during ENSO (Zebiak, 1985). These early results proved the importance of the changes in the equatorial subsurface thermal structure and zonal winds for El Niño generation. Thus, the novel forecasting model described in Petrova et al. (2016) and developed in the dissertation is also designed taking into account these earlier findings. In particular, this study identified from composite analysis of subsurface temperature (Figure 1.9) and zonal wind anomalies the WPAC as the region where the first warming associated with the origination of EN events occurs. It also showed the potential of these long-lead patterns of the subsurface temperature and surface wind stress in the WPAC and at a later stage in the CPAC as promising predictors for ENSO.

There are a couple of innovative approaches in this work that distinguish the use of precursors in Petrova et al. (2016) from earlier statistical modelling strategies. The first one is the definition of specific regions ("hot spots") for extraction of the predictor variables based on the dynamical processes occurring in the ocean-atmosphere system prior to EN. Since the warm anomalies are concentrated in particular areas in the subsurface, especially at the very long lead times (Figure 1.9), the traditional approach in statistical forecasting of using integrated measures of equatorial HC is not appropriate to successfully detect these fundamental early local processes. Therefore, using specific regions and depths, which is possible due to the availability of the Tropical Atmosphere Ocean-Triangle Trans-Ocean Buoy Network (TAO-TRITON) in the tropics (McPhaden et al., 1998), has allowed the vast extension in time of the predictive capacity of the forecasting model (Petrova et al., 2017). The TOGA Program and the placement of a high-resolution monitoring system in the equatorial Pacific was primarily driven by the fact that the very strong 1982/83 EN was not only unpredicted, but it was also noticed very late in its evolution (McPhaden et al., 1998). One of the main goals of the Program was to make possible the regular prediction of ENSO by operational and experimental forecasting systems, something that was already accomplished to a certain degree for the dynamical modelling of the phenomenon by the prominent 1997/98 EN (Kirtman and Schopf, 1998). For statistical modelling, though, the resources of the observing system have remained largely unexploited, as a longer reporting period for the data sets is typically required (Barnston et al., 2012). In Petrova et al. (2017) it is advocated that the data sets of detailed subsurface temperature are now sufficiently long to make use of them for long-lead forecasting. Moreover, it is recommended

that other statistical schemes should be updated in a similar manner to apply the subsurface information in a more efficient and target-oriented way. The second innovation of the statistical forecasting approach is the inclusion of different precursors at different lead times in agreement with the general evolution of a warm event described earlier. In addition to taking into a detailed account the EN dynamics, the selection of variables is also made based on rigorous statistical fitting criteria. Thus, depending on the time when a particular forecast is started, different

predictors are added to the core model configuration that will be explained next.

In the recent decades structural time series modelling and state space approaches have gained a lot of attention and applications, first in the area of engineering, and later in that of economics and econometrics. They have deservingly been established as highly useful forecasting tools in these fields, due to their advantageous features. One such feature is the formulation of the time series model in terms of components that have straightforward interpretations, for example trend, seasonal cycle and cycles on other relevant timescales (Harvey and Koopman, 2000). The inclusion of regression precursor variables in such configurations is also effortless and flexible. Furthermore, they are especially suited for modelling phenomena that is characterized by multiple variability modes, mixed frequencies and non-linearities (Harvey et al., 1998; Durbin and Koopman, 2012). ENSO is just such a phenomenon. Therefore, in Petrova et al. (2016) such a univariate structural time series model is developed with several dynamic components and regression precursor variables from the tropical ocean and atmosphere as described earlier in this Chapter. The dynamic components are stochastic functions of time, driven by independent disturbances, and extracted from the N34 time series using the Kalman Filter (Kalman, 1960). Forecasts and unknown parameter estimations are also generated with the Kalman Filter. The different components are first modelled separately and then put together in a single model using state space techniques (Petrova et al., 2016). The main assumption of the state-space approach is that the evolution in time of the system of interest is determined by an unobserved series of vectors, collectively the state-space vector, which contains the unknown parameters associated with the dynamic components and the regression variables (Durbin and Koopman, 2012). The state-space analysis is applied in order to extract the properties of this vector from the existing time series observations of reality, and this information is used for forecasting.

A very important aspect of time series modelling is to discover the physical laws that trigger its generation, which precludes a good understanding of the underlying dynamical processes. Since ENSO physics is now relatively well-understood, it is expected that a successful statistical model of the phenomenon incorporates the important dynamics that was described in the previous sections. Therefore, in order to account for as much of the ENSO variability as is needed to predict its future evolution, the following dynamic components are included in the proposed statistical model: trend, seasonal, three cycles, noise and various regression precursors as described previously (Petrova et al., 2016). A trend is necessary as there are slow changes in the mean background of the tropical ocean-atmosphere system. Proper representation of the seasonality and of the seasonally varying coupling between the atmosphere and the ocean is crucial for successful forecasting, as well (Barnett et al., 1993). For instance, dynamical models
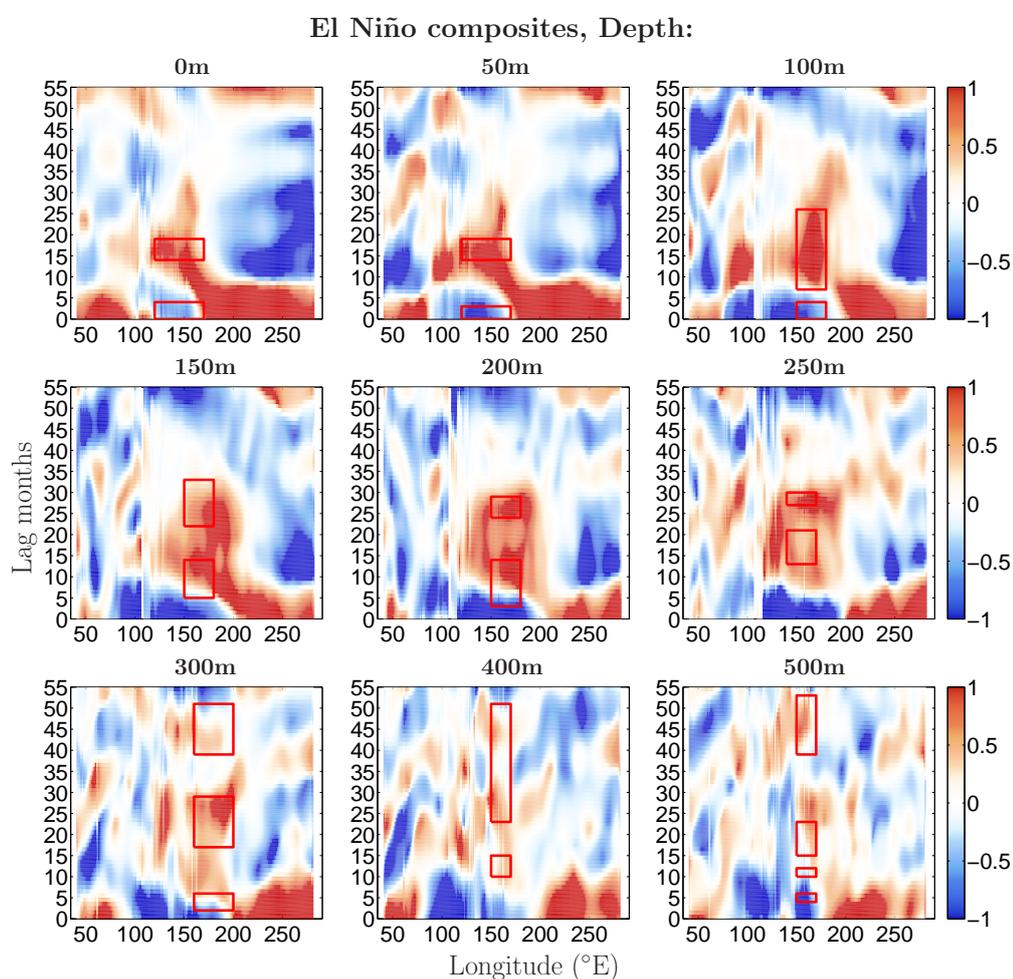
Figure 1.9: Hovmöller diagrams showing composites of interannual surface and subsurface temperature anomalies (in $[^\circ C]$, shading) from the Subsurface Temperature and Salinity Analyses by Ishii et al. (2005) between 0-500 metres depth. Red boxes indicate significant lag times of the precursors. Composites are with respect to all EN events in the period 1978-2012. Data is processed with a low-pass Butterworth filter, cut-off frequency 18, order 10.

with shifted and underscored mean annual cycle tend to simulate ENSO with shifts in its phase and with wrong amplitude, especially at the longer lead times (Jin et al., 2008). In the first simulation study with the model, Petrova et al. (2016), it became clear that seasonality was not sufficiently well reproduced, and there were some problems with the phase and amplitude of some of the ENSO events. Therefore, in the follow-up study, Petrova et al. (2017), this issue is addressed and seasonality is instead modelled with two independent time-varying cyclical components. Another cycle is also incorporated so that four cycle components are intended to capture the quasi-biannual, quasi-quadrennial and other dominant oscillatory modes in the tropical Pacific that have been shown to play major role in the formation of ENSO, such as its near-annual (An and Jin, 2004; An and Choi, 2009; Fashé and Dewitte, 2013),and decadal modulations, for example (Kleeman et al., 1999; Yeh and Kirtman, 2005), (Figure 1.8). Finally, since statistical models usually reduce the variance of the data they analyse (Barnett et al., 1993), an independent noise term is also included in the system. However, a portion of the variability appears to remain unaccounted for, as the spectra of forecasts and observations in Figure 1.8 reveal that the magnitude of variability is systematically underestimated also at the short lead time of 6 months.

Regardless, the shape of the variability of both the long- and short-term forecasts is in very good agreement with the observations and indicate a high level of forecasting performance. Although a rigorous direct intercomparison of this model, which constitutes a main development of the thesis, with the operational model forecasts is not possible, due to the fact that only experimental forecasts have been performed with the dynamic components model, still an imperfect parallel with other models is illustrative of the work presented here. In Table 1 are displayed *corr* with observations and *rmse* for the model and the same metrics for the best operational dynamical and statistical models as assessed by Barnston et al. (2012). It is worth highlighting that operational forecasts only extend to 8 months ahead, and longer lead times are currently not included in the forecasting plume (IRI, 2017). The values in Table 1 are for all seasons pulled together, and they suggest that the model performance is at least as good as that of the high-level dynamical models. However, this is a comparison only in very relative and not absolute terms. Our model is favoured by the fact that, although forecasts are made without any use of data from the predicted period, still they have not been issued in real time. At the same time, the operational models are also favoured by the shorter period over which they have been validated for the purposes of this study (Barnston et al., 2012). At the longer lead times at least up to 24 months *corr* values for our model plateau at about 0.6, while *rmse* values at about 0.8 (Figure 3 in Petrova et al. (2017), and predictions at this level are skilful (Barnston et al., 2012). Still, we remark that this is not the optimal skill of the dynamic components model as a fixed period is used for the calibration and forecasting in this case, while the real predictive power is reflected in Figure 1 of Petrova et al. (2017), where the *corr* at 24 months lead is 0.81 and components are continuously re-estimated whenever new information is available. For a comparison at these longer leads, Chen et al. (2004) reports *corr* values ∼ 0.6 and *rmse* ∼ 0.7 with the Lamont-Doherty Earth Observatory (LDEO) dynamical model for lead months up to

12, while Clarke and van Gorder (2003) reports *corr* ~ 0.6 for lead months up to 10 with the Florida State University (FSU) regression model.

Table 1.1: Correlation and root mean square error between observed and predicted values of the N34 Index as functions of lead time in months. Values are given for the dynamic components model (DCM) as in Figure 3 in Petrova et al. (2017) and the best dynamical and statistical models as displayed in Figure 6 and Figure 9 in Barnston et al. (2012).

	corr			*rmse*		
Lead time	**DCM** (1994-2015)	**Best dyn.** (2002-2011)	**Best stat.** (2002-2011)	**DCM** (1994-2015)	**Best dyn.** (2002-2011)	**Best stat.** (2002-2011)
1	0.97	~ 0.95	~ 0.90	0.26	~ 0.40	~ 0.55
2	0.92	~ 0.92	~ 0.80	0.41	~ 0.50	~ 0.63
3	0.87	~ 0.90	~ 0.75	0.53	~ 0.60	~ 0.71
4	0.80	~ 0.80	~ 0.68	0.61	~ 0.82	~ 0.76
5	0.74	~ 0.68	~ 0.60	0.69	~ 0.86	~ 0.80
6	0.70	~ 0.69	~ 0.50	0.73	~ 0.90	~ 0.85
7	0.67	~ 0.70	~ 0.40	0.77	~ 0.83	~ 0.90
8	0.63	~ 0.71	~ 0.30	0.80	~ 0.87	~ 0.93

1.5 ENSO Forecasts and Climate Services

Seasonal climate forecasts, in that account ENSO forecasts, are increasingly important in human life as climate at these timescales affects the planning of a great many of our activities. For example, it has a very high impact in the area of agriculture and food and water security, clean air, energy, and especially the renewable energy sector, transport, leisure, and human health (Dutton, 2002). Ultimately, it is a determining factor for our very existence. Fortunately, advancements in technology now provide opportunities for observations, modelling and computing capabilities that allow more and more credible predictions of atmosphere-ocean events. This increase of scientific skill enables a more quantitative and rigorous approach to the assessment of climate risk, and thus drives the evolution of the field of climate services where climate data is translated into specific tools and products in help of the public and decision makers. Climate services are in fact becoming more spread-out and diversified, not only as a result of the rapid development of technology, but also due to the increasing demand for such services in view of climate change. At the same time, in order to ensure the success of these services, there should be cooperation efforts on the part of the scientific, public and private sectors. For example, the health community encounters difficulties when struggling to access and interpret climate science information (Ballester et al., 2016c). Similarly, the climate community does not fully understand the needs of public health professionals and the authorities. Therefore, multi-disciplinary research should be promoted, with scientists from different fields collaborating to come up with the best possible climate services to answer this demand and achieve progress in this sphere. Key to this goal is the provision of accurate and relevant climate information, and the role of scientists is to provide decision makers or people at risk with reliable probabilistic climate information that would allow the management of resources and opportunities (Dutton, 2002).

The Global Framework for Climate Services (GFCS) is a mechanism for coordinated action that has set out to ensure smoother communication between the different professional areas, and to promote the development of trustworthy and sustainable climate services and tools at different timescales - from months to decades. One of the goals of GFCS is to extend the lead times for prevention and/or disaster-preparedness in cases of extreme events, disease outbreaks and other emergencies (GFCS, 2017). Climate information is provided by national and international scientific centres and institutions, and constitutes datasets of climatic variables such as temperature, precipitation, wind, pressure, soil moisture, SST, etc., along with risk-assessment maps and future projections and forecasts. Within the Framework, the main purpose of this information is a scientifically-supported and more effective decision-making process and better management of climate-related risk. One of the components of GFCS is *Research, Modelling and Prediction*, which aims to bring forward the area of climate simulation and forecasting. In order for the Framework to be properly implemented in this respect, more localized climate forecasting research and applications are required, and priority is given to developing countries that are highly vulnerable to climate variability (GFCS, 2017).

In this context, past, current and future information regarding ENSO events can be highly beneficial. Past information, for example, could provide users with the idea of what are the possible climatic conditions in the tropical Pacific during any given year, i.e. what is the climate variability, and what is the potential duration of anomalous conditions based on previous occurrences. This knowledge could be applied to prepare for any possible situation - EN, LN or normal conditions - or to adapt to the most probable and expected conditions, which is more efficient in terms of resource planning (Sarachik and Cane, 2010). The most common users of this kind of information are the nations in the vicinity of the equatorial Pacific, since the direct effect of ENSO events there is well-known and fairly consistent (Rasmusson and Carpenter, 1982). In addition, records of the past are also indicative of decadal and longer timescale variability, which could have implications about the frequency of cold and warm phases of ENSO, as already discussed. Finally, the track of past ENSO events allows for the construction of forecast systems, initialization techniques and validation of predictions to be made. The real-time information about the situation in the tropical Pacific provided by the monitoring system, on the other hand, raises the awareness of scientists as they know that a cold or warm event is underway several months in advance even when no explicit forecasting system is in operation. A typical example is the progression of the 1997/98 EN, which was followed in detail for almost a year before its peak (McPhaden et al., 1998). The tremendous benefits of infrastructure such as the observational array network, however, becomes immediately obvious when considering the kind of climate services and resource preservation that could arise from knowing the SST in the EPAC a couple of years in advance. The implications are for agricultural planning and insurance, water-management, fisheries, health, food aid, etc. Since the teleconnections of ENSO are now fairly well understood (Rasmusson and Carpenter, 1982; Ropelewski and Halpert, 1987; Kiladis and Diaz, 1989; Rodó et al., 2006), it is expected that the information from ENSO predictions would provide insight into the temperature and precipitation patterns in a number of regions throughout the globe. This emphasizes the role of ENSO in ensemble seasonal forecasting (Stockdale et al., 2011). An inherent problem of this kind of forecasts, however, is the fact that they are probabilistic (Goddard et al., 2001), which means that users should be aware of how to use the uncertain information in an educated and efficient way. Probabilistic seasonal climate forecasts could also be combined with risk management models and strategies that could then be used to avoid any unnecessary costs (Dutton, 2002). Therefore, although some of the negative direct effects of extreme ENSO events cannot be avoided, measures could still be taken in a timely manner in order to reduce drastically the main impacts on human life, as well as to avoid heavy financial losses.

The most cost-effective application of ENSO forecasts is enabled by making them relevant to the users on a regional level and targeting them at their local needs (Sarachik and Cane, 2010). One way to do this is to use the ENSO information within regional and local forecasting models of various kinds, such as infectious disease models for example. The last two studies included in this dissertation discuss this type of applications (Appendix 1 and 2). A connection has previously been shown between ENSO and interannual variability in dengue transmission in

southern coastal Ecuador through its influence on local precipitation and temperature rates. In fact, Ecuador is one of the countries in the tropics most heavily affected by ENSO (Bendix and Lauer, 1992; Moran-Tejeda et al., 2016). The first of the pair of studies explores the idea of applying long-lead forecasts of equatorial Pacific SST in the N34 within a dengue prediction model to evaluate if epidemics in the province of El Oro in Ecuador could be predicted well in advance using climate information. The ENSO forecasting model described earlier in the dissertation is used to predict past EN events about 2.5 years ahead, which provides a lead time of 2 years for the dengue model. In this way, it is shown that long-lead El Niño information could theoretically enhance long-lead predictions of dengue incidence in Ecuador and other tropical and subtropical countries. This could dramatically improve resource planning initiatives in such countries and remove some of the financial burden that they experience during dengue epidemics.

The second study, Lowe et al. (2017) (Appendix 2), attempts an operational forecast of dengue incidence in southern coastal Ecuador in 2016 using real-time seasonal climate forecasts in the region along with forecasts of El Niño produced with the prediction model discussed in the thesis. The resulting forecast of dengue incidence in El Oro is issued up to 10 months in advance, which is a relevant time scale for the decision-making process by public health officials in the country. The real-time forecasts are assessed retrospectively at the end of 2016 when the observational data became available. The predictions with climate forecast data are more accurate than those relying simply on benchmark risk thresholds (Lowe et al., 2017), and correctly simulate the timing of the peak of dengue in March 2016. The amplitude is also well-predicted in relative terms, indicating a very high chance of exceeding the mean of dengue incidence for the last 5 years.

The two studies represent important steps towards the design of an operational climate service for the health sector in Ecuador and the introduction of an early warning system for dengue epidemics in the region (Figure 1.10).
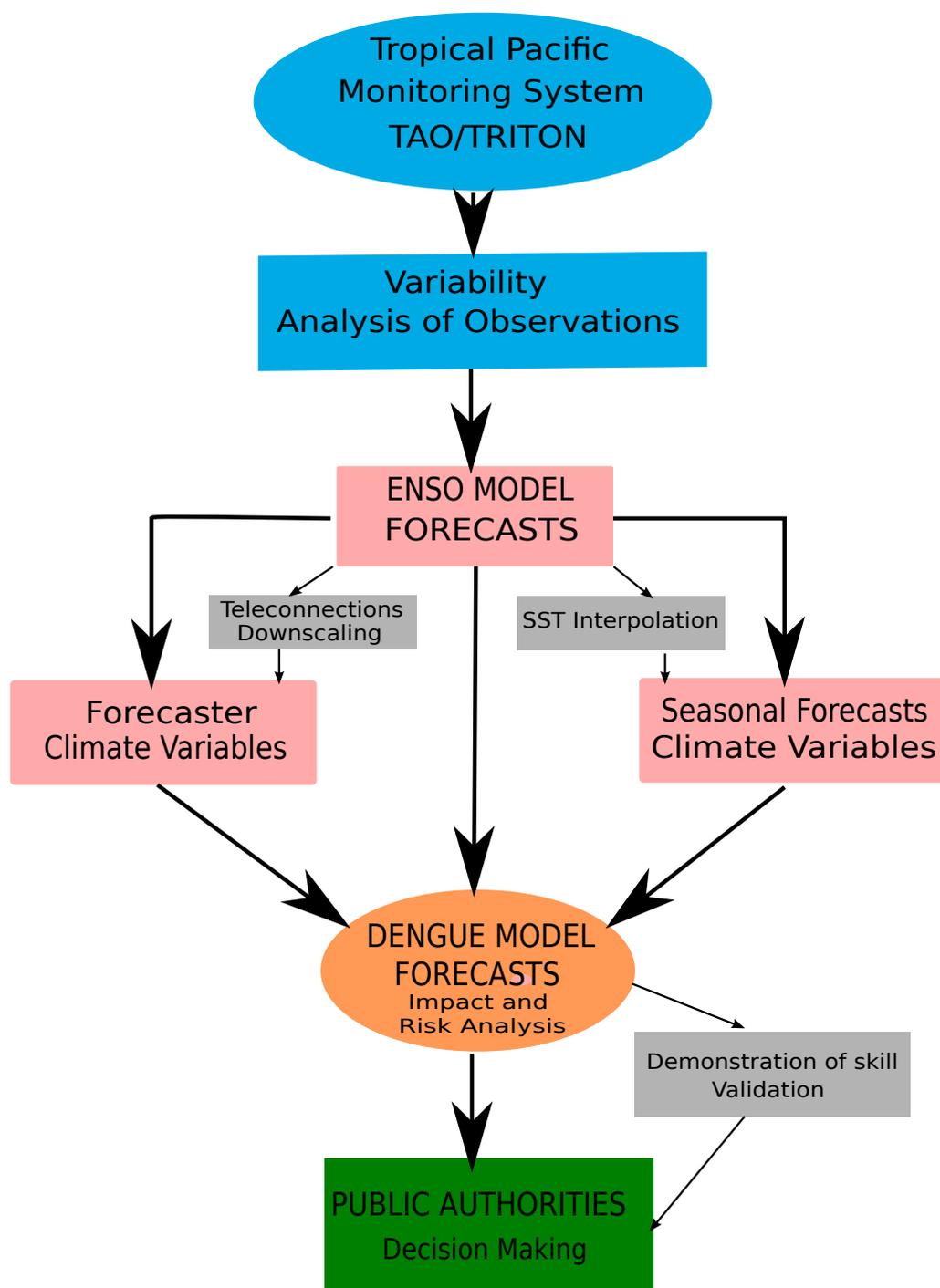
Figure 1.10: Schematic of a climate service tool for the health sector in El Oro, Ecuador. Shown is an international collaboration between climate and health scientists, as well as local public health authorities.

Chapter 2

Discussion and Conclusions

The studies included in this dissertation are dedicated to the advancement of the theoretical understanding and knowledge about the ENSO phenomenon. In particular, they are focused on unravelling the dynamical mechanisms leading to the accumulation and storage of heat in the WPAC ocean subsurface, as well as the propagation of this heat towards the CPAC and the growth of ENSO events in the EPAC. The implications of these subsurface and surface equatorial processes for the long-lead predictability of ENSO, as well as for the intensity and frequency of the events are also investigated. A new ENSO prediction framework is developed as a result, which incorporates these physical processes through a new dynamic components model and the addition of specially designed precursors. The forecast scheme is shown to be successful at very long-lead times of more than two years ahead of the winter peaks of the warm events. As a final chapter in this thesis, EN forecasts are implemented within a dengue prediction model to enhance the forecasts of dengue epidemics in southern coastal Ecuador. Discussed is, therefore, a prospect of a climate service tool (an early warning system) for the health sector.

In the first two studies the role of the WPAC as a fundamental area for the memory, origin and oscillatory nature of ENSO is analysed in detail. As the main equatorial region of deep atmospheric convection, and thus a major source of atmospheric teleconnections, it is not only a key domain for the local equatorial physical processes, but also has a large impact on the climate worldwide. As a primary development in this dissertation, it is shown that the onset of EN events is associated with strong trades and upper ocean warming first in the far western equatorial Pacific and some months after in the CPAC, and that this warming starts long before the actual events - at least 30 months earlier - with meridional and vertical convergence of mass in the subsurface. The validity and consistency of these results is confirmed by means of a number of data assimilation products. Moreover, it is demonstrated that the entrainment processes lead to slightly different intensity and exact areas of warming at different depth levels in the ocean, and that the warming is not restricted within the upper layer above the thermocline, but anomalies are present also at greater depths at least down to 500 metres. On the other hand, an

overall basinwide increase of HC does not occur until much later in the development of the events, i.e. about 9 months before the EN peak. Hence, the consideration of isolated heat patches at different depths as separate precursors of the events is considered for the first time in the second study of the dissertation, and this methodology is proved to strikingly increase the predictability of ENSO in a statistical dynamic components model. Therein, successful predictions are made more than two years in advance. Therefore, it is proved that zonally integrated HC traditionally used in statistical forecasting schemes is not an optimal precursor and could instead be limiting the real potential of such systems.

The third study further analyses the complex spatial and temporal structure of the heat buildup in the CPAC and its propagation eastward during the recharge and mature phases of EN. It is clarified that the warming tendency in the CPAC mainly results from the intensification of the EUC, which advects the original subsurface warming from the WPAC eastward along the thermocline. This process normally starts very early in the generation of the EN events, only a couple of months after the heat buildup begins in the WPAC (in the range between 25-36 months in advance). The peak of warm subsurface anomalies is identified at about 21 months ahead. In agreement with these findings, in Petrova et al. (2016) early precursors of the events in the subsurface ocean are defined in two general regions that span the WPAC and CPAC (Figure 2.1a). Since these predictors are oceanic features, they are conventionally more persistent and their time series are impacted to a lesser degree by the unreliable and shorter timescales of atmospheric noise. Several months later (between 6 and 15 months ahead), the EUC decreases in intensity in the WPAC and increases in the EPAC due to the subsidence of the trades at the surface and the occurrence of westerly winds. These westerly anomalies also have substantial impact on the timing and intensity of the warm events (Levine and McPhaden, 2016), (Eisenman et al., 2005), and are therefore also used as precursors at medium to short lead times (Figure 2.1b). Thus, the EUC advects the warm waters to the EPAC and they start to be upwelled to the surface and to warm the SST there. During the peak of the recharge stage zonal convergence and downwelling in the WPAC is decreased due to a decrease of the SEC and the surface westerly wind anomalies, which leads to consistent cold anomalies in the WPAC subsurface that are used as precursors at lead times shorter than 10 months. Another consistent cold anomaly in the northern off-equatorial region is found to anticipate EN by about 6-9 months in the second study of this dissertation and thus it is also defined as a precursor in the dynamic components model. In the third study it is shown that this anomaly in fact results from an inter-hemispheric asymmetry in the ocean vertical velocity, with upwelling in the Northern and downwelling in the Southern Hemisphere (Ballester et al., 2016b). Finally, the RossBell dipole defined in the extratropics in the Southern Ocean (Ballester et al., 2011) is also confirmed as a statistically significant precursor within the dynamics component model framework at lead times of 8-9 months, which is compatible with the previous results of Ballester et al. (2011). Its role as a tracer for EN events is important, considering that it is not directly affected by the ocean-atmosphere coupling in the equatorial Pacific. Preliminary results included here (see Appendix 3) point out that the RossBell dipole is generated by disturbances in the atmosphere

a)   Lag-21

Atmosphere

Ocean

100°E 120°E 140°E 160°E 180°E 200°E 220°E 240°E 260°E 280°E

b)   Lag-9

Atmosphere

Ocean
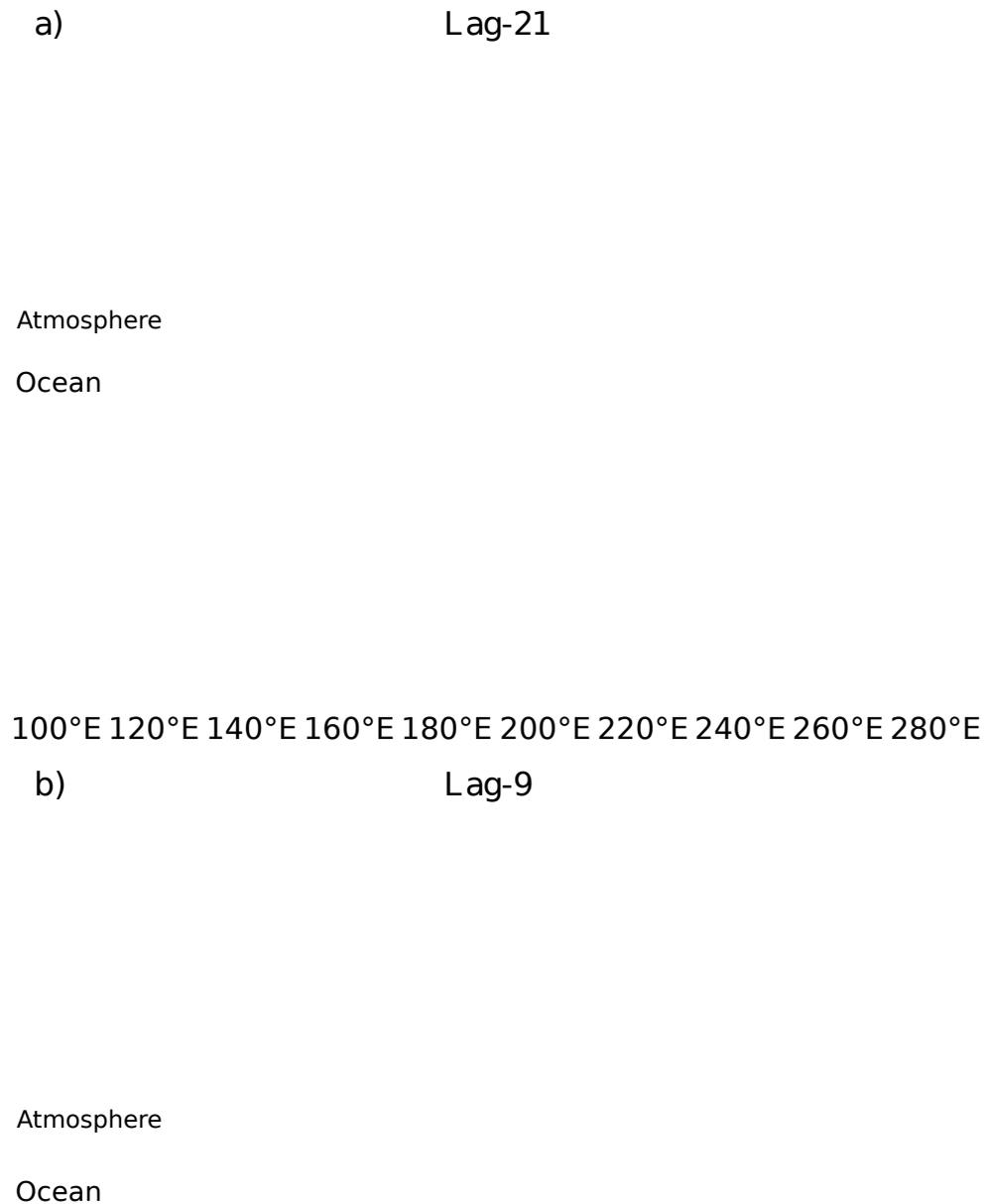
Figure 2.1: Schematic diagrams showing the processes and precursor regions in the tropical Pacific at lag- **a)** 21 and **b)** 9 months with respect to the peak of El Niño conditions, and corresponding to the early subsurface warming in the WPAC and CPAC, and the peak of subsurface heat content preceding EN.

in the extratropics. To study this effect better a similar atmospheric RossBell dipole is defined, which is shown to be partially driven by the ENSO forcing in the tropics and the annual cycle of the winds in the southern Pacific region. Further research is needed to consolidate these findings.

All these comprehensive results in the first three studies of the dissertation are not only valuable for the definition of ENSO precursors to be used in long-lead statistical forecasting, but could also be applied in the validation and improvement of intermediate and complex dynamical models that still fail to simulate some of the main features of ENSO (Barnston et al., 2012). By incorporating precursory covariates designed as a result of this analysis of coherent sequence of patterns starting in the WPAC, gradually moving towards the CPAC and then eastward to the EPAC, the forecasting scheme developed in the second study is shown to successfully predict the main EN events in the period 1996-2015 at least 29 months in advance. Up to our knowledge: **1.) forecasts at this long lead time have not been documented before, and precursory signals have neither been traced in specific subsurface regions, nor below the thermocline; 2.) moreover, the model predictive skill measured by standard metrics indicates that it outperforms other statistical models, and easily compares to some of the best dynamical ones; 3.) results there also demonstrate that the use of suitable subsurface information leads to the overcoming of the spring predictability barrier, 4.) while the use of specific regions as well as information from the surface, i.e. SST and wind stress, leads to the overcoming of the winter barrier associated with the predictability of equatorial HC; 5.) the simplicity of the model from a computational point of view renders yet another advantage to it.** Its flexibility provided by the state space approach and the re-estimation of component parameters with every piece of new information also ensures that the idiosyncrasies of individual events could be captured, especially by the time/frequency-varying cyclical components, and this usually represents a problem for statistical techniques. For example, the 2014/15 EN did not develop as a strong event partly because the atmosphere did not respond accordingly to the ocean warming and there was a strong easterly burst in the beginning of the summer in 2014 (Hu and Fedorov, 2000; Levine and McPhaden, 2016), while the usual westerlies preceding EN failed to appear afterwards. This is a typical example when atmospheric noise forcing significantly affects the amplitude of the event. Still, the dynamic components model correctly predicted this marginal EN event, since it incorporates the information from the atmosphere in the form of wind stress predictors. Because all components are estimated together in a dynamic way, specific precursors contribute to the accurate estimation of these varying frequencies of the cycle components in the case of every individual event.

2.1 Superimposed Frequencies of the Cyclical Components of ENSO

In a theoretical experiment, we could extend idealized time-invariant versions of the main frequency modes involved in the ENSO physics, as estimated with the dynamic components model (Figure 2.2). It can be observed that a number of the ENSO events coincide with simultaneous peaks of these idealized cyclical modes. In essence, this is a Fourier superposition of several waves of a particular phase relation, which results into focusing of the energy of the different modes, and thus into a sharp positive or negative peak through constructive interference. Such superposition is observed in this idealized framework for example in 1986/87, 1997/98, 2002/03, 2009/10 (Figure 2.2a,b), which are all officially recognized as EN years (CPC, 2017), and similarly for LN years in 1985/86, 1988/89, 1995/96, 1998/01, 2010/12. When the idealized decadal mode is also included (Figure 2.2b), peaks associated with the overlapping of energy modes coincide with some of the other ENSO events. For instance, simultaneous peaks of the 2-year cycle and the 4-year cycle are expected to peak in 1982/83. At the same time the decadal cycle peaks in 1982, which is the year when a prominent EN event actually occurred (Figure 2.2b). In addition, in 1991/92 there is a coincidental peak of the 5-year cycle and the decadal cycle, and this is also an official EN event. The decadal cycle remains in its peak phase until 1995, while the 4-year cycle peaks in 1994/95 when the 2-year cycle is at its rising phase as well and a weak El Niño is also documented. Notably, this particular positive phase of the decadal cycle corresponds to prolonged warm conditions in the EPAC between 1990 and 1995. A similar situation occurs during the next positive phase of the decadal cycle between 2002-2007. In the beginning of 2005 the decadal peak is superimposed onto a peak of the 2-year cycle and a growing phase of the 4-year cycle, and 2004/05 is defined as a weak EN event. Then in 2006/07 the 2-year cycle is again on the rise together with the 5-year cycle, while the decadal is still in its positive phase, and this is another EN event in the official records (CPC, 2017). The next decadal cycle positive phase starts in 2015 and is expected to peak in 2017. A similar decadal mode and entrance into a warm phase in 2015 is also found in Ramesh et al. (2016).

It is evident that these superposition inter-dependencies do not have a linear effect on the amplitude of the respective events, which evolve in a non-linear way. A super-event might occur when all of the discussed cycles peak together at the same time, as it is in the case of the 1997/98 EN event. Furthermore, there are also phase shifts of the main cyclical modes as a result of the modulation of the decadal cycle as for the 2015/16 EN when the frequency of the 5-year cycle is locked to that of the decadal cycle and its peak is delayed. In fact, it is the role of the predictor covariates to signal such phase and amplitude shifts and account for the variance in these idealized cyclical components through the dynamic estimation of parameters based on new observations. These features place the unique statistical modelling technique described here on the same plane with dynamical schemes, since the system could update itself as a result of changes in the physical process that describe the state of the atmosphere-ocean coupled system.
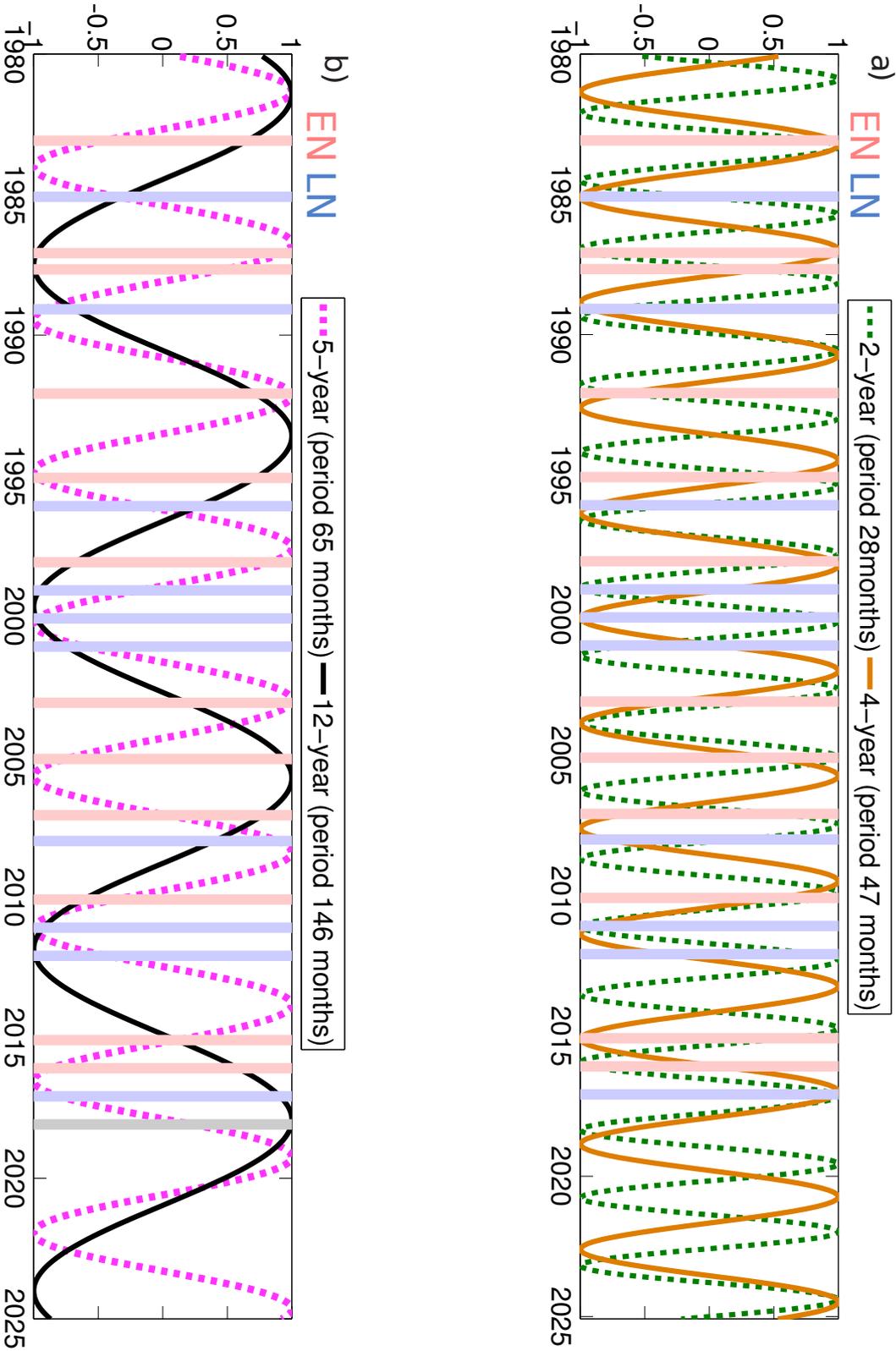
Figure 2.2: Schematic of idealized cycle modes involved in the ENSO dynamics. Shown are cycles of approximate periods of **a)** 2 (dashed green) and 4 (solid beige) years, **b)** 5 (dashed magenta) and 12 (solid black) years. Peak months of El Niño (pink lines) and La Niña (light blue lines) events from the observational record are indicated. Wave amplitudes are not realistic.

In light of the key modulation of cyclical components that has been found, in Petrova et al. (2017) the model is upgraded with an explicit additional decadal cycle component together with time-varying seasonal and annual cycles. Essentially, this study addresses and solves the majority of problems associated with the model that arose and were summarized as future work in Petrova et al. (2016). As expected, these improvements have resulted into further increase of the model skill and better performance metrics (see for example Figure 3 of Petrova et al. (2017)). In addition, the whole N34 time series from 1970 to 2016 is predicted 2 years ahead. Previous studies have only reported forecasts of individual EN events at such a long lead time (Chen et al., 2004; Gonzalez and Goddard, 2016), but not of neutral or LN years, which are normally harder to predict. This testifies that the model is not prone to produce false alarms. A major discovery is the dramatic improvement of overall skill after 1994, which is the year when the observing system in the tropical Pacific was fully deployed. The change is outstanding and cannot be simply explained by the variability in ENSO predictability itself. Moreover, it unequivocally corresponds to the period of the continuous positioning of the array network. Since the model depends heavily on subsurface data from the equatorial Pacific Ocean, the provision of such high-quality three dimensional data is crucial for its success. As seen in the paper, its availability has direct implications about the impact and significance that the precursors have on the predictions, and the model performance improves gradually with the evolution of the TAO-TRITON observational array and with the enrichment of the data records describing the ocean subsurface. Improvements in the regularity of surface data streaming also undoubtedly contribute to the boosted performance of the model. For the first time it is demonstrated that not only dynamical prediction systems (Stockdale et al., 2011), but also statistical models can benefit greatly from the constant and more accurate subsurface and surface data in the tropics for the correct modelling and prediction of ENSO. Therefore, this study has the potential to become a reference of paramount importance for the long-lead ENSO prediction and predictability, having implications for the overall provision of accurate seasonal-to-interannual climate forecasts. Furthermore, it strongly reinforces the complementary role of statistical ENSO forecasting. A second prominent finding of this study is that there is only a minor decrease in the skill of the model for forecasts issued during and beyond the preceding spring season, as well as the one in the year before the event, which supports the hypothesis proposed by previous work that this is not an inherent limit for ENSO predictability, but is rather a constraint conditioned upon the design and specificities of the forecasting schemes.

By and large, the results of the study not only push forward the theoretical potential for long-lead forecasting of the phenomenon, but also provide key evidence for the capacity of statistical long-term prediction that has so far been largely underestimated by the ENSO scientific community. In particular it is proven that long-lead statistical predictions are achievable when state-of-the-art models are adapted to incorporate the subsurface information in a more selective and less integrative way. These types of forecasts are also much faster to run computationally than those produced with complex physical models. Furthermore, the results highlight the huge

benefits of extensive climate monitoring systems, and the view that they make possible major advancements in both theoretical and practical climate science some decades later is strongly supported here.

In addition to isolating the different mechanisms involved in the buildup and propagation of heat prior to EN, the dissertation also explores the sensitivity of the event timing and intensity to the actual intensity of the heat buildup. In the fourth study, a nonlinear dependency between the two is found for the first time at the very long lead time of 21 months - the peak of warm subsurface anomalies preceding EN in the WPAC and CPAC as described earlier. Through numerical experiments with a GCM it is verified that a large increase of HC there at this time (by $\geq 80\%$) leads to warm SST anomalies over the whole equatorial basin in the winter one year earlier, which is also a pre-condition for the occurrence of more energetic deterministic westerly wind anomalies (Gebbie and Tziperman, 2009), and thus an extreme EN event the following winter. This scenario is in fact similar to the 2014/15 very weak EN when warm SST anomalies were present basinwide, and it was followed by the very strong 2015/16 EN. In the experiments, the strong EN is also followed by a strong LN the year after, unlike the weak LN that happened in 2016/2017. The explanation for this might be related to a modulation by the positive peak phase of the decadal mode, or perhaps to climate change, but this is beyond the scope of this research.

On the other hand, a substantial decrease in the HC at this time (by $\geq 80\%$) sets the initiation of a new recharge of heat in the WPAC, a weak LN in the winter before the event in the EPAC, and then a delay by one year of the EN in agreement with the seasonal phase-locking of ENSO. Furthermore, a decrease/increase of the HC by 40% does not result into changes in the timing of the event, but only into a decrease/increase of the EN amplitude by about 30%. It is also interesting that although in general the larger is the initial HC, the larger is the amplitude of the following warm event, this relationship is again not linear as an increase by 100% of the HC in the experiments results into an amplitude increase of only about 1 °C. This study as well as the previous ones substantiates the theory for self-sustainability and long-lead predictability of ENSO of more than one year, since it demonstrates that apart from the established association between tropical ocean HC and EN about 9 months before, such a significant association also holds 21 months before (Ballester et al., 2016b).

On average, the physical mechanisms described here are reverse for the early generation and development of the cold phase of the oscillation. However, some important asymmetries exist, especially in regards to the shorter time frame needed for the growing of LN that immediately follows a strong EN. In addition, some EN episodes do not seem to be directly preceded by LN or even LN-like conditions. Finally, by convention LN anomalies in the EPAC are predominantly of smaller amplitude than the warm events. These asymmetries will be studied in more detail in the future, and specific predictor variables will be designed with the aim of improving the LN forecasts. Some of the possible reasons for these differences might be related to the dynamic heating of the ocean or to the interdecadal modulation of ENSO. A question open to research is the actual origin of the ENSO fluctuation on decadal timescales. There are a few competing

theories: from mid-latitudinal temperature anomalies propagating to the equatorial region along the thermocline through advection processes, to decadal wind stress anomalies within the tropics modulating the thermocline (Ramesh et al., 2016), or to nonlinear effects due to the fact that the tropical ocean-atmosphere system is chaotic (Timmermann and Jin, 2003). Considering that this low-frequency mode not only has an important role to play in ENSO dynamics, but also has a modulating effect on the global mean surface temperature (Meehl et al., 2011), it would be pertinent to study its physical mechanism further in the future. An interesting result is that this mode is reproduced with the Zebiak-Cane model, which has as a domain only the tropical Pacific (Ramesh et al., 2016). This implies that whatever the exact dynamical mechanism is, it originates within the tropical region itself. Another possible avenue for future research is a study that outlines the limits of predictability with the dynamic components model developed in the dissertation. A study with a decadal ENSO forecasting system has suggested a possible predictability window extending up to 4 years ahead, albeit with necessary trend and bias corrections (Gonzalez and Goddard, 2016). An interesting outcome in this study, which we would like to test with the dynamic components model, is that the strongest events are not always predicted better than the weaker ones. This assertion clearly runs against previous results (e.g. Chen et al. (2004)).

Knowledge about the SST in the equatorial Pacific a year or more in advance could be of huge benefit to many countries in the world. This information could be utilised in weather forecasting, as well as seasonal climate predictions, annual planning of agriculture and fishing activities, storm and flood preparedness, organization of public health response and vaccination campaigns for fighting climate-sensitive diseases, and overall resource and state budget planning. In particular, it appears of considerable interest to the developing countries that directly border the area of major changes associated with ENSO, e.g. Peru and Ecuador, whose climate is particularly affected by the interannual ENSO variability. Thus, the final two studies of the dissertation demonstrate the possibility for the development of a climate-driven early warning system for dengue outbreaks in El Oro province in Ecuador. In the first of the pair of studies the ENSO dynamic components forecasting model is used to predict in retrospect the timing and amplitude of the prominent 1997/98 and 2009/10 EN events 2.5 years in advance. The predicted N34 is then used to drive a dengue incidence model, which successfully forecasts the two biggest recorded outbreaks of dengue in the region in 1998 and 2010 at least 30 months ahead (Figure 2.3). The role of ENSO as a precursor for dengue derives from its direct impact on the climate in El Oro, in particular its predictive capacity for the increase of minimum temperature and precipitation following El Niño events. These two variables at the same time are key drivers of mosquito population dynamics. The identified mechanistic explanation of the inter-linkage can be explained as follows: the normal climatic patterns in southern coastal Ecuador in boreal winter/spring (i.e. seasonal maximum in temperature and precipitation) are enhanced due to El Niño, and as a result, the dengue incidence (also with a seasonal maximum in spring) tends to be enhanced during El Niño years. It is thus demonstrated that the prediction window for dengue in the area could theoretically be extended much further in time than traditionally considered.
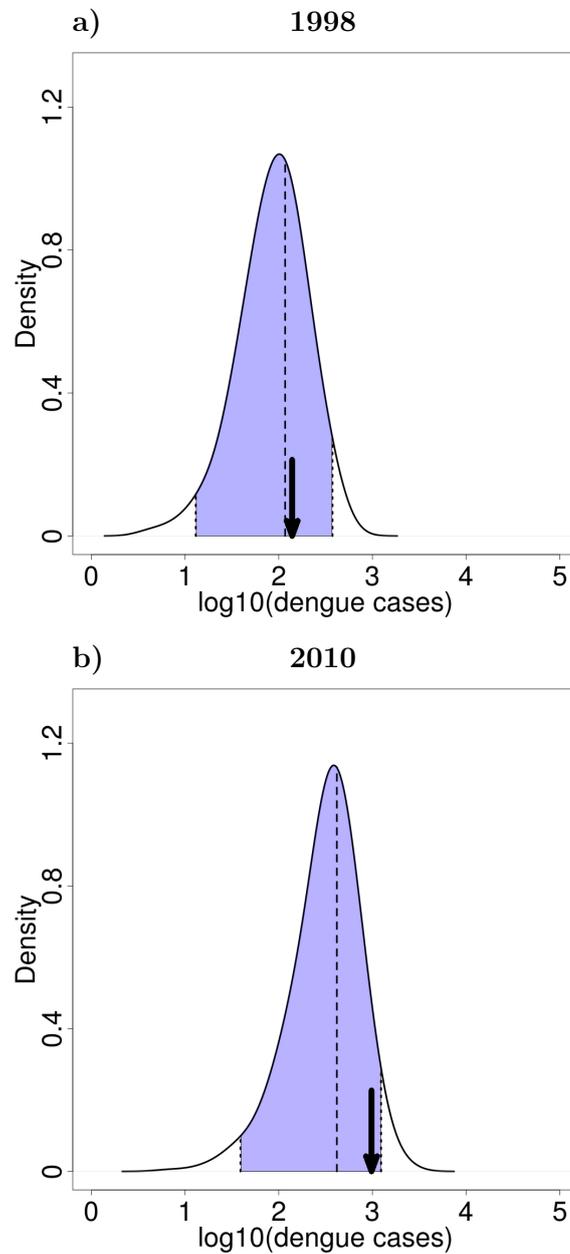
Figure 2.3: Posterior predictive distribution of dengue cases (base-10 logarithmic scale) for El Oro province, Ecuador for March **a)** 1998 and **b)** 2010 using forecast Oceanic Niño Index (3-month running mean N34) values for November-December-January at lead times of 30 months with respect to the peak in dengue. The posterior predictive mean (dashed line), 95% credible intervals of the predictive distribution (dotted lines) and observed dengue cases (arrow) are indicated.

While decision-makers customarily use forecast information for planning purposes only a few months before an epidemic occurs, the findings in this study show that it is possible to optimize public health budgets by planning a couple of years in advance. Moreover, the uncertainty related to the forecasts could be largely decreased by validating the predictions gradually and stepwise in real time, and thus refining the available information. This is emphasized by the fact that both the EN and dengue predictions are shown to converge to the observations uniformly with decreasing lead time. This is especially relevant when considering the limiting factors of this research, namely the admission that not all epidemics are driven by climate, that climate variables impact the disease in a nonlinear way, and that ENSO teleconnections are transient in nature and may change in the future with the progression of global warming.

The last study further explores the potential for an early warning system by proposing a tentative prototype (Lowe et al., 2017). N34 and seasonal climate forecasts are incorporated into an improved version of the dengue model, and dengue incidence in southern coastal Ecuador is predicted in real time for up to 10 months in the future. Dengue prediction studies usually use the observed climatic variables and are in this way restricted to only a few months lead time. Here a first step is made towards increasing the lead time in an operational framework with the inclusion of predicted climate precursors. Forecasts were issued on 1 January 2016 for all months until October 2016, and then validated at the end of 2016 when the epidemiological data became available. The predictions capture the peak of dengue incidence in March, and detect the very high chance of exceeding the mean incidence for the previous five years. They also correctly foresee the decrease below the mean of the number of cases during the second half of the year. Importantly, information from the ENSO and the seasonal climate forecasts not only extends the predictive capacity of the dengue model to longer lead times, but also improve the timing of peak incidence as compared to estimates based on the seasonal averages from the previous years, which foresee a peak in June rather than in March. As far as we are aware, this is the first study in which a dengue prediction utilising operational climate forecast information is performed in real time rather than in retrospective mode.

The pair of studies are in line with the recommendations of the World Health Organization to use climate forecasts to improve dengue outbreak predictions (Kuhn et al., 2005). Moreover, they are of direct usefulness for the public health sector in Ecuador in light of their potential to help authorities to optimise their available resources, especially in the planning and prevention phases when dealing with a dengue outbreak. The results further suggest that climate information should be included in the official surveillance activities of the country, given that even only awareness about a forthcoming EN event could improve their overall preparedness in regards to the timing and magnitude of an epidemic. Although policy makers in the tropics are very aware of the modulating effect of climate and ENSO in particular on the dynamics of diseases such as dengue, climate forecast information is not being exploited systematically yet in their prevention and mitigation efforts. Thus, the research work presented here implies that the collaboration between the meteorological and health services in Ecuador can now be extended further with a new climate service tool for health. Our goal for the future is to try to make it fully operational.

Future research could also identify the extent to which large-scale global teleconnections could be mapped applying the long-lead ENSO forecasts performed with the dynamic components model. Considering that ENSO is the most powerful generator of tropospheric interannual variability (Rasmusson and Carpenter, 1982; Ropelewski and Halpert, 1987), such an evaluation of the skill of SST forecasts in the EPAC to predict climate in various parts of the globe based on well-known physical mechanisms could be of tremendous value for seasonal forecasting systems, as well as for local climate monitoring and surveillance activities. As a further matter, more areas where ENSO forecasts could be incorporated into infectious disease or other types of risk models could be pinpointed as a result of such prospective scientific exploration with the aim of saving lives, ensuring better economic development and protecting the environment.

Chapter 3

Publications

Report of the director

The director of the doctoral thesis, doctor Xavier Rodó López, certifies the contribution of the candidate to the articles included in this dissertation, as well as their impact factor, as specified below.

Article 1

Ballester J., Bordoni S., Petrova D., Rodó X.

On the dynamical mechanisms explaining the western Pacific subsurface temperature buildup leading to ENSO events.

Geophysical Research Letters, doi: 10.1002/2015GL063701 (in press)

Impact factor 4.212 in 2015 (first quartile).

The applicant contributed with important analysis and discussions about the dynamical origin and evolution of ENSO, and helped to improve the manuscript style and structure. No other co-author has used or plans to use any of the results presented in this article for a PhD dissertation.

Article 2

Petrova D., Koopman S.J., Ballester J., Rodó X.

Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model.

Climate Dynamics, doi:10.1007/s00382-016-3139-y (2016).

Impact factor 4.708 in 2015 (first quartile).

The role of the applicant in this article is absolutely principal. She developed the model, ran the simulations, and performed all the analysis and diagnostics necessary for the results and conclusions of the study. The contributions of the other authors were on the guidance and supervision level. No other co-author has used or plans to use any of the results presented in this article for a PhD dissertation.

Article 3

Ballester J., Bordoni S., Petrova D., Rodó X.

Heat advection processes leading to El Niño events as depicted by an ensemble of ocean assimilation products.

Journal of Geophysical Research: Oceans, doi:10.1002/2016JC011718 (in press)

Impact factor 3.318 in 2015 (first quartile).

As a continuation of Article 1, this publication was also achieved with the important contributing role of the applicant in fundamental analysis and discussions about the dynamical evolution of ENSO and the accumulation of heat in the subsurface equatorial Pacific Ocean. She also helped to improve the manuscript style and structure. No other co-author has used or plans to use any of the results presented in this article for a PhD dissertation.

Article 4

Ballester J., Petrova D., Bordoni S., Cash B., García-Díez M., Rodó X.

Sensitivity of El Niño intensity and timing to preceding subsurface heat magnitude.

Scientific Reports, doi:10.1002/2016JC011718 (in press)

Impact factor 5.228 in 2015 (first quartile).

The applicant contributed significantly to this publication by discussing the design of the study and the results, and putting them in the context of the existing literature on ENSO predictability and recent El Niño events. No other co-author has used or plans to use any of the results presented in this article for a PhD dissertation.

Article 5

Petrova D., Ballester J., Koopman S.M., Rodó X.

Multi-year statistical prediction of ENSO enhanced by the tropical Pacific observing system.

Submitted for review in Proceedings of the National Academy of Sciences.

The applicant has the leading role in this publication as she was in charge of improving

the model, running all the experiment simulations, processing the forecasts and interpreting the results regarding the role of better observations for enhanced statistical forecasting. No other co-author has used or plans to use any of the results presented in this article for a PhD dissertation.

The director of the thesis,

Xavier Rodó López

3.1   On the dynamical mechanisms explaining the western Pacific subsurface temperature buildup leading to ENSO events

Ballester J (1), Bordoni S (1), Petrova D (2), Rodó X (2,3)

(1) California Institute of Technology, Pasadena, California, United States
(2) Institut Català de Ciències del Clima, Barcelona, Catalonia, Spain
(3) Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

Abstracto - *Summary in Spanish*

A pesar del progreso constante en el entendimiento de la naturaleza de El Niño-Oscilación Sur (ENSO) en el pasado durante décadas, siguen pendientes interrogantes sobre los mecanismos exactos que explican la acumulación de calor que conduce al El Niño (EN). Aquí utilizamos un conjunto de productos de asimilación de océanos y atmósferas para identificar mecanismos que son consistentemente identificados por todos los conjuntos de datos y que contribuyen a la acumulación de calor en el Pacífico occidental 18 a 24 meses antes de la aparición de eventos EN. Se encontró que la advección de calor meridional hacia el este, debido a la convergencia de la masa subsuperficial del ecuador y el transporte a lo largo de la corriente submarina ecuatorial, contribuye al calentamiento subsuperficial hacia los 170°E-150°W. En el *warm pool*, en cambio, la convergencia horizontal superficial y el movimiento descendente tienen un papel principal en el calentamiento subsuperficial. Los resultados obtenidos destacan una fuerte transición dinámica hacia los 170°E cerca del nivel de la termoclina.

Geophysical Research Letters

On the dynamical mechanisms explaining the western Pacific subsurface temperature buildup leading to ENSO events

Joan Ballester[1,2], Simona Bordoni[1], Desislava Petrova[2], and Xavier Rodó[2,3]

[1]Environmental Science and Engineering, California Institute of Technology, Pasadena, California, USA, [2]Institut Català de Ciències del Clima, Barcelona, Spain, [3]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

Abstract Despite steady progress in the understanding of El Niño–Southern Oscillation (ENSO) in the past decades, questions remain on the exact mechanisms explaining the heat buildup leading to the onset of El Niño (EN) events. Here we use an ensemble of ocean and atmosphere assimilation products to identify mechanisms that are consistently identified by all the data sets and that contribute to the heat buildup in the western Pacific 18 to 24 months before the onset of EN events. Meridional and eastward heat advection due to equatorward subsurface mass convergence and transport along the equatorial undercurrent are found to contribute to the subsurface warming at 170°E–150°W. In the warm pool, instead, surface horizontal convergence and downwelling motion have a leading role in subsurface warming. The picture emerging from our results highlights a sharp dynamical transition at 170°E near the level of the thermocline.

1. Introduction

El Niño–Southern Oscillation (ENSO) is the most energetic climate signal after the seasonal cycle, the major source of interannual variability worldwide and a dominant driver of climate teleconnections [*Ballester et al.,* 2013]. Its prominent amplitude in the tropical Pacific is essentially explained by the positive Bjerknes feedback [*Bjerknes,* 1969; *Wyrtki,* 1975], which involves a strong coupling between the Walker circulation, the zonal gradient of sea surface temperature and the longitudinal tilt of the thermocline [*Ballester et al.,* 2011]. Two main theories have been invoked to bound the amplitude and reverse the sign of interannual anomalies: the delayed oscillator theory, which explains this reversal through differential propagation speed of wind-induced oceanic Kelvin and Rossby waves [*Battisti,* 1988; *Schopf and Suarez,* 1988; *Wang,* 2002; *Fedorov and Brown,* 2009]; and the recharge oscillator, which emphasizes the time delay between anomalies in longitudinally averaged thermocline depth and eastern Pacific sea surface temperature [*Jin,* 1997a, 1997b; *Meinen and McPhaden,* 2000].

The western Pacific is a key region for understanding the oscillatory nature of ENSO and the generation of EN and La Niña (LN) events. It is characterized by the warm pool, an upper ocean area of very warm and well-mixed waters (supporting information Figures S1b and S2b), surface horizontal current convergence and subsurface divergence (supporting information Figure S3b) and therefore downwelling motion. This region represents the main energy source for deep atmospheric convection and basin-wide teleconnections [*Picaut et al.,* 1996], and as such, it modulates both the climate of the planet and the physical properties and dynamical processes occurring locally. The equatorial Pacific is characterized by a sharp upper ocean zonal salinity front near the warm pool edge around 160°E–180°E (supporting information Figure S1b), which results from the convergence of zonally advected low- (high-) salinity water masses from the western (central) Pacific (supporting information Figure S3b) [*Bosc et al.,* 2009]. The longitudinal contrast in surface salinity is in turn explained by both the intense thermally driven atmospheric convection and rainfall to the west (supporting information Figure S4b) and the strong trade winds and evaporation to the east (supporting information Figure S5b).

The interannual zonal migration of the Pacific warm pool is well known to be associated with the displacement of the areas of deep atmospheric convection, active precipitation, strong trade winds, and enhanced evaporation [*Singh et al.,* 2011]. These relationships explain the intimate link between ENSO variability and the zonal thermohaline structure of the upper equatorial Pacific, with increased (decreased) zonal contrast of temperature and salinity during LN (EN) events, when the area of maximum precipitation is enhanced in the western Pacific (is shifted to the central Pacific) and the trade winds and westward surface currents are strengthened

(relaxed) in the central Pacific (supporting information Figures S1, S2, S4, and S5). These associations determine the intensity and vertical extension of surface horizontal current convergence and downwelling motion in the warm pool, as well as the zonal position and vertical tilt of local isopycnals through simultaneous changes in both the warm pool edge and the salinity front (supporting information Figures S1 and S3).

2. Methods

Outgoing longwave radiation, latent heat flux and wind stress data are derived from the NCEP/NCAR reanalysis reanalysis [*Kalnay et al.,* 1996] and precipitation is from the Global Precipitation Climatology Project version 2.2 (GPCP v2.2) Combined Precipitation Data Set [*Adler et al.,* 2003]. Ocean variables are obtained from five assimilation products: GECCO, ORAS4, NEMOVAR-COMBINE, SODA2.1.6, and SODA2.2.6 [*Köhl and Stammer,* 2008; *Balmaseda et al.,* 2008, 2010, 2013; *Carton and Giese,* 2008]. Vertical velocity is diagnosed by integrating horizontal divergence down from the surface, with surface values assumed to be equal to the time tendency of sea surface height.

A 13 term running average (1/24, 1/12, …, 1/12, …, 1/12, 1/24) is used to calculate the interannual component of detrended monthly variables. EN events are chosen according to the classification of the Climate Prediction Center for the common period 1961–2001 of the assimilation data sets: December 1963, 1965, 1968, 1972, 1976, 1982, 1986, 1990, and 1997 [*Climate Prediction Center,* 2013]. In those cases in which EN conditions are observed in the tropical Pacific for two consecutive boreal winters (i.e.,1968/1969, 1976/1977, 1986/1987, and 1990/1991), only the first year is considered for the calculation of the composite anomalies, given that our main objective here is the description of the onset of these events. The 1994 event was excluded from the analyses because it was the continuation of a previous warm event starting in 1990, with warm sea surface temperature anomalies persisting in the central and eastern tropical Pacific for almost 6 years [*Trenberth and Hoar,* 1996]. A similar criterion is used for the selection of LN years: December 1964, 1970, 1973, 1975, 1984, 1988, 1995, and 1998.

The temperature tendency equation is analyzed with regard to the contribution of the zonal (U_{adv}), meridional (V_{adv}), and vertical (W_{adv}) heat advection components. These terms are expressed as

$$U_{adv}' = -\overline{u}\frac{\partial \theta'}{\partial x} - u'\frac{\partial \overline{\theta}}{\partial x} - u'\frac{\partial \theta'}{\partial x} + \overline{u'\frac{\partial \theta'}{\partial x}}, \tag{1}$$

$$V_{adv}' = -\overline{v}\frac{\partial \theta'}{\partial y} - v'\frac{\partial \overline{\theta}}{\partial y} - v'\frac{\partial \theta'}{\partial y} + \overline{v'\frac{\partial \theta'}{\partial y}}, \quad \text{and} \tag{2}$$

$$W_{adv}' = -\overline{w}\frac{\partial \theta'}{\partial z} - w'\frac{\partial \overline{\theta}}{\partial z} - w'\frac{\partial \theta'}{\partial z} + \overline{w'\frac{\partial \theta'}{\partial z}}, \tag{3}$$

where the overbar and the prime denote the decomposition of potential temperature (θ) and the zonal (u), meridional (v), and vertical (w) current velocities into their climatology and the nonclimatological anomalies, respectively.

3. Results

Figure 1 shows the longitude-depth composite of equatorial potential temperature and its time tendency before EN events, averaged for the ensemble of assimilation products. Note that the stippling denotes areas where anomalies are consistent among all assimilation products. The initial accumulation of subsurface warm waters in the western Pacific (Figure 1a) and the subsequent eastward movement along the equatorial thermocline (Figures 1b–1d) characterize the onset of EN events [*Jin,* 1997a, 1997b]. This propagation links opposite phases of the tilting mode at lags −21 and +00 (Figures 1a and 1d) with a transition period of increased zonally integrated equatorial subsurface heat content at lag −09 [*Meinen and McPhaden,* 2000] (Figure 1c). After reaching the upper eastern equatorial Pacific, the released heat is rapidly amplified by the Bjerknes feedback, leading to the fast growth of an EN event [*Ramesh and Murtugudde,* 2013].

The present study describes the ocean-atmosphere mechanisms explaining this initial subsurface warming in the western equatorial Pacific, with special emphasis on the spatial characterization of the
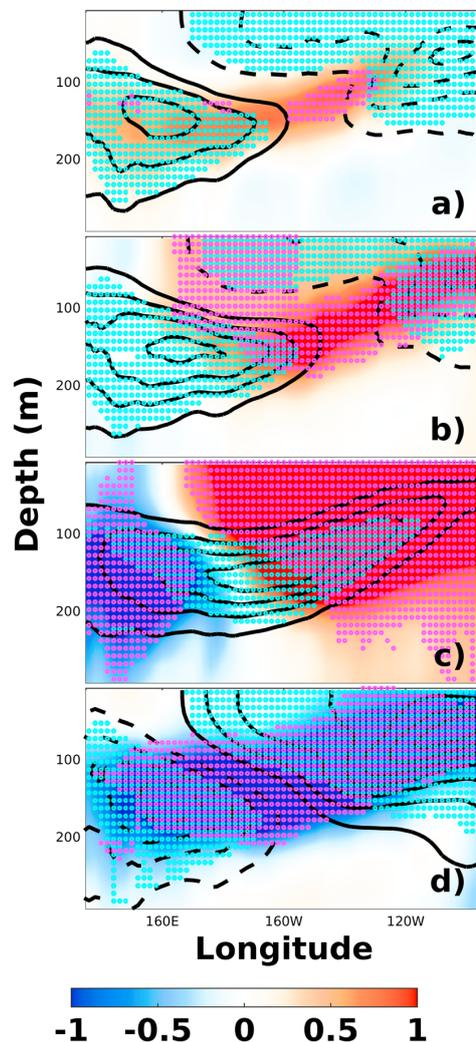
Figure 1. Evolution of equatorial temperature and its tendency during the onset of El Niño events. Multiproduct average of potential temperature (°C, contours) and its time tendency (°C/yr, shading) averaged over the equatorial band (2°S–2°N), and shown for lags 21 (a), 15 (b), 09 (c), 00, and (d) months before the major El Niño events. The temperature contour interval is 0.25°C, with solid (dashed) lines depicting positive (negative) anomalies. The purple (cyan) stippling denotes areas where shading (contour) anomalies have the same sign and magnitude larger than ±0.025°C/month (±0.25°C) for all members of the ensemble.

relative role that different processes play in this heat buildup. To this aim, Figure 2 shows the longitude-depth composites of ocean temperature, salinity, zonal velocity, and zonal, horizontal, and meridional divergence of ocean currents in the equatorial Pacific at 21 months before the mature phase of EN, while Figure 3 depicts the contribution of zonal, meridional, and vertical heat advection to the temperature tendency leading to the anomalies occurring at this lag.

On average, the peak of interannual LN-like conditions is found in the upper equatorial ocean at lag −21, with colder-than-normal waters in the central and eastern Pacific (contours in Figure 2). The subsequent shift of convection and precipitation to the far western Pacific increases the surface salinity in the easternmost edge of the climatological warm pool near 160°E–180°E (shading in Figure 2a), sharpening the zonal differences in surface salinity near the warm pool edge and the salinity front. Changes in both temperature and salinity increase the zonal contrast in upper ocean density and the local tilt of the isopycnals. At the same time, the enhanced trade winds in the central Pacific strengthen the westward south equatorial current near the warm pool edge and the salinity front (Figure 2b). This results in anomalous surface zonal (0–60 m) and horizontal (0–75 m) convergence in the western equatorial Pacific (shading in Figures 2c and 2d). The surface buoyancy loss, additional isopycnal tilting and enhanced surface westward circulation and

horizontal current convergence result into the intensification of the downwelling motion, which contributes to the growth of the subsurface warm buildup.

The strengthening of the trades during LN-like conditions is also associated with the enhancement of the clockwise (counterclockwise) wind stress curl in the central off-equatorial north (south) Pacific, which together with the rise in dynamic height due to the accumulation of warm waters in the western Pacific, drives anomalous geostrophic Sverdrup transport toward the equator (shading in Figure 2e) [*Wyrtki*, 1981]. This subsurface-intensified meridional convergence is particularly strong in the western and central Pacific, where the strongest interannual ENSO-like zonal wind stress anomalies are found (cf. with supporting

Figure 2. Multiproduct average of equatorial variables during the heat buildup leading to El Niño events. Potential temperature (°C, contours), (a) salinity (g/kg), (b) zonal velocity (m/s), and (c) zonal, (d) horizontal, and (e) meridional current divergence (10^{-6} 1/s). Composite anomalies are averaged over the equatorial band (2°S–2°N), and shown for lag 21 months before the major El Niño events. The temperature contour interval is 0.25°C, with solid (dashed) lines depicting positive (negative) anomalies. The purple (cyan) stippling denotes areas where shading (contour) anomalies have the same sign for all the members of the ensemble.

information Figure 5d). It is important to note that the meridional Sverdrup subsurface convergence is already present in the western and central Pacific at lag −30, when incipient LN-like conditions start to reinforce the trade winds (not shown). In the upper Ekman layer, the stronger-than-normal easterly wind stress generates anomalous meridional Ekman divergence (shading in Figure 2e). This configuration with surface meridional divergence and subsurface convergence results into an area of anomalous upper ocean upwelling motion to the east of the dateline (shading in Figure 2d). In the warm pool, anomalous zonal currents are strongly convergent in the surface and divergent in the subsurface (shading in Figure 2c). These zonal signals are locally stronger than the expected meridional Ekman-induced surface divergence and the meridional Sverdrup subsurface convergence, therefore inducing anomalous local downward motion to the west of 170°E (shading in Figure 2d).

The mechanisms described in Figure 2 are consistently observed in all the assimilation products considered in this study (supporting information Figure S6). Interestingly, the magnitude of surface temperature, salinity, and current anomalies and the associated subsurface warming vary coherently between products, so that the stronger these anomalies, the warmer the subsurface buildup. For example, the interproduct correlation between surface salinity (156°–166°E, 0–50 m) and subsurface potential temperature (156°–166°E, 100–130 m) is 0.98 ($p < 0.001$), with a regression coefficient of 3.91°C per g/kg. The only exception is GECCO, which underestimates the zonal surface convergence (subsurface divergence) in the warm pool compared to other products, and does not produce the patterns of meridional surface Ekman divergence (subsurface Sverdrup convergence) in the western and central equatorial Pacific observed in the other assimilation data sets. Interestingly, this is the assimilation product with the weakest subsurface heat buildup. The fact that GECCO is nevertheless able to reproduce the vertical structure of horizontal divergence in the warm pool, and therefore, the associated downwelling motion suggests that the contribution by the subsurface meridional mass convergence and subsurface eastward advection might be underestimated. The associated spread in the simulation of meridional divergence across the different assimilation products allows a rough estimate of the relative dependency between these variables in the western equatorial Pacific. Thus, the interproduct correlation between meridional upper ocean divergence (156°–166°E, 0–200 m) and subsurface potential
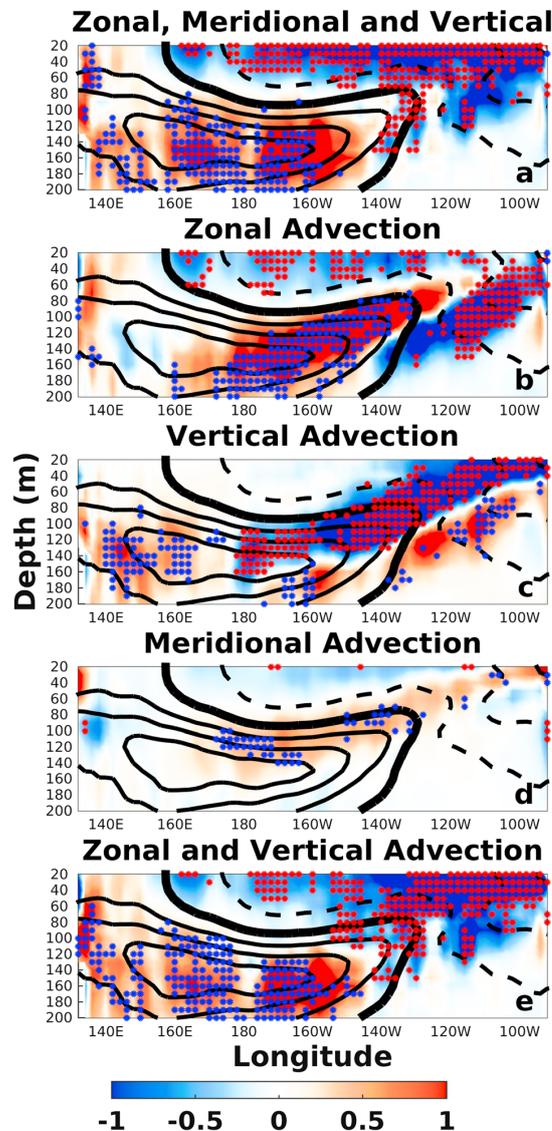
Figure 3. Multiproduct average of heat advection preceding the heat buildup leading to El Niño events. Tendency in potential temperature (°C/yr, contours) and zonal (Figures 3a, 3b, and 3e), meridional (Figures 3a and 3d), and vertical (Figures 3a, 3c, and 3e) heat advection (°C/yr). Composite anomalies are averaged over the equatorial band (2°S–2°N), and shown for lag 30 months before the major El Niño events. Note that this lag shows the temperature tendency that determines the anomalies that are later observed at lag −21. The contour interval is 0.25°C/yr, with solid (dashed) lines depicting positive (negative) anomalies. The stippling denotes areas where shading anomalies have the same sign for all the members of the ensemble.

temperature (156°–166°E, 100–130 m) is −0.90 ($p < 0.001$), with a regression coefficient of $-3.36 \cdot 10^7$°C per 1/s. Instead, surface horizontal convergence (156°–166°E, 20–60 m) is consistently observed in all the assimilation products, with a mean relationship of $-4.38 \cdot 10^7$°C per 1/s with subsurface potential temperature.

The analysis of the temperature tendency equation reveals that the sum of the zonal and vertical advection terms explains a large fraction of the spatial structure of the tendency in surface and subsurface heat anomalies in the equatorial Pacific (cf. shading and contours in Figures 3a and 3e). Vertical advection alone determines the subsurface warming to the west of 170°E (Figure 3c). Instead, to the east of the sharp transition at 170°E, the largest contribution comes from the combination of the zonal (Figure 3b) and vertical (Figure 3c) advection components (Figure 3e), as a result of the strengthening of the equatorial undercurrent and the advection of incipient subsurface temperature anomalies from the western to the central Pacific. Nevertheless, the meridional advection of heat is already contributing to the warming at the level of the thermocline to the east of 170°E (Figure 3d), even if at this early stage of the ENSO oscillation the thermocline has a tendency toward shallower depths in the eastern Pacific (contours in Figure 3). Note that to the west of 170°E, where the subsurface equatorward convergence of mass is as strong as in the central Pacific, the meridional advection of heat is weak because the meridional temperature gradient at the equator is small (not shown).

4. Summary and Discussion

The present work shows that the enhancement of the trade winds in the central equatorial Pacific increases the intensity of the westward circulation of the south equatorial current [*Yu and McPhaden*, 1999], strengthening the surface zonal and thus horizontal current convergence and downwelling motion

near the anomalously tilted warm pool edge and salinity front. Additionally, we show that the vertical advection of heat plays a key role in the redistribution of water masses in the warm pool. Zonal and vertical currents are indeed intimately connected through the energy balance, because a significant
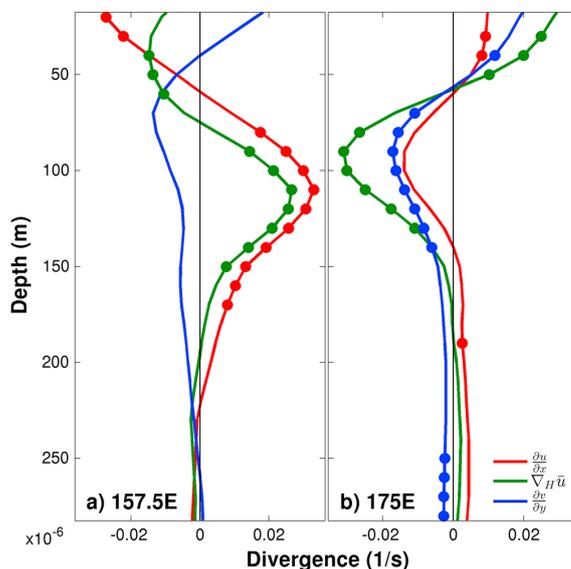
Figure 4. Vertical profiles of equatorial current divergence during the heat buildup leading to El Niño events. Multiproduct average of zonal (red), horizontal (green), and meridional (red) current divergence (1/s) at longitudes (a) 157.5°E and (b) 175°E, averaged over the equatorial band (2°S–2°N), and shown for lag 21 months before the major El Niño events. Dots denote depths where anomalies have the same sign for all the members of the multiproduct ensemble.

fraction of the wind power (whose interannual signal is dominated by the zonal component) is converted into buoyancy power [*Brown and Fedorov, 2010*]. This transfer explains how the energy supplied by enhanced trade winds to the westward south equatorial current in the central Pacific is converted into vertical mass fluxes in the western Pacific that distort local ocean isopycnals and deepen the thermocline [*Brown et al., 2011*].

In this regard, results presented here provide a comprehensive description of the interaction between wind stress forcing in the central Pacific, oceanic circulation in the warm pool, and subsurface temperature and thermocline depth in the western equatorial Pacific starting about 2 years before the peak of EN. Figure 4 summarizes the relevant anomalies in horizontal current divergence for two representative equatorial locations to the west of the dateline. Near the warm pool edge and salinity front (i.e., 157.5°E, Figure 4a), we find downwelling motion from an upper layer (0–75 m) with horizontal convergence to a subsurface layer (75–190 m) with horizontal divergence. The decomposition of horizontal divergence into its zonal and meridional components reveals how at these long lead times different mechanisms contribute in an intricate way to the subsurface heat buildup. On the one hand, horizontal convergence near the surface (0–75 m) is to a large extent explained by surface (0–60 m) zonal convergence and to a lower extent by subsurface (40–75 m) meridional convergence. On the other hand, the subsurface layer (75–190 m) of horizontal divergence is almost entirely explained by the zonal component, characterized by the strengthening of the equatorial undercurrent in the central Pacific. Near the dateline (i.e., 175°E, Figure 4b), vertical profiles of horizontal divergence reveal a different regime, dominated by surface (0–60 m) zonal and meridional divergence and subsurface (60–190 m) zonal and meridional convergence.

These results are based on an ensemble of ocean assimilation products and to the extent these products can be regarded as faithfully capturing the observed dynamics, the anomalies that are consistently simulated by all the data sets can be used to validate previous modeling studies. For instance, *Yu and Mechoso* [2001] used a coupled ocean-atmosphere general circulation model to show that vertical advection contributes to the subsurface (80–145 m) cooling during the mature phase of El Niño (see their Figure 18c). For the sake of this discussion, we will here assume opposite anomalies for LN conditions, which would correspond to our results in Figure 3c. Despite the similarities with our results, the subsurface cooling due to vertical advection in *Yu and Mechoso* [2001] occurs in a broad longitudinal band extending from the western boundary to 150°W, while in the analysis products the subsurface warming due to vertical advection is confined to a very narrow range to the west of 170°E. Coherently, the distribution of horizontal divergence inferred from the assimilation products also shows a sharp transition at 170°E from surface (subsurface) horizontal convergence (divergence) to the west and divergence (convergence) to the east. The spatial characterization emerging from this work might prove useful in determining the relative role of the different processes involved in the generation of the heat buildup in the western Pacific, and ultimately in improving the predictability of ENSO at long lead times beyond the spring barrier.

References

Adler, R. F., et al. (2003), The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present), *J. Hydrometeorol.*, *4*, 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Ballester, J., M. A. Rodríguez-Arias, and X. Rodó (2011), A new extratropical tracer describing the role of the western Pacific in the onset of El Niño: Implications for ENSO understanding and forecasting, *J. Clim.*, *24*, 1425–1437, doi:10.1175/2010JCLI3619.1.

Ballester, J., J. C. Burns, D. Cayan, Y. Nakamura, R. Uehara, and X. Rodó (2013), Kawasaki disease and ENSO-driven wind circulation, *Geophys. Res. Lett.*, *40*, 2284–2289, doi:10.1002/grl.50388.

Balmaseda, M. A., A. Vidard, and D. L. T. Anderson (2008), The ECMWF ocean analysis system: ORA-S3, *Mon. Weather Rev.*, *136*, 3018–3034, doi:10.1175/2008MWR2433.1.

Balmaseda, M. A., K. Mogensen, F. Molteni, and A. T. Weaver (2010), The NEMOVAR-COMBINE ocean re-analysis, *COMBINE Tech. Rep. No. 1*.

Balmaseda, M. A., K. Mogensen, and A. T. Weaver (2013), Evaluation of the ECMWF ocean reanalysis system ORAS4, *Q. J. R. Meteorol. Soc.*, *139*, 1132–1161, doi:10.1002/qj.2063.

Battisti, D. S. (1988), Dynamics and thermodynamics of a warming event in a coupled tropical atmosphere–ocean model, *J. Atmos. Sci.*, *45*, 2889–2919, doi:10.1175/1520-0469(1988)045<2889:DATOAW>2.0.CO;2.

Bjerknes, J. (1969), Atmospheric teleconnections from the equatorial Pacific, *Mon. Weather Rev.*, *97*, 163–172, doi:10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2.

Bosc, C., T. Delcroix, and C. Maes (2009), Barrier layer variability in the western Pacific warm pool from 2000 to 2007, *J. Geophys. Res.*, *114*, C06023, doi:10.1029/2008JC005187.

Brown, J. N., and A. V. Fedorov (2010), How much energy is transferred from the winds to the thermocline on ENSO time scales?, *J. Clim.*, *23*, 1563–1580, doi:10.1175/2009JCLI2914.1.

Brown, J. N., A. V. Fedorov, and E. Guilyardi (2011), How well do coupled models replicate ocean energetics relevant to ENSO?, *Clim. Dyn.*, *36*, 2147–2158, doi:10.1007/s00382-010-0926-8.

Carton, J. A., and B. S. Giese (2008), A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA), *Mon. Weather Rev.*, *136*, 2999–3017, doi:10.1175/2007MWR1978.1.

Climate Prediction Center (2013), Cold and warm episodes by season. [Available at http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml Last accessed in January 2014.]

Fedorov, A. V., and J. Brown (2009), Equatorial waves, in *Encyclopedia of Ocean Sciences*, 2nd ed., edited by J. Steele, pp. 3679–3695, Academic Press. San Diego, Calif.

Jin, F. F. (1997a), An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model, *J. Atmos. Sci.*, *54*, 811–829, doi:10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2.

Jin, F. F. (1997b), An equatorial ocean recharge paradigm for ENSO. Part II: A stripped-down coupled model, *J. Atmos. Sci.*, *54*, 830–847, doi:10.1175/1520-0469(1997)054<0830:AEORPF>2.0.CO;2.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Köhl, A., and D. Stammer (2008), Variability of the meridional overturning in the North Atlantic from the 50-year GECCO state estimation, *J. Phys. Oceanogr.*, *38*, 1913–1930, doi:10.1175/2008JPO3775.1.

Meinen, C. S., and M. J. McPhaden (2000), Observations of warm water volume changes in the equatorial pacific and their relationship to El Niño and La Niña, *J. Clim.*, *13*, 3551–3559, doi:10.1175/1520-0442(2000)013<3551:OOWWVC>2.0.CO;2.

Picaut, J., M. Ioualalen, C. Menkes, T. Delcroix, and M. J. McPhaden (1996), Mechanism of the zonal displacements of the Pacific warm pool: Implications for ENSO, *Science*, *274*, 1486–1489, doi:10.1126/science.274.5292.1486.

Ramesh, N., and R. Murtugudde (2013), All flavours of El Niño have similar early subsurface origins, *Nat. Clim. Change*, *3*, 42–46, doi:10.1038/nclimate1600.

Schopf, P. S., and M. J. Suarez (1988), Vacillations in a coupled ocean–atmosphere model, *J. Atmos. Sci.*, *45*, 549–566, doi:10.1175/1520-0469(1988)045<0549:VIACOM>2.0.CO;2.

Singh, A., T. Delcroix, and S. Cravatte (2011), Contrasting the flavors of El Niño-Southern Oscillation using sea surface salinity observations, *J. Geophys. Res.*, *116*, C06016, doi:10.1029/2010JC006862.

Trenberth, K. E., and T. J. Hoar (1996), The 1990–1995 El Niño-Southern Oscillation event: Longest on record, *Geophys. Res. Lett.*, *23*, 57–60, doi:10.1029/95GL03602.

Wang, B. (2002), Kelvin waves, in *Encyclopedia of Atmospheric Sciences*, edited by M. Shankar, pp. 1062–1068, Elsevier, Amsterdam.

Wyrtki, K. (1975), El Niño—The dynamic response of the equatorial Pacific Ocean to atmospheric forcing, *J. Phys. Oceanogr.*, *5*, 572–584, doi:10.1175/1520-0485(1975)005<0572:ENTDRO>2.0.CO;2.

Wyrtki, K. (1981), An estimate of equatorial upwelling in the Pacific, *J. Phys. Oceanogr.*, *11*, 1205–1214, doi:10.1175/1520-0485(1981)011<1205:AEOEUI>2.0.CO;2.

Yu, J. Y., and C. R. Mechoso (2001), A coupled atmosphere–ocean GCM study of the ENSO cycle, *J. Clim.*, *14*, 2329–2350, doi:10.1175/1520-0442(2001)014<2329:ACAOGS>2.0.CO;2.

Yu, X., and M. J. McPhaden (1999), Seasonal variability in the equatorial Pacific, *J. Phys. Oceanogr.*, *29*, 925–947, doi:10.1175/1520-0485(1999)029<0925:SVITEP>2.0.CO;2.

3.2 Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model

Petrova D (1), Koopman SJ (2), Ballester J (1,3), Rodó X (1,4)

(1) Institut Català de Ciències del Clima, Barcelona, Catalonia, Spain
(2) Vrije Universiteit Amsterdam, Amsterdam, Netherlands
(3) California Institute of Technology, Pasadena, California, United States
(4) Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

Abstracto - *Summary in Spanish*

El Niño (EN) es una característica dominante de la variabilidad del clima en escalas de tiempo interanuales que impulsa cambios en el clima a lo largo del globo, y tiene consecuencias naturales y socioeconómicas ampliamente difundidas. En este sentido, su pronóstico es una tarea importante, y las predicciones se emiten de manera regular por una amplia gama de esquemas de predicción y centros climáticos en todo el mundo. Este estudio explora un nuevo método para la previsión de EN. Sin embargo, la técnica estadística ventajosa de modelado de series temporales con componentes no observados, también conocido como modelado de series temporales estructurales, no se ha aplicado para la predicción de El Niño. Por lo tanto, hemos desarrollado un modelo en el que el análisis estadístico, incluyendo la estimación de parámetros y la predicción, se basa en los métodos de espacio-estado, e incluye el conocido filtro de Kalman. La característica distintiva de este modelo dinámico es la descomposición de una serie temporal en una gama de componentes estocásticamente variables en el tiempo, tales como nivel (o tendencia), estacional, ciclos de diferentes frecuencias, irregular y los efectos de regresión incorporados como covariables explicativas. Estos componentes se modelan por separado y en última instancia combinados en un único esquema de previsión. Modelos de predicción estadísticas habituales de EN usan esencialmente SST y el estrés del viento en el Pacífico ecuatorial. Además de esto, introducimos un nuevo dominio de variables de regresión que explican el estado de la temperatura subterránea del océano en el Pacífico ecuatorial central y occidental, motivado por nuestro análisis, así como por la investigación reciente y clásica, mostrando que procesos subsuperficiales y la acumulación de calor son fundamentales para el génesis de EN. Una característica importante del esquema es que los diferentes predictores de regresión se utilizan en diferentes meses de antelación, capturando así la evolución dinámica del sistema y haciendo pronósticos más eficientes. El nuevo modelo ha

sido probado con la predicción de todos los eventos calientes ocurridos en el período 1996-2015. Las previsiones retrospectivas de estos eventos se hicieron con mucho tiempo de antelación de al menos dos años y medio. Por lo tanto, el presente estudio demuestra que el límite teórico de la predicción de ENSO debería ser mucho más largo de la comúnmente aceptada "Barrera de Primavera". La alta correspondencia entre las previsiones y las observaciones indica que el modelo propuesto supera todos los modelos estadísticos operacionales y se comporta de manera comparable a los mejores modelos dinámicos utilizados para la predicción EN. Así, la nueva manera en que el esquema de modelización se ha estructurado también podría mejorar otros sistemas de modelización estadísticos y dinámicos.

CrossMark

Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model

Desislava Petrova[1,2] · Siem Jan Koopman[3] · Joan Ballester[1,4] · Xavier Rodó[1,5]

Abstract El Niño (EN) is a dominant feature of climate variability on inter-annual time scales driving changes in the climate throughout the globe, and having wide-spread natural and socio-economic consequences. In this sense, its forecast is an important task, and predictions are issued on a regular basis by a wide array of prediction schemes and climate centres around the world. This study explores a novel method for EN forecasting. In the state-of-the-art the advantageous statistical technique of unobserved components time series modeling, also known as structural time series modeling, has not been applied. Therefore, we have developed such a model where the statistical analysis, including parameter estimation and forecasting, is based on state space methods, and includes the celebrated Kalman filter. The distinguishing feature of this dynamic model is the decomposition of a time series into a range of stochastically time-varying components such as level (or trend), seasonal, cycles of different frequencies, irregular, and regression effects incorporated as explanatory covariates. These components are modeled separately and ultimately combined in a single forecasting scheme. Customary statistical models for EN prediction essentially use SST and wind stress in the equatorial Pacific. In addition to these, we introduce a new domain of regression variables accounting for the state of the subsurface ocean temperature in the western and central equatorial Pacific, motivated by our analysis, as well as by recent and classical research, showing that subsurface processes and heat accumulation there are fundamental for the genesis of EN. An important feature of the scheme is that different regression predictors are used at different lead months, thus capturing the dynamical evolution of the system and rendering more efficient forecasts. The new model has been tested with the prediction of all warm events that occurred in the period 1996–2015. Retrospective forecasts of these events were made for long lead times of at least two and a half years. Hence, the present study demonstrates that the theoretical limit of ENSO prediction should be sought much longer than the commonly accepted "Spring Barrier". The high correspondence between the forecasts and observations indicates that the proposed model outperforms all current operational statistical models, and behaves comparably to the best dynamical models used for EN prediction. Thus, the novel way in which the modeling scheme has been structured could also be used for improving other statistical and dynamical modeling systems.

Keywords El Niño Southern Oscillation · Prediction · Predictability · Subsurface dynamics · Time series

✉ Desislava Petrova
desislava.petrova@ic3.cat

1 Climate Dynamics and Impacts Unit, Catalan Institute of Climate Sciences (IC3), Carrer del Dr. Trueta, 203, 08005 Barcelona, Catalonia, Spain

2 Department of Physics, University of Barcelona (UB), Barcelona, Catalonia, Spain

3 Department of Econometrics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

4 California Institute of Technology (Caltech), Pasadena, CA, USA

5 Institució Catalana de Recerca i Estudis Avancats (ICREA), Barcelona, Catalonia, Spain

1 Introduction

The equatorial Pacific region, and especially the western equatorial Pacific (WPAC), is fundamental for the

D. Petrova et al.

modulation of the weather patterns worldwide through changes in the surface-boundary condition and the thermal heating of the atmosphere, while the signal of El Niño Southern Oscillation (ENSO) variability in this region is the strongest after the seasonal cycle (Sarachik and Cane 2010). Therefore, it is not surprising that ENSO is a subject of intense scientific research and schemes for its prediction are plentiful with various rates of success (Barnston et al. 2012). At the same time, the limitations of its predictability are debated, with some studies arguing that high-frequency atmospheric noise prevents accurate forecasts to be made at long lead times (Penland 1996; Thompson and Battisti 2000, 2001; Fedorov et al. 2003). Others, instead, imply that EN is to a larger extent the result of an internal self-sustained dynamics, which allows for its prediction well ahead of its typical December peak (Zebiak and Cane 1987; Goswami and Shukla 1991; Jin et al. 1994; Tziperman et al. 1994; Chen et al. 2004).

It is interesting that despite the major improvements of dynamical models, initialization techniques, and the longer and more accurate data sets available for statistical predictions, verification of the real-time ENSO prediction skills during 2002–2011 indicates skills somewhat lower than those found for the less advanced models of the 1980s and 1990s (Barnston et al. 2012). In addition, the so-called spring barrier continues to represent a major difficulty for all types of ENSO forecasting models, and their skill for predictions made in the months March, April, May is still non-significant. A potential explanation could lie in the seasonal cycle, because during boreal spring the tropical Pacific is warmest and this seasonal warming is erroneously taken by the models for the growing phase of EN (Sarachik and Cane 2010). Moreover, the low signal-to-noise ratio in spring also accounts for the lower prediction skill of model forecasts issued in these months (Torrence and Webster 1998). Thus, advances in the understanding of the dynamical mechanisms involved in the initiation of El Niño and the transition to La Niña, as well as the introduction of new and improved ENSO prediction models remain necessary.

The majority of the existing statistical forecasting schemes are based on sea surface temperature, sea level pressure, and wind stress data in the equatorial Pacific (Barnston and Ropelewski 1992; Penland and Magorian 1993; Xue et al. 1994; Kondrashov et al. 2005). Data sets directly capturing subsurface temperature changes in the tropical Pacific region, however, have not been incorporated in the existing empirical models. As three-dimensional observations of the tropical Pacific (TAO-TRITON data) are only available since the 1990s, and as the statistical models used for operational ENSO forecasting have remained predominantly unchanged over the last 10–15 years, the subsurface indicators that play a fundamental role in ENSO variability have not been included in the existing forecasting skills (Barnston et al. 2012). At the same time, several studies have already highlighted the significance of the accumulation of heat in the subsurface of the western tropical Pacific for the onset of EN (Ramesh and Murtugudde 2013; Ballester et al. 2015).

Wyrtki 1975 has proposed a theory, later further developed in the dynamical recharge-discharge oscillatory theory by Jin (1997a, b), according to which an intensification of the trade winds in the central equatorial Pacific leads to the accumulation of heat in the western tropical Pacific subsurface. Then, with the relaxation of the trade winds and the occurrence of westerly anomalies in the western/central tropical Pacific, this heat is released and allowed to move to the east during the growing phase of EN. What has not been explicitly discussed in this theory, though, is that the propagation of the warm waters/heat towards the east occurs mainly in the subsurface near the thermocline (Ramesh and Murtugudde 2013; Ballester et al. 2015), and warm anomalies appear on the surface in the eastern equatorial Pacific only at a later stage of this growing phase. Moreover, warm anomalies in the subsurface (down to 400–500 m depth) occur on average as early in the WPAC as 26–30 months prior to the peak of El Niño, and are present there at different levels of depth up to 10–11 months before the peak, when the heat starts to be advected eastwards.

These patterns at longer lags demonstrate that subsurface temperature and surface wind stress indices in the WPAC and the central Pacific (CPAC) could be promising as regression predictor variables for EN. At the same time, sea surface temperature anomalies (SSTAs) there exist at lags 15–17 months, but they are short-lived and not as strong in magnitude as the subsurface anomalies, so their role as predictors could be more limited.

Considering the above, in the present study we propose a statistical unobserved component time series model for El Niño forecasting based on wind stress, surface and subsurface ocean temperatures in the western and central tropical Pacific. The lead/lag times for these variables are between 30 and 0 months prior to the peak of EN, and as it is customary, predictions are made for the SSTs in the Niño 3.4 region [5°N–5°S, 170°W–120°W, (Barnston et al. 1997); black time series in Fig. 1a]. Section 2 of this manuscript describes the data sets and statistical methods used for the analysis; in Sect. 3 we give a detailed explanation of the proposed ENSO modeling technique, while in Sect. 4 we clarify how the model is applied; in Sect. 5 we present some forecasting results, and then we end with a discussion of the results in Sect. 6, and conclusions in Sect. 7.

2 Data and statistical analysis

Wind stress data has been derived from the NCEP/NCAR reanalysis (Kalnay et al. 1996). Sea surface temperature
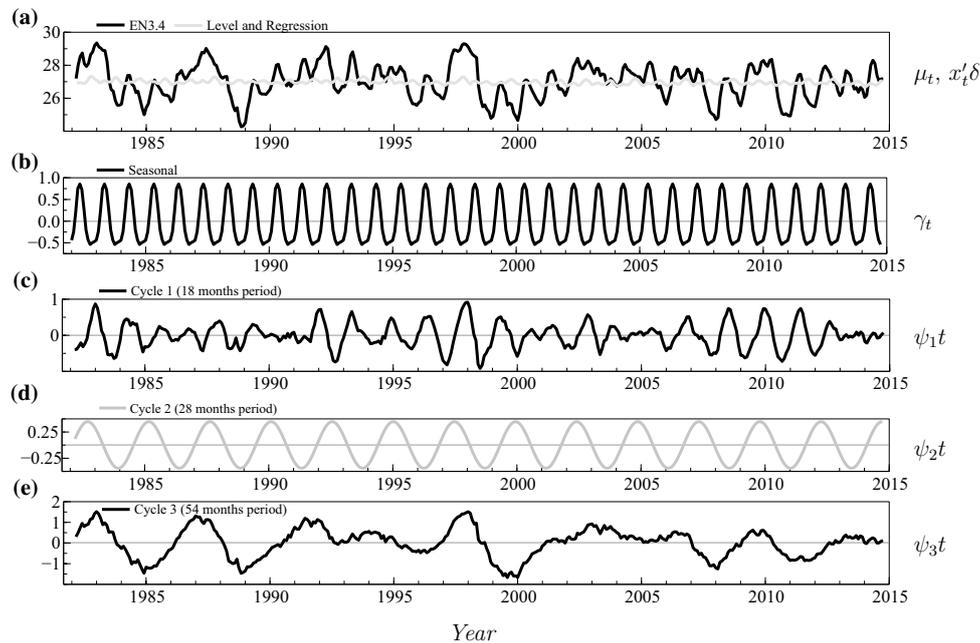
Fig. 1 Components graphics of the model. Shown are temperature (°C) time series of the **a** level and regression components together with the N3.4 index, **b** the seasonal component, and the three cycle components of periods: **c** 1.5, **d** 2.5, **e** 4.5 years

is the NOAA-ERSST-V3 before 1982 and the NOAA-OI-SST-V2 data afterwards, provided by the NOAA/OAR/ESRL PSD (www.esrl.noaa.gov/psd/). Subsurface temperature data before 2012 is from the Subsurface Temperature And Salinity Analyses by Ishii et al. (2005), archived at the National Center for Atmospheric Research, Computational and Information Systems Laboratory (www.rda.ucar.edu/datasets/ds285.3/), and from the Hadley Centre EN4.0.2 analyses data (Good et al. 2013) afterwards. The EN events used for the composite anomalies are based on the definition of the Climate Prediction Center for the period 1978–2012: December 1982, 1991, 1997, 2002, 2006, 2009 (CPC 2015). The 1986–1987 event was excluded from the composite calculations because the peak of this event was in September–October rather than in December–January. However, calculations with this event included do not produce significantly different results, and the temperature and zonal wind anomalies at the respective lag times are qualitatively the same (not shown).

A Multi Taper Method (MTM) was used for the spectrum density estimation and signal reconstruction of the N3.4 time series shown in Fig. 2. MTM is a nonparametric method that reduces the variance of spectral estimates

through the use of a small set of orthogonal tapers, which are multiplied by the data to minimize spectral leakage (Ghil et al. 2002).

A recursive Butterworth procedure was used (Moron and Plaut 2003; Ballester et al. 2011) to filter SST and windstress data, so that only frequencies corresponding to periods of 14–18 months (Fig. 3), 24–28 months (Fig. 4), and 46–63 months (Fig. 5) have been kept. A Butterworth filter was also applied to remove high-frequency variability and thus filter out the annual cycle in the composites of sea surface and subsurface temperatures, and windstress anomalies in Figs. 6, 7 and 8.

Complex Orthogonal Function (CEOF) analyses was applied to the area-weighted filtered SSTAs and windstress anomalies in the equatorial Pacific (Figs. 3, 4, 5). This technique decomposes variability into real and imaginary spatial patterns that are amplified by real and imaginary time-varying coefficients, respectively. The patterns represent the main modes of variability of the data as a function of the phase (ϕ) in periodic spatial coefficients and periodic temporal scores (Ballester et al. 2011). The temporal scores were correlated with the filtered spatio-temporal SSTAs and windstress anomalies in the tropical Pacific area.
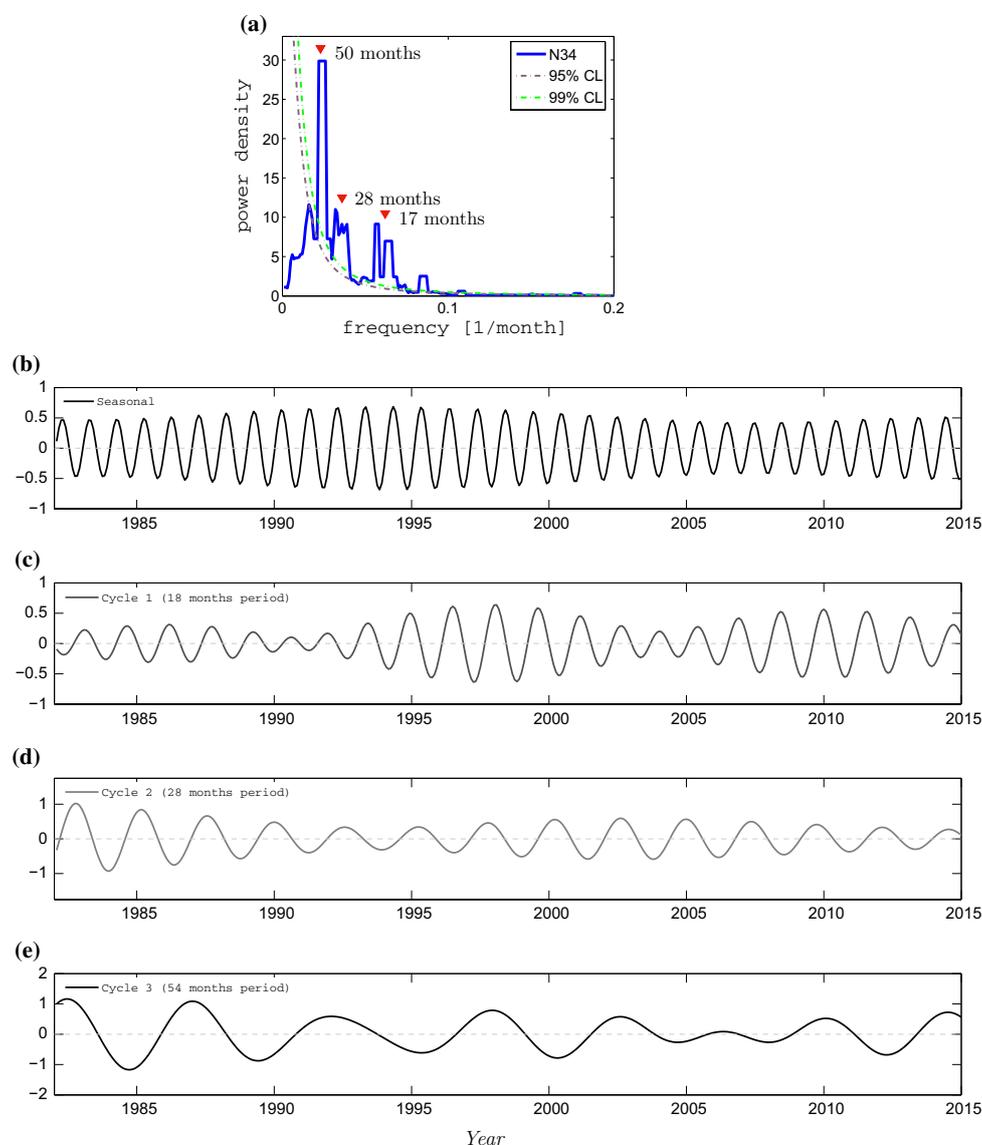
Fig. 2 a Multi Taper Method (MTM) power spectra for the observed N3.4 time series. The *solid line* indicates the power density and *dashed lines* the respective confidence level (CL) based on a *red* noise null hypothesis. The *red* indicators correspond to the near-annual, biannual, and quasi-quadrennial ENSO modes of variability. Reconstructed components from the multitaper decomposition in **a**, corresponding to the **b** seasonal, **c** near-annual, **d** biannual, **e** quasi-quadrennial modes

3 Model

Structural time series models are formulated in terms of unobserved components of interest that could be directly interpreted (Harvey 1989). For example, unobserved components may represent time series features such as trend, seasonal and cycles. This type of models can also be extended naturally with regression effects. In this way, they are able to provide a description of the salient features of a given time series and, at the same time, a basis for making predictions of future observations (Harvey and Shephard 1993). Moreover, time series models based on
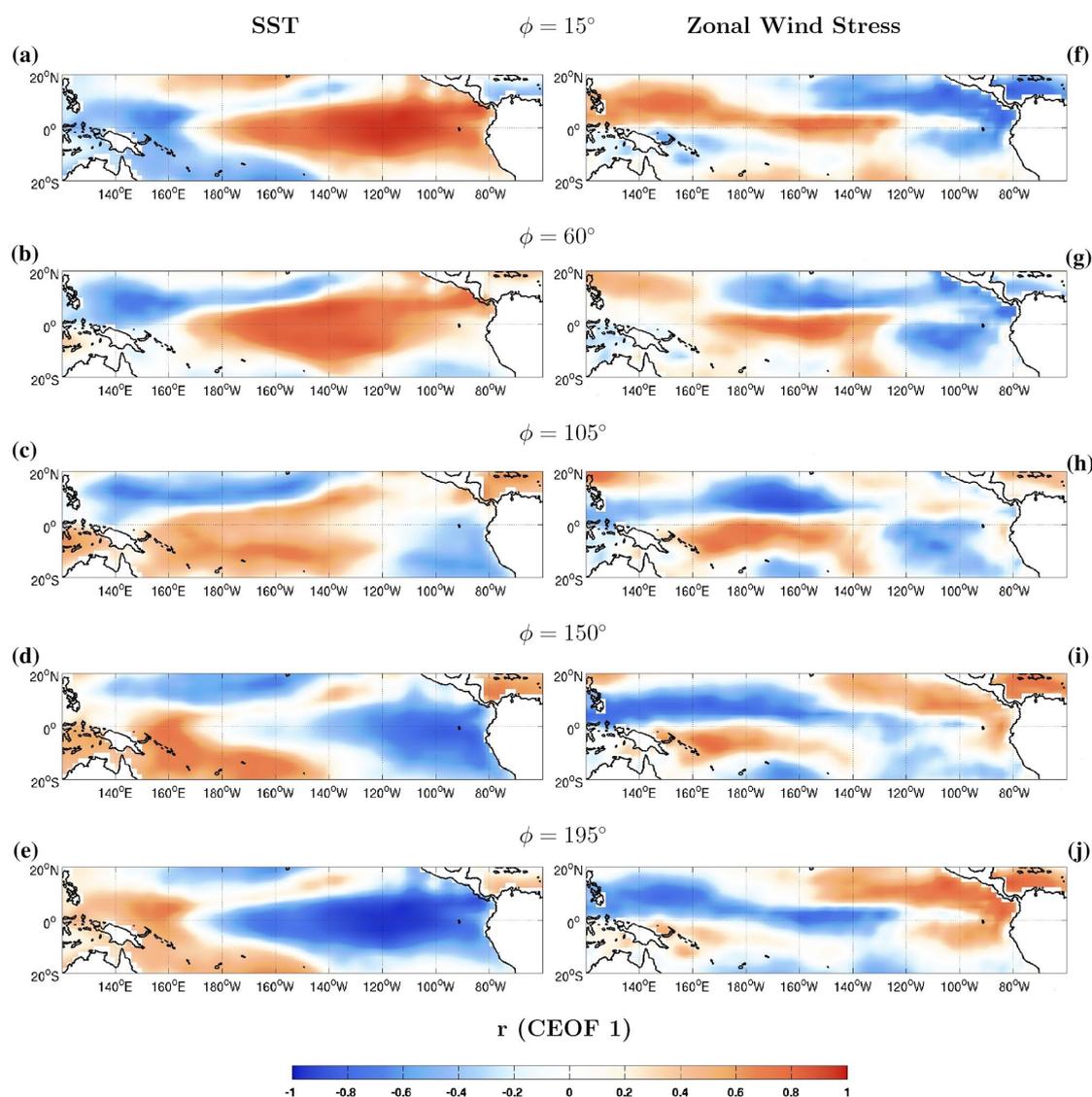
Fig. 3 Pearson correlations between the temporal scores of the first CEOF modes of filtered SSTs and surface wind stress anomalies, and filtered spatio-temporal SST anomalies and wind stress anomalies in the equatorial Pacific region. A Butterworth filter has been applied to the SST and wind stress data sets, so that only frequencies corresponding to periods between 14 and 18 months (associated with the near-annual mode of variability) have been kept. *Panels* correspond to the respective phases of the CEOF shown on the figure. *Shaded areas* indicate significant anomalies

unobserved components are particularly effective when complex features are present in the time series, such as mixed frequencies, multiple modes of variability of the time series, outliers, structural breaks and nonlinear and/ or non-Gaussian processes (Tong 1990; Harvey et al. 1998).

The flexibility provided by this modeling approach makes it a suitable framework for treating time series with complex features and nonlinearities (Durbin and Koopman 2012), common in the climate system and in ENSO in particular. The linear univariate unobserved components time series model that we consider is given by

D. Petrova et al.
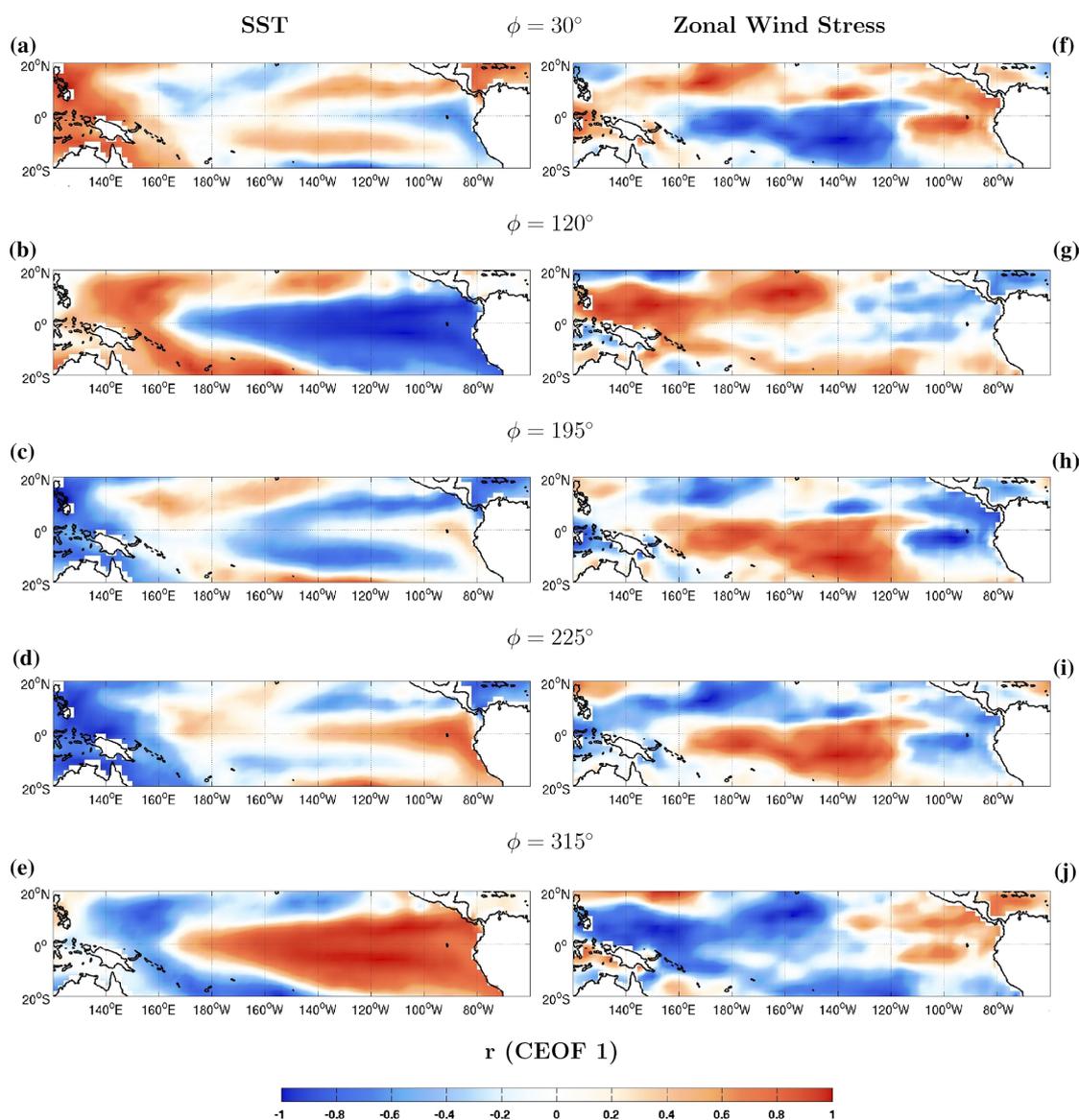
Fig. 4 Same as Fig. 3, but the Butterworth filter has been applied so that only frequencies corresponding to periods between 24 and 28 months (associated with the biannual mode of variability) have been kept

$$y_t = \mu_t + \gamma_t + \psi_{1t} + \psi_{2t} + \psi_{3t} + x_t'\delta + \varepsilon_t \qquad (1)$$

where y_t represents observations on a single variable y, in this case the monthly N3.4 index at time t; μ_t is a level component that we specify as a random walk process; γ_t is a seasonal component with seasonal length $S = 12$; ψ_{1t}, ψ_{2t} and ψ_{3t} are three cycle components with different parameter

values for frequency λ, persistence φ_ψ and variance σ_κ^2; $x_t'\delta$ represents a predictor regression variable based on covariate vector x_t and coefficient vector δ; and ε_t is the remainder irregular term. The trend, seasonal, and cycle components are modeled by linear dynamic stochastic processes, which depend on disturbances (Harvey and Koopman 2000). For more information the reader is referred to Harvey (1989)
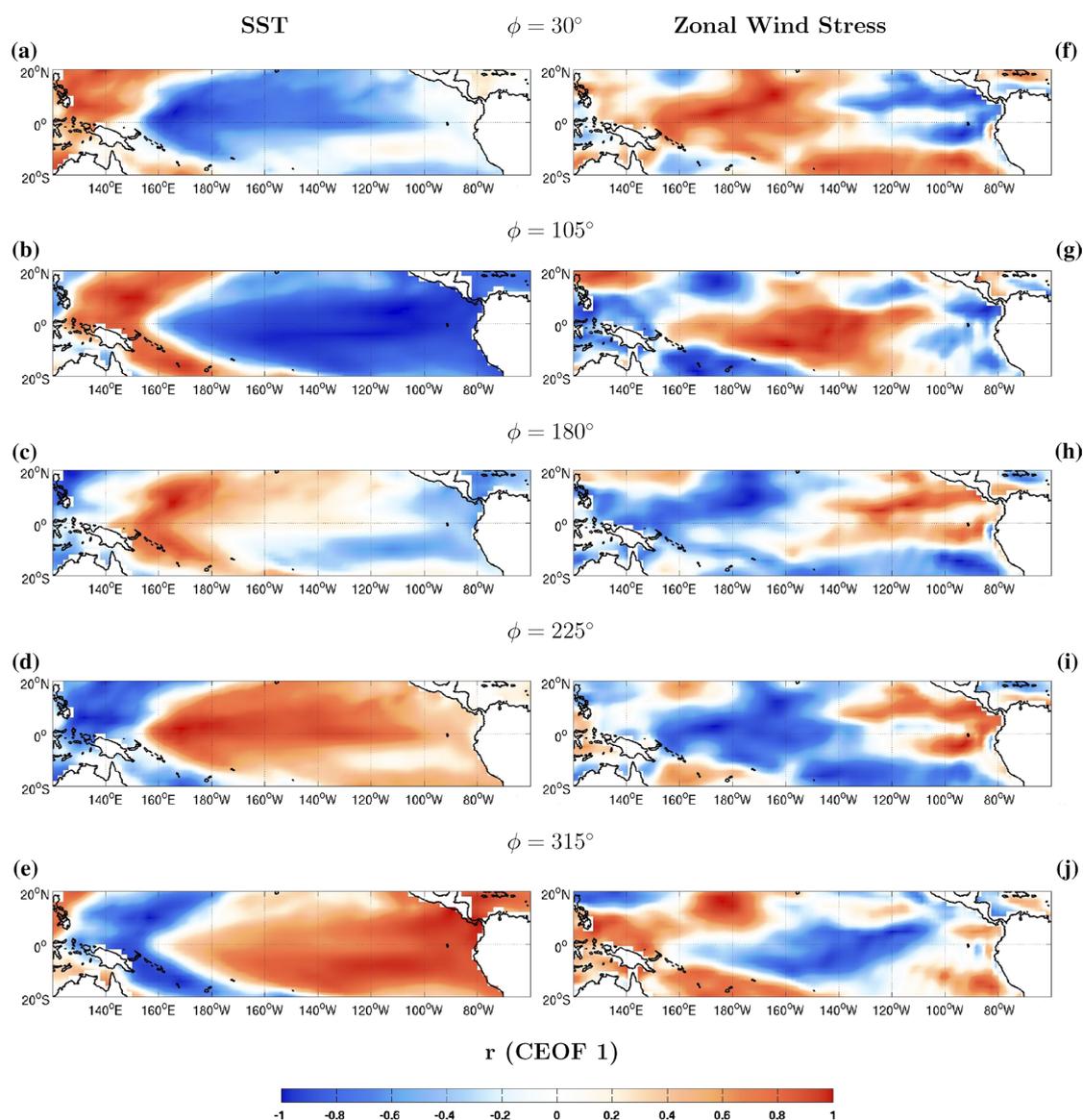
Fig. 5 Same as Fig. 3, but the Butterworth filter has been applied so that only frequencies corresponding to periods between 46 and 63 months (associated with the low-frequency mode of variability) have been kept

and Durbin and Koopman (2012). The components are formulated in a flexible way as stochastic functions of time. The disturbances driving the components are independent of each other.

This structural time series model can be represented in state space form (Durbin and Koopman 2012), and relies on the state and disturbance vectors as given by

$$\alpha_t = \left(\mu_t, \gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-10}, \psi_{1t}, \psi_{1t}^+, \psi_{2t}, \psi_{2t}^+, \psi_{3t}, \psi_{3t}^+, \delta'\right)',$$
$$\epsilon_t = \left(\varepsilon_t, \eta_t, \omega_t, \kappa_{1t}, \kappa_{1t}^+, \kappa_{2t}, \kappa_{2t}^+, \kappa_{3t}, \kappa_{3t}^+\right)',$$

where we assume that the dimension of the regression coefficient vector δ is $k \times 1$. The corresponding system matrices and details of the state space representation of the model are given in Appendix 1. The state vector contains the

D. Petrova et al.

Fig. 6 Composites of surface zonal wind stress (Nm^{-2}, *arrows*) anomalies with respect to all EN events in the period 1978–2012. Shown are anomalies **a** 24, **b** 13, **c** 7 months before the winter peak of EN. The *red boxes* indicate the three zonal wind stress regions from Table 1—**a** Region I, **b** Region II, **c** Region III
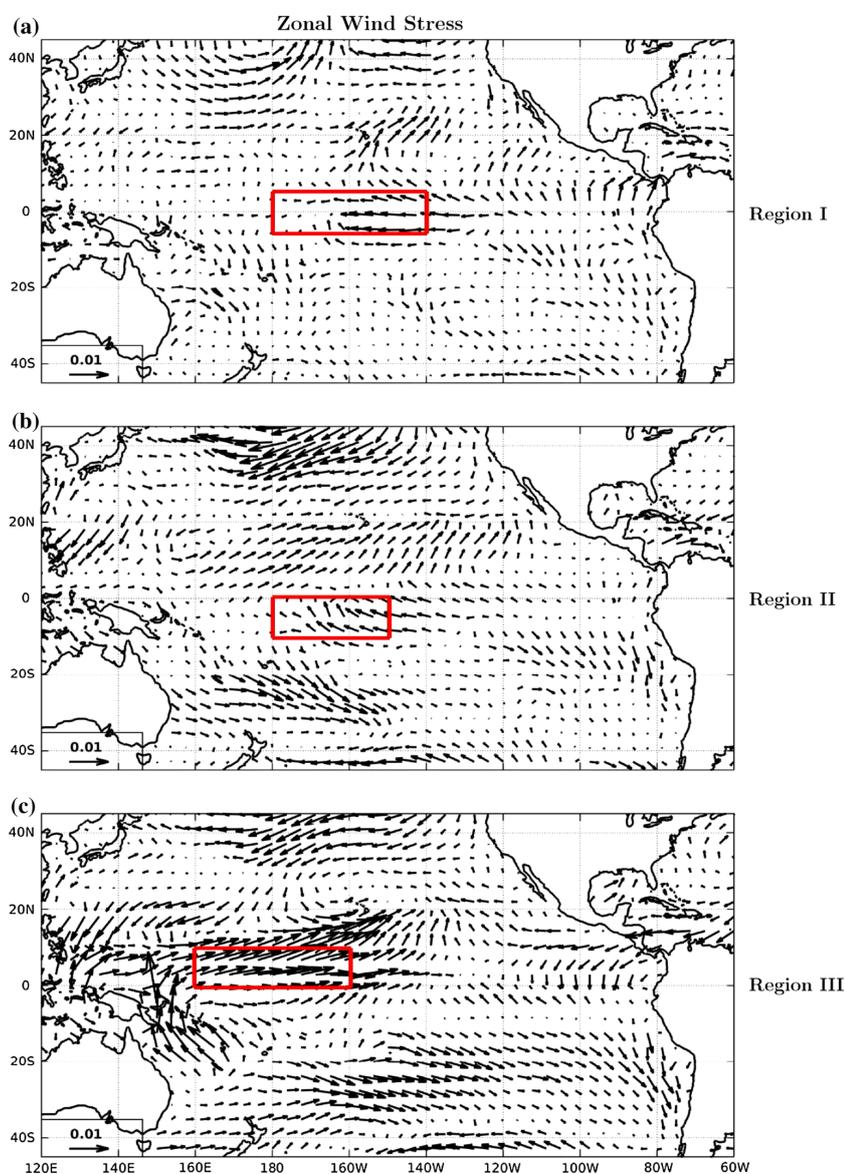
components, auxiliary variables necessary for the dynamic formulations of the components, and the regression coefficient vector. The disturbance vector ϵ_t contains all the disturbances required for the linear stochastic functions of the components. All unknown disturbance variances need to be estimated together with the discount factor $\varphi_{\psi,j}$ and cycle frequencies $\lambda_{c,j}$ for $j = 1, 2, 3$, for the three cycles. For the estimation of the variances, we consider log-transformations to enforce positive values only. The restrictions

$0 < \varphi_{\psi,j} < 1$ and $0 < \lambda_{c,j} < \pi$, for $j = 1, 2, 3$, are also enforced via transformations.

The unknown parameters are collected in the 12×1 parameter vector θ that is given by

$$\theta = (a_\varepsilon, a_\eta, a_\omega, a_{\kappa,1}, a_{\kappa,2}, a_{\kappa,3}, b_{\psi,1}, b_{\psi,2}, b_{\psi,3}, d_1, d_2, d_3)',$$

where

$$\sigma_\ell = \exp(a_\ell), \quad \varphi_{\psi,j} = \exp(b_{\psi,j})/[1 + \exp(b_{\psi,j})],$$
$$\lambda_{c,j} = 2\pi/[2 + \exp(d_j)],$$

Region I

(a) Composites for 150m depth

(b) Composites for 250m depth

(c) Composites for 400m depth
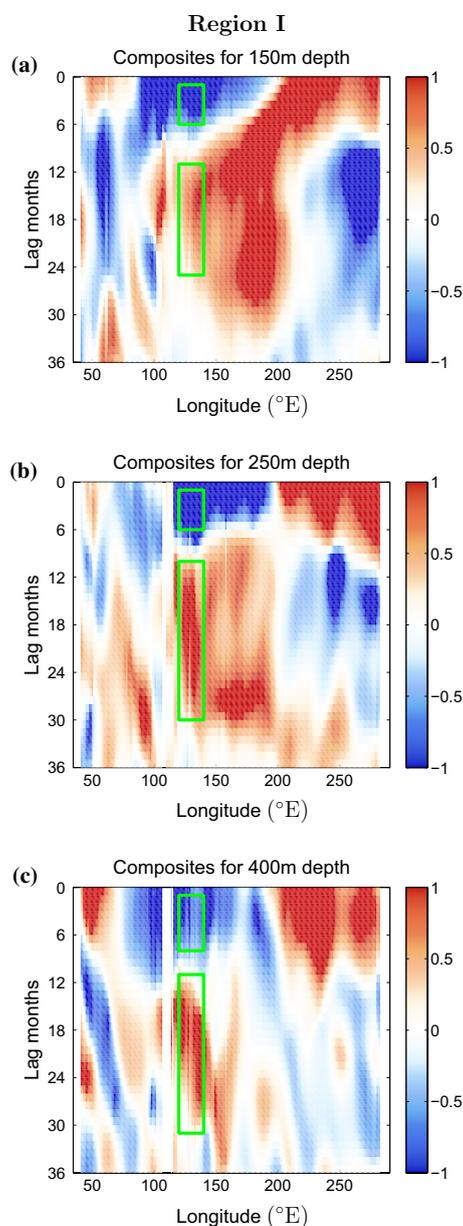
Fig. 7 Composites of subsurface temperature (°C, *shading*) anomalies at **a** 150, **b** 250, **c** 400 m depth with respect to all EN events in the period 1978–2012 in Region I (see Table 2). Data is filtered using a low-pass Butterworth filter (cut-off frequency 18, order 10)

Region II

(a) Composites for 0m depth

(b) Composites for 100m depth

(c) Composites for 200m depth

Fig. 8 Same as Fig. 7, but at **a** surface, **b** 100, **c** 200 m depth in Region II (see Table 2)

Steady state is reached after weak convergence relative to 1e–007.

The statistical treatment of the model relies heavily on the celebrated Kalman Filter (Kalman 1960), which enables the signal extraction of the components, likelihood

for $\ell = \varepsilon, \eta, \omega, \{\kappa, j\}$ and $j = 1, 2, 3$. Estimation is carried out via the numerical maximization of the likelihood function with respect to the transformed parameters.

D. Petrova et al.

evaluation, and forecasting. A general treatment of state space modeling is presented in Durbin and Koopman (2012). For our specific model, we provide the details in Appendix 1. All estimations and forecasts are generated by STAMP, SsfPack and OxMetrics (Koopman et al. 2008, 2010; Doornik 2013). The dynamic formulations of the components, together with the regression effects, are discussed next.

3.1 Level component (long-term variability), μ_t

The level component is modeled as a random walk process and is given by

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathrm{NID}\left(0, \sigma_\eta^2\right), \tag{2}$$

where $\mathrm{NID}(0, \sigma^2)$ refers to a normally independently distributed series with mean zero and variance σ^2. The disturbance series η_t is serially independent and mutually independent of all other disturbance series related to y_t. The initial trend μ_1 is for simplicity treated as an unknown coefficient that needs to be estimated together with the unknown variance σ_η^2. This component is included in the model to account for the long-term fluctuations of the N3.4 time series around its mean, assuming that it is stationary (see Fig. 1a). A simple interpretation is that the level at the current step is equal to the level in the previous step plus a white noise disturbance (Harvey 1989).

3.2 Seasonal component, γ_t

To account for the monthly variation in the annual cycle of the N3.4 time series, the component γ_t is included in the model. More specifically, γ_t represents the seasonal effect at time t that is associated with season $s = s(t)$ for $s = 1, \ldots, S$, where S is the seasonal length ($S = 12$ for monthly data). In particular, we adopt the notation

$$\gamma_t = \gamma_t^{[s]},$$

if we need to emphasize that γ_t represents the seasonal effect for season s.

We use a seasonal pattern that is fixed over time (see Fig. 1b), i.e. we have a set of S seasonal effects $\gamma_1, \ldots, \gamma_S$ which are taken as unknown coefficients that need to be estimated together with the other coefficients in the model. The seasonal effects must have the property that they sum to zero over the full year to make sure that they are not confounded with the trend component, that is

$$\gamma_1 + \cdots + \gamma_S = 0, \quad \gamma_t = \gamma_{t-S}, \quad t = S + 1, \ldots, n. \tag{3}$$

The seasonal pattern could also change slowly over time, by relaxing the summing-to-zero requirement with a stochastic equation

$$\gamma_{t+1} = -\gamma_t - \cdots - \gamma_{t-S+2} + \omega_t, \quad \omega_t \sim \mathrm{NID}(0, \sigma_\omega^2), \tag{4}$$

In the present study the disturbance variance $\omega_t = 0$. In this way we have $S - 1$ unknown seasonal coefficients that need to be estimated by the Kalman Filter.

There is a marked seasonal cycle in the tropical Pacific, which has a substantial impact on the ENSO cycle and the evolution of its phases (Tziperman et al. 1997; Krishnamurthy et al. 2015). This effect is known as phase-locking to the annual cycle (Rasmusson and Carpenter 1982; An and Choi 2009; Stein et al. 2011), as ENSO phases normally grow in the boreal summer and autumn and the peak of the associated anomaly is in the winter months of December, January and February (DJF; Sarachik and Cane 2010). Thus, including an annual cycle component in the forecasting model is necessary for the correct seasonal development of the signal and, therefore, for a more accurate prediction. It is in fact a fundamental part of the ENSO dynamics, especially as the atmospheric conditions should be appropriate in order for an initial surface and subsurface ocean temperature perturbation to grow and propagate. The seasonal variations of the atmospheric convergence zones modulate unstable ocean-atmosphere interactions (Philander 1989), which in turn favour the development of EN or LN.

3.3 Cycle components, $\psi_i t$

In order to account for interannual variability present in the ENSO phenomenon (Fig. 2), we further include a number of additional cycle components. A stochastic formulation of a cycle component can be based on a time-varying trigonometric process, but with frequency λ_c associated with the typical length of a cycle. We have

$$\begin{pmatrix} \psi_{t+1} \\ \psi_{t+1}^+ \end{pmatrix} = \varphi_\psi \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{pmatrix} \psi_t \\ \psi_t^+ \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^+ \end{pmatrix}, \tag{5}$$

where the discount factor $0 < \varphi_\psi < 1$ is introduced to enforce a stationary process for the stochastic cycle component. The disturbances and the initial conditions for the cycle variables are given by

$$\begin{pmatrix} \kappa_t \\ \kappa_t^+ \end{pmatrix} \sim \mathrm{NID}(0, \sigma_\kappa^2 I_2), \quad \begin{pmatrix} \psi_1 \\ \psi_1^+ \end{pmatrix} \sim \mathrm{NID}\left(0, \frac{\sigma_\kappa^2}{1 - \varphi_\psi^2} I_2\right),$$

where the disturbances κ_t and κ_t^+ are serially independent and mutually independent, also with respect to disturbances that are associated with other components. The coefficients φ_ψ, λ_c and σ_κ^2 are unknown and need to be estimated together with the other parameters. This stochastic cycle specification is further discussed by Harvey (1989).

Generally, the ENSO oscillation is said to have a period of between 2 and 7 years. As we already pointed out, seasonality has an important role in the evolution of the overall

irregular cycle (Tziperman et al. 1997; An and Choi 2009). At the same time, several other inter-annual spectral peaks at different time scales can be noted in Fig. 2a, showing the power density spectrum of the entire N3.4 time series. It will be demonstrated in the next sections that the modes of variability that correspond to these spectral peaks are key not only for understanding ENSO, but also for its simulation.

3.3.1 Near-annual cycle component, $\psi_1 t$

Previous studies have identified one significant signal at a frequency close to the annual cycle, recognized as a separate coupled mode of variability of the equatorial Pacific ocean-atmosphere system (Zebiak 1985; Neelin 1990; Mantua and Battisti 1995; Jiang et al. 1995; Fedorov and Philander 2001; Jin et al. 2003). Its relevance to the evolution of the ENSO cycle and its prediction has been demonstrated especially by Jin et al. (2003), where this mode is called a "near-annual" mode and defined to have a period of 12–18 months. It is characterized by a westward propagation of SST and zonal wind stress anomalies from the eastern equatorial Pacific (EPAC) to the WPAC (Fig. 3). These anomalies are similar to patterns of the mean annual cycle in the EPAC, but enhanced and propagating all the way to the western boundary (Mantua and Battisti 1995; Jin et al. 2003). Westward coupled modes of this kind, driven by near-surface advective processes (mostly anomalous zonal advection of mean temperature gradients), have also been identified in the tropical Pacific in the theoretical work of Philander et al. (1984); Gill (1985); Neelin (1991), and some others. In the present model we too found that the period of one of the cycle components estimated with the Kalman Filter is approximately 1.5 years (16–18 months). The graphical representation of this cycle is shown in the time series in Fig. 1c.

This near-annual mode has been shown to be modulated both by the annual cycle and ENSO. It has been hypothesized by Jiang et al. (1995) that it is produced through the nonlinear resonance of the low-frequency ENSO mode, which is discussed later, with the annual cycle. Then, its interaction with the main ENSO cycle and its phase-locking with the mean annual cycle result in fluctuations in its amplitude, which was low in the late 1980s, for example, and then increased in the second half of the 1990s (Jin et al. 2003; Fig. 1c). When its period is close to 12 months (i.e. to the annual cycle), phase- and frequency-locking likely occur and lead to the greater amplitude of the mode, and to near-annual LN-like events (as between 1996 and 2004, see Figs. 1c, 3d, e, i, j; Jin et al. 2003). However, the continued long warm background period taking place in the tropical Pacific during 1985–1991 (due to ENSO variability, Fig. 1e) has resulted in the lower frequency of this mode, reducing the opportunity for its non-linear phase-locking

with the main oceanic Rossby mode (Mantua and Battisti 1995), and leading to "mini El Niño events" that have occurred between all major EN events in that period (Jin et al. 2003; Figs. 1c, 3a, b, f, g).

Thus, this fast mode of variability has important implications for the evolution of ENSO, and hence, for the accurate prediction of EN and LN events. Therefore, its inclusion into the proposed prediction scheme could significantly improve the model performance and skill. In fact, it has been demonstrated by Mantua and Battisti (1995) that the irregularity of ENSO reproduced by the Zebiak–Cane (ZC) model (Zebiak 1985) is due to a large extent to the presence of a "mobile mode" with similar characteristic features as the near-annual mode discussed here.

3.3.2 Quasi-biannual cycle component, $\psi_2 t$

A number of studies have discussed a biannual/quasi-biannual (QB) (24–28 months period) component of ENSO variability (Trenberth 1976; Rasmusson and Carpenter 1982; Lau and Shen 1988; Yasunari 1989; Rasmusson et al. 1990; Jiang et al. 1995), which has been detected in both the equatorial surface zonal wind and SSTs. Rasmusson et al. (1990) discovered through an SVD analysis of the zonal winds in the tropics a biannual cycle characterized by consistent eastward-propagating anomalies across the WPAC and the CPAC. In the EPAC it was found to have a meridional propagation also affecting local SSTs. Moreover, a strong and close relationship between the biannual components of the zonal winds in the WPAC and the SST time series in the EPAC was demonstrated. From this analysis, the conclusion has been reached that all well-defined warm phases of ENSO coincide with a distinctive warm phase of the biannual mode.

Similarly to the fast-frequency component, the biannual oscillation also tends to be phase-locked to the annual cycle (Rasmusson et al. 1990). Typically associated with it are westerly wind anomalies in the WPAC (between 130° and 160°E) and north of the Equator in March–April of the first year of the cycle (Fig. 4f, g). These patterns progress eastward and peak by September–October near the date line, moving south of the Equator by that time, testifying for a meridional displacement (Fig. 4h, i). After October they propagate to the EPAC and begin to disappear. Then, by November easterlies start to form west of 130°E and north of the Equator. These develop in DJF and move towards the east (Fig. 4j). By March–April of the second year a reversal of the wind pattern is at place.

The evolution of SSTs in the biannual mode, on the other hand, develops from near neutral anomalies in March–April (Fig. 4a) to a typical ENSO-like structure in September–October of the first year of the cycle (Fig. 4b). In the second year, the SSTA pattern is reversed and the opposite phase of the ENSO-like structure develops (Fig. 4d, e). In

this way, equatorial negative (positive) SSTAs in the CPAC and EPAC are preceded by easterly (westerly) surface zonal wind anomalies in the WPAC and CPAC (Fig. 4).

Thus, the estimation of a period close to 2 years (between 26 and 30 months; Fig. 1d) for the second cycle component of the proposed model corresponds to this mode of variability of the equatorial Pacific ocean-atmosphere system. Importantly, the phase evolution and amplitude of this cycle apear to be approximately regular (see Fig. 1d). On the other hand, Rasmusson et al. (1990) has noted that even though the cycle is regular, the amplitude and phase of the mode are subject to low-frequency changes, which will be discussed later in this manuscript. In any case, the QB mode has been shown to be a fundamental element of the ENSO dynamics, essential for its prediction (Rasmusson et al. 1990; Jiang et al. 1993, 1995).

3.3.3 Low-frequency cycle component, $\psi_3 t$

The main period of ENSO has been established to be between 4 and 5 years, which, as expected, corresponds to the most pronounced spike in the power spectral density of the N3.4 time series (Fig. 2a). Moreover, as shown by Jiang et al. (1995), the dominant mode of variability of SSTs and surface zonal winds in the tropical Pacific is a low-frequency oscillation with a period of approximately 46–63 months. It has been referred to in the literature as the low-frequency ENSO mode (Rasmusson et al. 1990), or as the quasi-quadrennial (QQ) mode of variability (Jiang et al. 1995).

For SST the QQ component represents an eastward propagating oscillation with maximum anomalies close to 120°W (Fig. 5b, e). For the zonal wind it represents an oscillation with an eastward propagation in the WPAC and with maximum anomalies close to 170°E (Fig. 5f, i). Actually, its spatial characteristics for both SST and zonal wind closely resemble the main spatial characteristics of the biannual mode (Figs. 4, 5). Most El Niño events have been found to correspond to the warm phases of these two modes (Jiang et al. 1995), a few correspond to only that of the QQ mode, while warm years in the time series that were not termed as warm events officially have been shown to correspond to a warm phase of the biannual cycle alone, but out-of-phase low frequency component (Rasmusson et al. 1990).

Thus, the third cycle included in the model coincides with this main oscillatory mode, and its estimated period is about 52–56 months. Unlike the biannual cycle, its phase and amplitude are clearly irregular and time-varying (Fig. 1e).

The robustness of the analysis about the significance of these modes of variability holds both in the case of using raw data or anomalies (Jiang et al. 1995). Our results confirm this, as the estimated frequencies of the three cycle components were the same regardless of filtering the mean annual cycle (not shown). As can be seen in Fig. 2, they are associated with the main frequency signals in the spectrum of the N3.4 time series (also see Fig. 1). We then assume that these are fundamental time scales of the ENSO phenomenon, and that its irregularity could be to a large degree explained in terms of the relative phase positions and amplitudes of the three cycle components described here.

3.4 Regression variables, $x_t' \delta$

Additionally, we have included a set of explanatory variables in the model for capturing specific dynamic variations in the time series, not explained by the components discussed above. Such variables could also be used to allow for outliers and breaks in the model, or to account for part of the seasonality of the dependent variable in case it is not fully captured by the seasonal component. In fact, it has been proposed that there is a partial 2:1 phase synchronization of ENSO to the annual cycle, and that El Niño is typically associated with a weaker annual cycle (Stein et al. 2011). Since the seasonal component here has been estimated as fixed, i.e. with constant amplitude and periodicity (Sect. 3.2; Fig. 1b), we incorporate regression variables that not only aim to explain more variability in the N3.4 time series, but also to model more subtle seasonal fluctuations.

Thus, we extend the decomposition with a multiple regression effect $x_t' \delta$ for $t = 1, \ldots, n$ (grey time series in Fig. 1a), where x_t is a $K \times 1$ vector of predetermined covariates and δ is a $K \times 1$ vector of regression coefficients. Elements of δ could also be allowed to change over time. However, as we want to establish stable relationships between the dependent variable and a set of explanatory variables, δ is kept constant for the full sample length or, for at least, a large part of the sample. Therefore, for the present study we have kept the coefficients δ fixed in time, and we will explore the time-varying case in a future study.

In very general terms, ENSO represents the alternation of heat buildup and release from the equatorial region (Wyrtki 1985; Zebiak 1989; Jin 1997a). The forcing of the zonal wind causes fluctuations in this heat content through Kelvin and Rossby wave dynamics, and SSTs are affected by these waves through downwelling and upwelling processes (Ballester et al. 2015). SST anomalies, on the other hand, influence the atmosphere and result in anomalies in the zonal wind patterns—the so-called Bjerknes feedback (Bjerknes 1969). Therefore, we can say that ENSO could be deemed predictable due to the intrinsic ocean variability, as well as the deterministic wind-driven ocean dynamics and the resulting equatorial heat content redistribution (McPhaden 2004). Based on this assumption, we use zonal

equatorial wind stress, and surface and subsurface ocean temperature as regression variables in our model. It should be noted that some of these covariates can be highly correlated between each other, and when including a number of them, the problem of co-linearity can occur. In this situation, the estimation may be based on inversions of matrices that are close to singularity (reduced rank matrices). In our approach for model configuration we account for co-linearity by an appropriate selection strategy that we discuss in the next Sect. 4.

3.4.1 Equatorial wind stress

Tropical wind stress is among the main drivers of surface and subsurface temperature anomalies in the equatorial Pacific. Wyrtki (1975) was the first to hypothesize that El Niño is preceded by abnormally strong southeasterly trade winds in the CPAC, more precisely between 180°E and 220° E, starting approximately 2 years before the event (Fig. 6a). Moreover, such anomalies in the zonal wind component represent one of the factors preceding the occurrence of an EN that has an essential role in the triggering of its whole mechanism (Wyrtki 1975; Jin 1997a; Ballester et al. 2015). Thus, following our analysis (Fig. 6), as well as previous studies (Wyrtki 1975; McPhaden 2004; Sarachik and Cane 2010), we have chosen two regions in the CPAC to calculate zonal wind stress indices and obtain time series to be used as regression variables in the proposed model (Fig. 6a, b; Regions I and II in Table 1). It is interesting to note that one of these regions (Region II) is located entirely in the Southern Hemisphere, as this is the area of maximum easterly zonal wind anomalies at about 11–14 months before the boreal peak of EN (Fig. 6b). This is also in agreement with the results of McGregor et al. (2012), where it is shown that the zonal wind anomalies suddenly shift southward at the end of the calendar year.

The strengthening of the westward surface trade winds, on the other hand, increases the intensity of the westward circulation of the South Equatorial Current (SEC; Yu and McPhaden 1999), which deepens the thermocline in the WPAC (Sarachik and Cane 2010). This process piles up warm water at those latitudes and accumulates heat, which often leads to the spreading of the warm pool toward the CPAC (both on the surface and in the subsurface). The warming induces an area of anomalous low surface pressure, over which westerly wind anomalies tend to appear at the Equator near the date line (Fig. 6c; McPhaden 2004).

In fact, SST-modulated westerly wind anomalies are critical for the ocean-atmosphere coupling before a warm event (Eisenman et al. 2005), and have occurred prior to every significant El Niño in the last several decades (Fig. 6c; McPhaden 2004). As mentioned earlier (Sect. 3.3.2) and shown by other studies (Tziperman and Yu 2007; Gebbie and Tziperman 2009), there is a deterministic portion of these anomalies, which is in interplay with the biennial ENSO cycle of SSTs, and could be used as a precursor for it (Gebbie and Tziperman 2009). Therefore, another critical region for the model is located over the easternmost part of the WPAC and over the CPAC (see Fig. 6c; Region III in Table 1). Over this region we average zonal wind stress to obtain a time series used as a regression variable at short lead times (5–8 months before the typical December peak).

3.4.2 Equatorial Pacific surface and subsurface temperatures

The western Pacific plays a fundamental role in the oscillatory behaviour of ENSO and in the generation of warm and cold phases. It is where the warm pool develops—an upper-ocean area of very warm and well-mixed waters, surface horizontal current convergence and subsurface divergence, and therefore downwelling motion (Brown and Fedorov 2010; Ballester et al. 2015). As mentioned earlier and shown by Wyrtki (1975), and Jin (1997a, b), the abnormally strong trade winds increase the east–west slope of sea level in the CPAC, intensify the westward SEC, and accumulate warm water (heat buildup) in the surface and subsurface WPAC region (Ballester et al. 2016). Right after a relaxation of the trades, this stored warm water would be allowed to move eastward through the Equatorial Undercurrent (EUC) and the North Equatorial Countercurrent (NECC) in a dynamically consistent way (Jin 1997a; Ballester et al. 2016), and to lead to an EN event in the EPAC (Wyrtki 1985; Ramesh and Murtugudde 2013).

Substantial heat content increase along the equator has preceded all El Niños since 1980 by at least two seasons, and it is a necessary condition for its occurrence (McPhaden 2004). In addition, it has been shown by Ramesh and Murtugudde (2013) that regardless of the regime shifts in the tropics, the buildup of warm water in the WPAC (Fig. 7) and the onset of its eastward displacement has remained unchanged, always starting at a particular stage of the cycle—about 18–20 months before the peak (Fig. 8), when the heat buildup in the WPAC has

Table 1 Regions over which zonal wind stress is averaged to obtain time series used as predictor regression variables in the model

Regression variable	Region I	Region II	Region III
Zonal wind stress	$[180e - 220e] \times [4s - 4n]$	$[180e - 210e] \times [10s - 0]$	$[160e - 200e] \times [0 - 10n]$

grown substantially (Fig. 7). This is also the time when the subsurface anomalies begin to appear on the surface in the far western Pacific (Fig. 8a), and start to weaken the trades (see previous Sect. 3.4.1).

In this way, a variable that accounts for this heat buildup and the onset of its eastward migration could be used as a precursor for EN. Based on this mechanism we defined two sets of regions over which to average surface and subsurface temperatures in the WPAC/CPAC and obtain time series indices for regression variables (Table 2). The first set of regions corresponds to the early subsurface warming in the WPAC (between latitudes 12°S and 8°N; Fig. 7; Table 2). As can be seen in Fig. 7a–c anomalous warming of the subsurface exists on average as early as 25–30 months prior to the typical peak of EN in boreal winter, at depths between 150 and 400 m.

The second set of regions corresponds to the onset of the advection of warm water eastward along and below the thermocline, and the subsidence of the trades on the surface (Figs. 6, 8). During this stage the warm pool extends towards the CPAC in the subsurface (100–300 m depth) through the strengthened EUC (Ballester et al. 2016), therefore, this set of regions is located closer to the CPAC (see Fig. 8; Table 2). Warm anomalies start to appear in these regions at approximately 25 months or even before (Fig. 8b, c), intensify and stay there until about 8–9 months prior to the peak of EN, when they are further advected eastwards to enter the EPAC (Fig. 8a). At the same time, cold anomalies start to develop in the WPAC region instead, corresponding to the onset and rapid development of an EN event (Figs. 7, 8). This coincides with the occurrence of westerly wind anomalies as discussed earlier, which trigger upwelling Rossby waves and decrease the subsurface temperature in the WPAC due to shoaling of the thermocline there (McPhaden 2004). Thus, the temperature indices averaged over the first set of regions are also used

as regression variables during the mature phase, when there are strong cold anomalies in the WPAC (lead/lag times 0–9, Fig. 7a–c; Yu and Mechoso 2001).

Equatorward subsurface (100 m depth and below) advection from the northern off-equatorial region at this time (5–9 months before the EN peak) due to anomalous upwelling between 5°N and 10°N and downwelling between 3°S and 1°S, also results into an area of strong cold anomalies in the box [140°E–210°E] × [5°N–10°N] (Supplementary Figure 1). This inter-hemispheric asymmetry has also been found by McGregor et al. (2012), where it is revealed in the EOF2 of the integrated upper ocean heat content and shown to lead the typical EN pattern (EOF1) by about 8–9 months. These vertical velocity anomalies and inter-hemispheric asymmetry result from the latitudinal distribution of zonal (Yu and Mechoso 2001) and meridional (McGregor et al. 2012) wind stress anomalies. Hence, we calculate a temperature time series from this box and use it as a regression variable in the model at the corresponding significant lead/lag times to further account for the dynamical processes happening before an El Niño, and improve the overall forecast skill of the model.

4 Model configuration

Based on the recharge-discharge oscillatory theory (Jin 1997a), and as discussed previously, the buildup of heat at the equatorial Pacific is a prerequisite for El Niño, and the event itself discharges this excess heat poleward. In addition, the time between two consecutive events could generally be determined by the time needed for the tropical ocean to accumulate enough heat, and normally the amplitude of the event is in a direct proportion to the magnitude of the excessive heat content (McPhaden 2004). The component configuration of our structural model is then synchronized

Table 2 Regions over which surface and subsurface temperatures are averaged to obtain time series used as predictor regression variables in the model

Regression variable	Region I	Region II
SST WPAC	$[140e - 160e] \times [5s - 5n]$	
SST WPAC2	$[140e - 180e] \times [10s - 5n]$	
SST WPAC3	$[120e - 170e] \times [10s - 5n]$	
SST WPAC4	$[140e - 160e] \times [10s - 0]$	
Temperature at 50 m depth	$[120e - 170e] \times [10s - 5n]$	
Temperature at 100 m depth "cold"	$[140e - 210e] \times [5n - 10n]$	
Temperature at 100 m depth	$[120e - 140e] \times [10s - 5n]$	$[150e - 180e] \times [7s - 7n]$
Temperature at 150 m depth	$[120e - 140e] \times [10s - 5n]$	$[150e - 180e] \times [7s - 7n]$
Temperature at 200 m depth	$[120e - 140e] \times [10s - 7n]$	$[150e - 180e] \times [7s - 7n]$
Temperature at 250 m depth	$[120e - 140e] \times [7s - 7n]$	$[140e - 170e] \times [7s - 7n]$
Temperature at 300 m depth	$[120e - 140e] \times [7s - 7n]$	$[160e - 200e] \times [10s - 3n]$
Temperature at 400 m depth	$[120e - 140e] \times [5s - 5n]$	$[150e - 170e] \times [10s - 0]$
Temperature at 500 m depth	$[120e - 140e] \times [5s - 5n]$	$[150e - 170e] \times [10s - 0]$

with these assumptions about the warm buildup in the WPAC and its migration eastwards.

It has been shown that the correct representation of the feedback between SST and westerly winds has a high impact on the dynamical regime of ENSO (Gebbie and Tziperman 2009). The correct inclusion of the SSTs, subsurface temperature, and zonal wind stress time series at the respective lead/lag times of importance of these variables could lead to a more accurate forecast. In this way, using the dynamics of EN described earlier (see Sects. 3.4.1, 3.4.2; Figs. 6, 7, 8), as well as rigorous statistical indicators for goodness of fit (see Harvey 1989 for details), such as the prediction error variance (pev), the prediction error mean deviation (pemd), the information criteria Akaike (AIC) and Bayesian Schwartz (BIC), and the coefficient of determination (Rs^2) (see Tables 3, 4, 5, 6 in Appendix 2), we have configured the model in such a way that the different regression variables are only used at their respective monthly lead/lag times of dynamical relevance (see Tables 7, 8 in Appendix 2). These lead/lag times have also been determined based on the estimated coefficients and p values of the explanatory covariates (Tables 3, 4, 5, 6 in Appendix 2; Harvey 1989). The core of the model described in Sect. 3 is kept constant, but the regression variables given by $x'_t\delta$ are varied depending on the time before the peak when a forecast is started. The selection of regression variables also depends on the amount of co-linearity that is encountered in the regression. When several explanatory covariates have been selected for one lead/lag time during the fitting procedure (and they are all significant), we only use one or two of them when we perform the actual forecasts, only the ones that are the most significant for the particular forecasting event. In this way we avoid co-linearity in the estimation and forecasting procedures.

As seen in Tables 7 and 8 in Appendix 2, for every lead/lag time the most statistically relevant predictors (zonal wind stress or temperature at specific depths in Regions I, II or III) are added to the model. These variables are associated with low p values when fitted at these specific lead/lag times, and the overall coefficient of determination of the model R^2, as well as the coefficient of determination based on differences around seasonal means Rs^2, increase when these particular variables are included (see Tables 3, 4, 5 and 6 in Appendix 2). Additionally, the main anomalies of the covariates at the respective lead/lag times are also taken into consideration (the composite anomalies shown in Figs. 6, 7, 8) when determining the model configuration. In summary, depending on the relative time before the potential peak of EN that a forecast is started, different regression variables are added to the model, in order to enhance its forecasting performance.

5 Forecasting results

In order to test and cross-validate the proposed prediction scheme we attempted to forecast all El Niños that occurred in the period 1996–2015 at long (20–34 months), medium (10–19 months), and short (3–9 months) lead times (Figs. 9, 10). EN events were chosen according to the classification of the Climate Prediction Center for the period 1996–2015: December 1997, 2002, 2006, 2009, 2014 (CPC 2015). As a final test, we proceeded with an experiment, in which we predicted the whole ENSO time series in the period 1982–2014 (Fig. 10).

5.1 Forecasts of individual El Niño events

Long-lead (20–34 months in advance; magenta, light blue, dark green and beige curves in Fig. 9) forecasts of the five events are shown in Fig. 9a–e. Some shifts and mismatches in the exact onset and magnitude of the events can be noticed, but their occurrences are properly forecast, and an indication for a developing El Niño is clearly present even at these very long lead times. The 21 months ahead predictions (beige curves in Fig. 9a–e) already represent the events very well in terms of timing and amplitude. It is important to note that at 26 months lead the peak of the 1997/98 EN is predicted to reach an anomaly of $+2°$C (dark green curve in Fig. 9a and blue curve in Supplementary Figure 2; Table 9), and that an event of extreme magnitude is foreseen more than 2 years in advance, long before the appearance of a series of westerly wind bursts in early 1997, assumed by some studies to be the prime reason for the strength of this El Niño (McPhaden and Yu 1999). This is in support of the theory proposed by Chen et al. (2004) that the evolution especially of the big events is to a greater extent determined by the initial condition of the ocean, and the atmospheric forcing is rather a secondary modulator or consequence. In addition, the 2014/15 weak EN is forecast with high accuracy 24 months ahead (dark green curve in Fig. 9e, black curve in Supplementary Figure 2; Table 9 in Appendix 2), while a large number of the operational models failed and predicted an extreme event similar in size to the 1997/98 one (McPhaden et al. 2014; Glantz 2015).

Medium-range (10–19 months in advance; red, blue and green curves in Fig. 9) forecasts of the same events are depicted in Fig. 9f–j. It is evident that for some of the events the forecast improves as the lead time becomes shorter, as in the case of the 2002/03 and 2009/10 El Niños (Fig. 9g, i). For the 1997/98, the 2006/07 and the 2014/15 events the skill of the model remains similar (Fig. 9f, h, e). Again, we want to highlight the accurate prediction of the 2014 EN at 16–17 months lead time (blue curve in Fig. 9j;
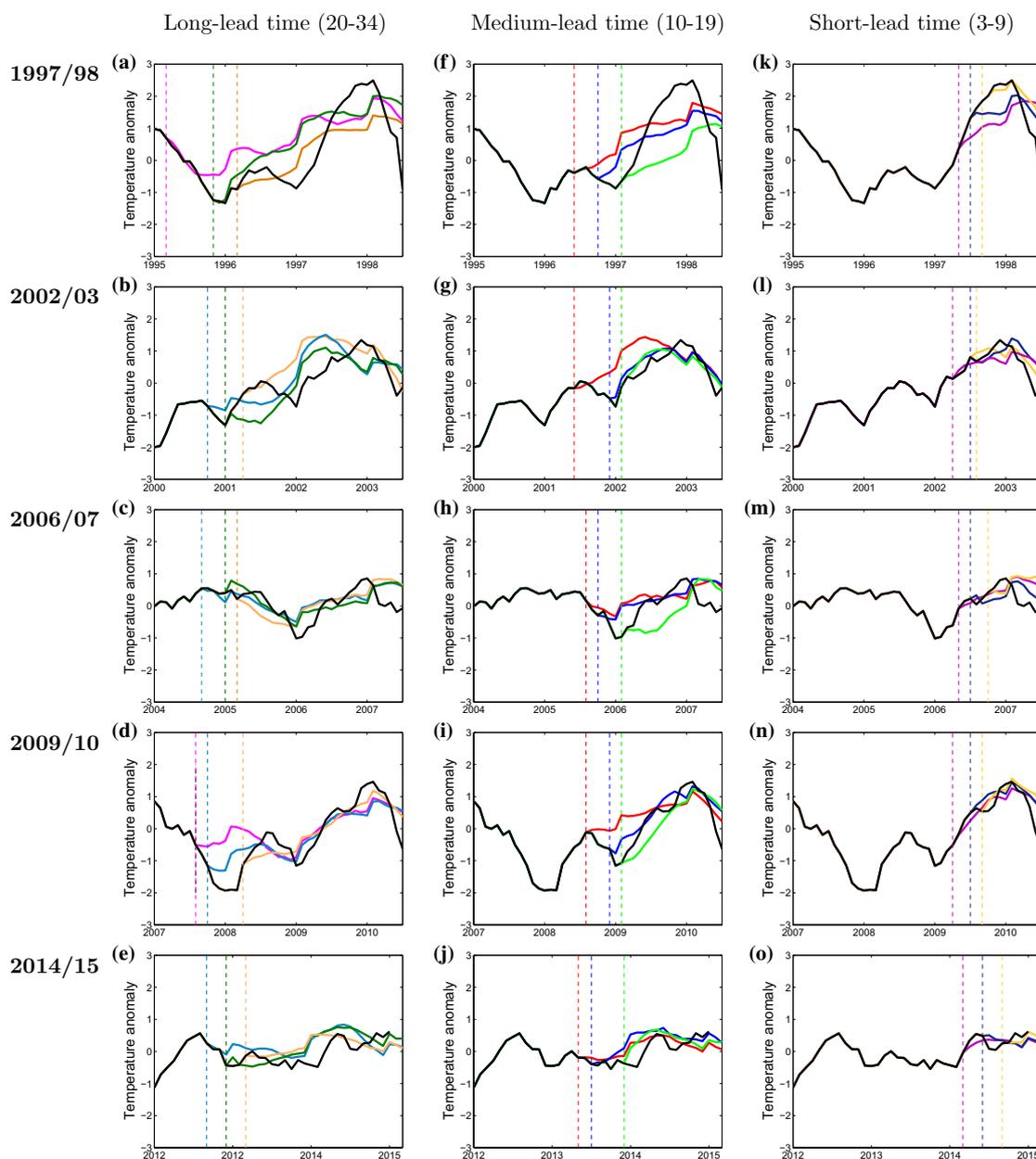
Fig. 9 Time series of area-averaged sea surface temperature (°C) anomalies in the Niño 3.4 region. Shown are forecasts of the **a**, **f**, **k** 1997/98, **b**, **g**, **l** 2002/03, **c**, **h**, **m** 2006/07, **d**, **i**, **n** 2009/10, and **e**, **j**, **o** 2014/15 EN events, starting 29–34 (*magenta* in **a** and **d**), 27–28 (*light blue* in **b**–**e**), 24–26 (*dark green* in **a**–**c** and **e**), 21–22 (*beige* in **a**–**e**), 17–19 (*red* in **f**–**j**), 13–16 (*blue* in **f**–**j**), 11–12 (*green* in **f**–**j**), 8–9 (*velvet* in **k**–**o**), 6 (*dark blue* in **k**–**o**), and 3–5 (*dark yellow* in **k**–**o**) months before the peak of El Niño, respectively. *Vertical dotted lines* indicate the month in which the respective forecasts are started. Observations are in *black*

(a)
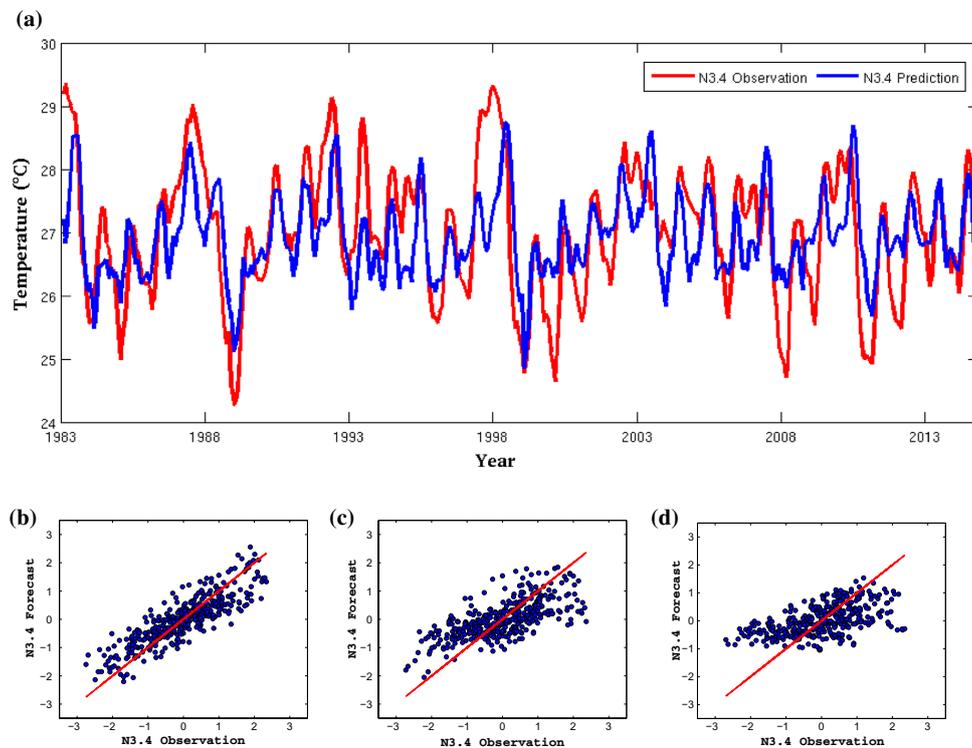
Fig. 10 a Retrospective forecast of the EN3.4 time series in the period 1983–2014. The EN3.4 observation is in *red* and the model prediction at 6 months lead time is in *blue*. Scatterplots of the EN3.4 time series observation against forecast at **b** 3, **c** 6, **d** 18 months lead time. The respective regression coefficients are 0.70, 0.45 and 0.30

Table 9), and the consistency of the forecast indicating a minor event (black curve in Supplementary Figure 2) instead of a major one.

As expected, the overall performance of the model is improved at short lead times (3–6 months; dark blue and dark yellow curves in Fig. 9k–o). The peak of the temperature anomaly is already accurately forecast in all five cases (Table 9; Supplementary Figure 2). An interesting feature is that most events (the 1997/98, the 2002/03, and the 2009/10) are better forecast 12–14 months in advance than 9–11 months ahead (velvet curves in Fig. 9k–o, blue, red and yellow curves in Supplementary Figure 2; Table 9). We relate this issue to the fact that 8–9 months before a typical boreal winter peak is the time of the spring barrier, when the signal to noise-ratio is lower and when most models tend to lose their predictability skills (Barnston et al. 2012). This result confirms that some El Niños are more reliably forecast at medium- and long-lead times (Izumo et al. 2014), rather than at shorter ones, something

that could seem counterintuitive and calls for further investigation.

Table 9 and Supplementary Figure 2 summarize the results for the prediction of the target month of January for the five EN cases. Roughly one year and a half to two years in advance the forecasts successfully indicate whether the respective event is expected to be strong, moderate or weak. Moreover, for some lead months the temperature anomaly value of the prediction is almost exactly the same as the value in the observations—for instance 12–14 and 21–23 months ahead of the 2002/03 event, 27 months ahead of the 2006/07 event, and 6–8 and 12–14 months ahead of the 2009/10 event.

Finally, a feature to be noted is that even at short lead times the the evolution of the temperature anomaly during some of the events, for instance the 2002/03 and the 2006/07, appears slightly lagged by 1–2 months in the model than in the observations (Fig. 9b, g, l, c, h, m, respectively). This delay could be linked to the fact that in

the current configuration of the model the seasonal cycle has been estimated as a non-varying fixed cycle with a seasonal peak in the months of April–May (Fig. 1b). Even though in general terms this representation is correct, we know that in reality the seasonality varies especially in EN years. In the present scheme we attempted to model this additional variability through the regression explanatory covariates and the near-annual cycle component, however, this issue should be better addressed in the future. Thus, one avenue for improvement would be to test the model with a time-varying seasonal component, which we expect to result in an even more accurate evolution and timing of the predicted events.

So far it has been shown how the proposed model is highly skillful in forecasting the warm phase of ENSO, of those El Niños that occurred in the last 20 years. In the next subsection we demonstrate its potential to predict the overall phenomenon, including neutral years.

5.2 ENSO time series forecast

The proposed forecasting scheme has been especially targeted at the prediction of both the timing and magnitude of El Niño events, and in the previous subsection its skill in that respect was shown. In the present study the regression covariates used in the model and described earlier (Sect. 3.4) have been incorporated with the aim to improve the forecast of the warm phase of the oscillation, without considering the cold LN phase. Indeed, the rest of the components of the model are formulated in such a way as to address the prediction of the entire ENSO phenomenon. Therefore, it is expected that the model is also skillful at predicting neutral and LN years, albeit with less precision.

In Fig. 10a a forecast across the whole interval spanning 1983–2014 at 6 months lead time can be seen. The data between 1952–1982 was used for the estimation of the model components, and then prediction was attempted for the remaining period. This forecast is more complex than the ones presented in the previous subsection when components were estimated on the basis of the available observations prior to every warm event. Instead here, general conditions for the fitting period are applied, and not specific ones related to each ENSO event. In addition, most of this fitting interval covers a time before the ENSO regime shift in 1977 (Hare and Mantua 2000), while the forecast is made for the years after this shift. As expected the model is capable of predicting all EN events and neutral years in the period, but some of the LN events are missed, for example the cold ones occurring in the years 2000 and 2008.

From the scatterplots of the observed Niño 3.4 index against forecasts at lead times of 3, 6 and 18 months (Fig. 10b–d) it becomes clear that the model has

correctly predicted the LN events of extreme amplitude at a lead time of 3 months (Fig. 10b). With its current formulation, however, the cold anomalies exceeding –2 °C cannot be captured at a lead time of 6 months (Fig. 10c), while at the longer lead time of 18 months anomalies exceeding –1 °C are missed (Fig. 10d). Furthermore, the positive skewness coefficients of 0.43 and 0.45 for the forecast time series in Fig. 10c and d, respectively, indicate an asymmetry of the probability distribution functions with longer tail toward the warm ENSO events. In reality, the rapid termination of El Niño after boreal winter is on average followed by the development of La Niña in the following year (Fig. 1a; Ballester et al. 2015). In these cases the ocean dynamics follows a different pattern than a symmetric pattern (with opposite sign) to the one that precedes El Niño. The eastward propagation of the subsurface cold anomalies and their surface appearance in the EPAC occur faster, for example. More aspects of the asymmetry between the two ENSO phases have been discussed in detail by Kang and Kug (2002), An and Jin (2004), An and Choi (2009), Yu et al. (2011) and Ballester et al. (2015). Therefore, the current predictor regression variables of the model cannot properly signal the development of some of the LN events. Hence, specific regression covariates related to the dynamics and playing a role during and before La Niña will be added to the modeling scheme in the future with the objective to improve the representation of the cold ENSO phase.

In Table 10 the correlation between forecasts and observations, as well as root mean square error (RMSE) are given as functions of lead time for a number of lead times and for all seasons combined. These are indicative of the high skill of the proposed model to predict the Niño3.4 index. For lead times of 3 and 6 months the correlation is 0.86 and 0.68, respectively, which compares to the best two operational dynamical models described in Barnston et al. (2012) (see Figure 6 in Barnston et al. 2012). The RMSE for the same lead times are 0.54 and 0.77, respectively, and thus for this particular measure of skill and lead times the model outperforms all other operational models (see Figure 9 in Barnston et al. 2012). A more detailed and extensive comparison with existing ENSO models will be provided in a follow-up study.

6 Discussion

The results presented in the previous section suggest that the structural time series model (Sect. 3) described in this study outperforms the statistical ENSO models included in the cross-validation exercise, and it is comparable to some of the most accurate dynamical models. Our prediction

scheme successfully goes beyond the spring barrier, and also through the one after that, which is an implication for the much longer predictability limit of ENSO than generally accepted. This result, to our knowledge, has only been accomplished by Chen et al. (2004) so far. Still, we report forecasts started two and a half years before the boreal EN peak, which is even longer than the lead time of the predictions shown therein. Also, in Chen et al. (2004) a complete account of the reasons underlying this performance is not provided, while here a dynamical understanding of the EN-associated processes accompanies the definition of our model structure and regional covariates.

Additionally, we want to highlight the simplicity of the proposed prediction scheme, as it uses readily available data directly as input—monthly values of SSTs, subsurface ocean temperature in the vicinity of the equatorial Pacific, and zonal wind stress; and it does not require extensive computations. Therefore, our model appears to have a clear advantage over the more complicated statistical schemes, as well as over the computationally more expensive dynamical models. We would also like to discuss one important advantage that is especially due to the state space approach applied here.

Two of the forecast events in Fig. 9 were unusual in the respect that the frequencies of the three cycle components estimated with the Kalman Filter were different. In the case of the 2006/07 event (Supplementary Figure 3), the quasi-biannual and quasi-quadrennial cycles were preserved, but the third cycle was estimated to have a frequency associated with decadal scales of variability [Supplementary Figure 3(c)]. Actually, when the model was fixed as described in Sect. 3 and shown in Fig. 1, the event was shifted and forecast for the next year 2007/08 (not shown). This confirms the important role of the decadal tropical Pacific variability in the modulation of the ENSO amplitude, which was reported by previous studies (Kleeman et al. 1999; Yeh and Kirtman 2004, 2005). As the model in its current configuration has only three cycle components, we cannot deduce whether the near-annual cycle component still has a vital contribution to the evolution of this event.

In Supplementary Figure 3(a), however, it can be seen that the estimated 2-year oscillation is not regular as in Fig. 1d. We have plotted together the cycles from Fig. 1c and d (the near-annual and the biannual oscillations), and we superimposed the irregular quasi-biannual component from Supplementary Figure 3(a) (see Supplementary Figure 4). The irregular 2-year cycle appears to result from the convolution of the other two cycles. This could explain the occasional shift in the amplitude and phase of the otherwise regular biannual cycle, which was mentioned earlier

(see Sect. 3.3.2). In fact, it has been difficult to clearly isolate the peaks in the spectrum of both zonal wind and SSTs associated with the oscillations at periods of 16–18 and 24–28 months in the analysis done by Rasmusson et al. (1990) and Jiang et al. (1995), favouring this kind of conclusion. In any case, a model configuration in which both cycles are present in an explicit way would most probably result into an even more precise forecast of the 2006/07 El Niño, but this is not in the scope of the present article.

The other event, for which the frequencies of the cycle components were estimated differently was the prominent 1997/98 event. In the case of this El Niño, which is also the most extreme on record, the quasi-biannual cycle was calculated as for the 2006/07 event [see Supplementary Figures 3(a) and 5(a)], while the other two cycle frequencies corresponded to periods of 43–45 and 62–65 months [Supplementary Figure 5(b), (c)]. In this way, the event is characterized by a "double" or "split" quasi-quadrennial cycle. In order to shed more light onto this distinction, we examined these two cycles in more detail. In Supplementary Figure 5 we can clearly see that the cycle with 65 months period has its peak in 1992/93 [Supplementary Figure 5(c)], while the one with 45 months peaks in 1991 [Supplementary Figure 5(b)]. On the one hand, in 1991/92 there was an official El Niño event. On the other hand, the year 1993 was indeed exceptionally warm (Fig. 1a). These two cycles then peak coincidentally in 1998, which preconditions an event of higher magnitude. It should also be taken into consideration that the prediction of the 1997/98 event is made by fitting the data between 1982 and 1995, which is a short period of time when only one event of similar, but still lower magnitude, occurred. Thus, the model has performed very well by forecasting this exceptional event.

These preliminary results lead to the conclusion that at least some of the bigger and more extreme EN events occur as a result of the superposition of these two cycles in combination with the quasi-biannual cycle, the near-annual cycle and/or a cycle on decadal timescales. A recent study (Chen et al. 2015) classified EN events into three different categories—extremely strong ones, weak warm events reminiscent of the "warm pool El Niño", and the canonical ones with moderate warming in the central/eastern equatorial Pacific. Thus, taking into account this classification, we presume that on average for the canonical events, the cycles that have an important role in EN development are the ones discussed in Sect. 3 and shown in Fig. 1. On the other hand, for the more specific smaller events, and for the very extreme ones, the dynamical mechanisms seem to be complicated by the effect of additional cycles superimposed over these basic ones.

Considering the results presented here, we believe that the parameter re-estimation of the various model components, and especially of the three cycle frequencies, based on the observational data prior to each event, is crucial for the accurate forecasting of El Niño. Hence, the long-lead time prediction capabilities of our scheme come from the correct representation of the ocean-atmosphere system at each stage of its evolution. The fact that the state space form and the Kalman Filter allow for the unknown parameter re-estimation whenever an updated information about the coupled system becomes available is a fundamental advantage of our model.

7 Concluding remarks and future directions

A new ENSO forecasting scheme has been developed, based on structural unobserved components time series modeling with a state space approach. The unknown parameters of the unobserved components are re-estimated before every event, thus rendering the unique flexibility of the model. Regression variables especially designed to account for ocean surface and subsurface anomalous processes at the equatorial Pacific prior to individual El Niño events are incorporated at their specific time of relevancy. Since all unknown parameters are calculated together in a dynamic way, the addition of these covariates also affects the estimation of all the other model components in an efficient way, helping for the correct update of the whole system.

The model has been tested and cross-validated through the retrospective forecasts of the El Niños events in the recent period 1996–2015, and it has successfully predicted all the events that occurred at both long lead times (2 years and a half in advance), as well as at shorter lead times before and after the spring predictability barrier. Essentially, the model has accurately forecast a moderate EN for the end of 2014/beginning of 2015, unlike the majority of the operational models that warned of an event of substantial magnitude similar to that of the 1997/98 one.

A direct conclusion of high relevance is that the predictability limit of ENSO could be extended to at least two and a half years before the El Niño peak. Our study also clearly demonstrates that there is still room for improvement in the prediction of ENSO, not only in the forecasting of the cold phase of the oscillation, but also in that of EN itself. One immediate avenue for future work should be the inclusion of explanatory covariates especially aimed at improving the prediction of LN events, by taking into account the La Niña specific dynamics. Another area for improvement would be the transformation of the modeling scheme, so that all the important cycles discussed here - the additional quasi-quadrennial cycle, the decadal cycle, as well as independent near-annual and biannual cycles, could be separately incorporated in it. Finally, special attention should be paid to the incorporation of seasonality as discussed in Sect. 5, by changing the seasonal component of the model from being fixed to a time-varying one.

Appendix 1

The linear Gaussian state space model form that we have used is as in de Jong (1991), that is:

$$y_t = Z_t \alpha_t + G_t \epsilon_t, \quad \alpha_{t+1} = T_t \alpha_t + H_t \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, I) \tag{6}$$

for $t = 1, \ldots, n$, and where ϵ_t is a vector of serially independent disturbance series. The $m \times 1$ state vector α_t contains the unobserved components and their associated variables.

$$\alpha_t = \left(\mu_t, \gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-10}, \psi_{1t}, \psi_{1t}^+, \psi_{2t}, \psi_{2t}^+, \psi_{3t}, \psi_{3t}^+, \delta'\right)',$$
$$\epsilon_t = \left(\varepsilon_t, \eta_t, \omega_t, \kappa_{1t}, \kappa_{1t}^+, \kappa_{2t}, \kappa_{2t}^+, \kappa_{3t}, \kappa_{3t}^+\right)',$$

The measurement equation is the first equation in (6) and it relates the observation y_t to the state vector α_t through the signal $Z_t \alpha_t$. The transition equation is the second equation in (6) and it is used to formulate the dynamic processes of the unobserved components in a companion form. The deterministic matrices T_t, Z_t, H_t and G_t, are time-invariant except the matrix Z_t, and referred to as system matrices that are sparse and known:

$$
T_t = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p(1) & q(1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -q(1) & p(1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & p(2) & q(2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -q(2) & p(2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p(3) & q(3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -q(3) & p(3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p(4) & q(4) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -q(4) & p(4) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p(5) & q(5) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q(5) & p(5) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c(1) & s(1) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -s(1) & c(1) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c(2) & s(2) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -s(2) & c(2) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c(3) & s(3) & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -s(3) & c(3) & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_k
\end{bmatrix},
$$

$$
H_t = \begin{bmatrix}
0 & \sigma_\eta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma_\omega & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma_{\kappa,1} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma_{\kappa,1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma_{\kappa,2} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\kappa,2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\kappa,3} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\kappa,3} \\
0_k & 0_k & 0_k & 0_k & 0_k & 0_k & 0_k & 0_k & 0_k
\end{bmatrix},
$$

$$
Z_t = \left(1\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ x_t' \right),
$$

$$
G_t = \left(\sigma_\varepsilon\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \right),
$$

where 0_k is a $k \times 1$ vector of zeros, I_k is a $k \times 1$ vector of ones, $p(i) = \cos \lambda_i$ and $q(i) = \sin \lambda_i$ for $\lambda_i = \frac{2\pi i}{S}$, $i = 1, 2, \ldots, \lfloor \frac{S}{2} \rfloor$; $c(j) = \varphi_{\psi,j} \cos \lambda_{c,j}$ and $s(j) = \varphi_{\psi,j} \sin \lambda_{c,j}$ for $j = 1, 2, 3$.

Appendix 2

Every explanatory regression variable that has been used in the analysis (surface temperature, subsurface temperature at different depth levels and regions, and zonal wind stress at different regions) has been tested separately with the model described in Sect. 3 during the fitting procedure. In this way each variable was fitted at every lag time between 0 and 35 months. Based on the in-sample estimations for each of these fittings (the sample spanned the data from January 1982 to December 2012), the p value (an indicator of statistical significance, showing the probability that the coefficient of the predictor is equal to zero, and thus that the predictor does not add value), the significance level (SL) based on it (SL = 90 % for $p \leq 0.10$, SL = 95 % for $p \leq 0.05$, SL = 99 % for $p \leq 0.01$), and the Rs^2 value (modified coefficient of determination based on seasonal means—the ratio between the variance explained by the model and the variance of the seasonally differenced time series), were used to determine significance of the regression variable at the respective lag time. The significant values obtained in this way are summarized in the following Tables 3, 4, 5 and 6.

D. Petrova et al.

Table 3 Diagnostics of wind stress predictor regression variables

Regression variable	Time series region	Lag (months)	Coefficient	p value (SL)	Rs^2
Zonal wind stress	Region I	7	−2.80	0.07 (90 %)	0.3220
		8	−3.91	0.01 (99 %)	0.3110
		9	3.87	0.01 (99 %)	0.3125
		16	3.75	0.02 (95 %)	0.3048
		20	−4.24	0.01 (99 %)	0.3167
		21	4.05	0.01 (99 %)	0.3141
		25	2.73	0.10 (90 %)	0.3119
		27	4.77	0.00 (99 %)	0.3229
		28	−3.75	0.02 (95 %)	0.3038
		29	2.94	0.08 (90 %)	0.3007
		30	−2.56	0.13 (<90 %)	0.3042
	Region II	13	3.08	0.04 (95%)	0.3036
		25	3.07	0.04 (95 %)	0.3151
		27	4.83	0.00 (99 %)	0.3254
		28	−4.17	0.01 (99 %)	0.3080
		30	−4.46	0.01 (99 %)	0.3122
	Region III	7	−2.20	0.10 (90%)	0.3045
		8	−3.24	0.01 (99 %)	0.3105
		15	−6.15	0.00 (99 %)	0.3455
		16	3.42	0.01 (99 %)	0.3072
		19	−4.83	0.01 (99 %)	0.3304
		20	−6.19	0.00 (99 %)	0.3402
		28	−2.18	0.10 (90 %)	0.3025

Table 4 Diagnostics of surface and subsurface temperature predictor regression variables

Regression variable	Time series region	Lag (months)	Coefficient	p value (SL)	Rs^2
SST	WPAC	2	−0.14	0.02 (95 %)	0.2981
		14	−0.15	0.02 (95 %)	0.2898
		15	0.17	0.01 (99 %)	0.2888
		17	0.11	0.07 (90 %)	0.2716
	WPAC2	2	−0.20	0.01 (99 %)	0.3004
		14	−0.18	0.03 (95 %)	0.2882
		15	0.20	0.01 (99 %)	0.2878
		17	0.16	0.04 (95 %)	0.2759
	WPAC3	2	−0.28	0.00 (99 %)	0.3118
		15	0.21	0.01 (99 %)	0.2889
		17	0.12	0.13 (<90 %)	0.2725
	WPAC4	2	−0.24	0.00 (99 %)	0.3218
		5	0.15	0.01 (99 %)	0.2922
		15	0.16	0.00 (99 %)	0.2940
		17	0.10	0.08 (90 %)	0.2739
Temperature at 50 m depth	Region I	2	−0.45	0.00 (99 %)	0.3466
		3	−0.27	0.04 (95 %)	0.3352
		14	−0.20	0.13 (<90 %)	0.3232
		17	0.26	0.06 (90 %)	0.3079
		19	−0.19	0.15 (<90 %)	0.3092

Improving the long-lead predictability of El Niño using a novel forecasting scheme based· · ·

Table 4 continued

Regression variable	Time series region	Lag (months)	Coefficient	p value (SL)	Rs^2
Temperature at 100 m depth	Region I	2	−0.23	0.00 (99 %)	0.3512
		3	−0.19	0.00 (99 %)	0.3435
		6	−0.10	0.07 (90 %)	0.3341
		7	−0.11	0.05 (95 %)	0.3349
		8	−0.13	0.02 (95 %)	0.3331
		10	−0.11	0.05 (95 %)	0.3239
		19	−0.16	0.01 (99 %)	0.3181
		20	−0.10	0.07 (90 %)	0.3108
		35	−0.15	0.02 (95 %)	0.3275
	Region II	2	−0.14	0.04 (95 %)	0.3069
		4	−0.13	0.05 (95 %)	0.2981
		7	0.13	0.05 (95 %)	0.3061
		9	0.15	0.03 (95 %)	0.2680
		13	0.16	0.02 (95 %)	0.2886
		22	−0.21	0.00 (99 %)	0.2953
		25	0.22	0.00 (99 %)	0.2879
		26	0.15	0.03 (95 %)	0.2878
		35	−0.17	0.02 (95 %)	0.2956
		36	0.11	0.10 (90 %)	0.2953
Temperature at 100 m depth "cold"	Region I	6	0.07	0.08 (90 %)	0.2907
		8	−0.09	0.02 (95 %)	0.2930
		9	−0.07	0.08 (90 %)	0.2827
		26	0.06	0.09 (90 %)	0.2798

Table 5 Diagnostics of subsurface temperature predictor regression variables

Regression variable	Time series region	Lag (months)	Coefficient	p value (SL)	Rs^2
Temperature at 150 m depth	Region I	2	−0.17	0.00 (99 %)	0.3135
		3	−0.16	0.00 (99 %)	0.3102
		8	−0.17	0.00 (99 %)	0.3121
		9	−0.12	0.03 (95 %)	0.2968
		10	−0.14	0.01 (99 %)	0.2994
		19	−0.11	0.05 (95 %)	0.2861
		26	−0.10	0.06 (90 %)	0.2957
	Region II	4	0.08	0.05 (95 %)	0.3252
		5	0.11	0.01 (99 %)	0.3294
		10	0.06	0.14 (<90 %)	0.3196
		14	0.07	0.09 (90 %)	0.3225
		22	−0.07	0.09 (90 %)	0.2815
		25	0.08	0.05 (95 %)	0.3142
		35	−0.07	0.09 (90 %)	0.3211
		36	−0.08	0.05 (95 %)	0.3267
Temperature at 200 m depth	Region I	8	−0.14	0.04 (95 %)	0.3315
		9	−0.18	0.01 (99 %)	0.3296
		10	−0.13	0.07 (90 %)	0.3234
		19	−0.14	0.04 (95 %)	0.3127
		20	−0.12	0.09 (90 %)	0.3104
	Region II	3	0.14	0.01 (99 %)	0.3392
		4	0.17	0.00 (99 %)	0.3363
		5	0.16	0.00 (99 %)	0.3344

Table 5 continued

Regression variable	Time series region	Lag (months)	Coefficient	p value (SL)	Rs^2
		13	0.08	0.14 (<90 %)	0.3197
		14	0.12	0.03 (95 %)	0.3256
		35	−0.09	0.11 (<90 %)	0.3209
		36	−0.11	0.04 (95 %)	0.3284
Temperature at 250 m depth	Region I	5	0.19	0.04 (95 %)	0.3272
		6	0.15	0.10 (90 %)	0.3343
		8	−0.29	0.00 (99 %)	0.3428
		9	−0.20	0.03 (95 %)	0.3269
		12	0.19	0.05 (95 %)	0.3217
		24	0.15	0.10 (90 %)	0.3079
		28	−0.26	0.04 (95 %)	0.3492
		29	0.18	0.10 (90 %)	0.2818
		35	−0.15	0.10 (90 %)	0.3212
	Region II	5	0.39	0.00 (99 %)	0.3380
		11	−0.22	0.07 (90 %)	0.3227
		13	0.20	0.11 (<90 %)	0.3210
		21	−0.39	0.00 (99 %)	0.2859
		27	−0.27	0.03 (95 %)	0.3239
		34	−0.28	0.02 (95 %)	0.3234

Table 6 Diagnostics of subsurface temperature predictor regression variables

Regression variable	Time series region	Lag (months)	Coefficient	p value (SL)	Rs^2
Temperature at 300 m depth	Region I	9	−0.27	0.07 (90 %)	0.3240
		24	0.25	0.10 (90 %)	0.3078
		27	−0.25	0.10 (90 %)	0.3200
		28	−0.35	0.02 (95 %)	0.3170
		29	0.28	0.07 (90 %)	0.3163
		34	−0.24	0.12 (<90 %)	0.3179
		35	−0.44	0.00 (99 %)	0.3321
	Region II	2	0.31	0.12 (<90 %)	0.3185
		4	0.53	0.01 (99 %)	0.3182
		6	0.44	0.03 (95 %)	0.3025
		17	0.45	0.03 (95 %)	0.3023
		32	−0.35	0.09 (90 %)	0.3211
Temperature at 400 m depth	Region I	17	−0.32	0.12 (<90 %)	0.3077
		24	0.32	0.12 (<90 %)	0.2850
		28	−0.32	0.12 (<90 %)	0.2843
		29	0.38	0.07 (90 %)	0.3038
	Region II	10	−0.38	0.10 (90 %)	0.3215
		12	−0.41	0.08 (90 %)	0.3240
		15	0.38	0.10 (90 %)	0.3088
		29	−0.49	0.04 (95 %)	0.3033

Improving the long-lead predictability of El Niño using a novel forecasting scheme based· · ·

Table 4 continued

Regression variable	Time series region	Lag (months)	Coefficient	*p* value (SL)	Rs^2
Temperature at 500 m depth	Region I	2	−0.60	0.07 (90 %)	0.3192
		3	−0.58	0.08 (90 %)	0.3202
		6	0.74	0.02 (95 %)	0.3080
		10	−0.68	0.04 (95 %)	0.3109
		11	−0.74	0.02 (95 %)	0.3124
		36	−0.49	0.14 (<90 %)	0.3156
	Region II	4	0.77	0.01 (99 %)	0.3164
		5	1.24	0.00 (99 %)	0.2648
		6	0.90	0.00 (99 %)	0.3432
		10	−0.78	0.01 (99 %)	0.3298
		11	−1.22	0.00 (99 %)	0.3416
		12	−0.50	0.11 (<90 %)	0.3222
		15	0.65	0.04 (95 %)	0.3244
		17	0.60	0.05 (95 %)	0.3122

Table 7 Predictor regression variables added to the model at lead times between 17 and 36 months, based on the criteria and results shown in Tables 3–6

Lag/Lead	Regression variables
36	Temperature at 500 m depth Region I, temperature at 100, 150, 200 m depth Region II
35	Temperature at 100, 250, 300 m depth Region I, temperature at 100, 150, 200 m depth Region II
34	Temperature at 250 m depth Region II
32	Temperature at 300 m depth Region II
30	Zonal wind stress Region I and II
29	Temperature at 250, 300, 400 m depth Region I
28	Temperature at 250, 300, 400 m depth Region I, zonal wind stress Region I and III
27	Temperature at 300 m depth Region I, temperature at 250 m depth Region II, zonal wind stress Region I and II
26	Temperature at 100 m depth Region "cold", temperature at 150 m depth Region I, temperature at 100 m depth Region II
25	Temperature at 100, 150 m depth Region II, zonal wind stress Region I and II
24	Temperature at 250, 300, 400 m depth Region I
22	Temperature at 100, 150 m depth Region II
21	Temperature at 250 m depth Region II, zonal wind stress Region I
20	Temperature at 100, 200 m depth Region I, zonal wind stress Region I and III
19	Temperature at 50, 100, 150, 200 m depth Region I, zonal wind stress Region III
17	SST WPAC, WPAC2, WPAC3, WPAC4, temperature at 50, 400 m depth Region I, temperature at 300, 500 m depth Region II

D. Petrova et al.

Table 8 Predictor regression variables added to the model at lead times between 0 and 15 months, based on the criteria and results shown in Tables 3–6

Lag/Lead	Regression variables
15	SST WPAC, WPAC2, WPAC3, WPAC4, temperature at 250, 400, 500 m depth Region II, zonal wind stress Region III
14	SST WPAC, WPAC2, Temperature at 50 m depth, temperature at 150, 200 m depth Region II
13	Temperature at 100, 200, 250 m depth Region II
12	Temperature at 250 m depth Region I, temperature at 400, 500 m depth Region II
11	Temperature at 500 m depth Region I, temperature at 250, 500 m depth Region II, zonal wind stress Region II
10	Temperature at 100, 150, 200, 500 m depth Region I, temperature at 150, 400, 500 m depth Region II
9	Temperature at 150, 200, 250, 300 m depth Region I, temperature 100 m depth Region II, temperature at 100 m depth Region "cold"
8	Temperature at 100 m depth Region "cold", temperature at 100, 150, 200, 250 m, depth Region I, zonal wind stress Region I and III
7	Temperature at 100 m depth Region I and II, zonal wind stress Region I and III
6	Temperature at 100 m depth Region "cold", temperature at 100, 250, 500 m depth Region I, temperature at 300, 500 m depth Region II
5	SST WPAC4, Temperature at 250 m, depth Region I, temperature at 150, 200, 250, 500 m depth Region II
4	Temperature at 100, 150, 200, 300, 500 m depth Region II
3	Temperature at 50 m depth, temperature at 100, 150, 500 m depth Region I, temperature at 200 m depth Region II
2	SST WPAC, WPAC2, WPAC3, WPAC4, temperature at 50 m depth, temperature at 100 m depth Region I and II, temperature at 150, 500 m depth Region I, temperature at 300 m depth Region II
1	SST WPAC, WPAC3, WPAC4, temperature at 50 m depth, temperature at 100, 150, 200, 250, 300, 400 m depth Region I and II, 500 m depth Region II
0	SST WPAC2, temperature at 50 m depth, temperature at 100, 150, 200, 250 m depth Region I and II, temperature at 300, 400, 500 m depth Region II

Table 9 Predictions of the January target month for all EN events shown in Fig. 9

Lead month	1997–1998	2002–2003	2006–2007	2009–2010	2014–2015
27–29	*1.420 (10 %)*	0.644 (28 %)	*0.645 (28 %)*	*0.950 (19 %)*	0.223 (41 %)
24–26	*2.000 (4 %)*	0.784 (24 %)	*0.602 (29 %)*	*1.162 (15 %)*	*0.549 (32 %)*
21–23	*1.401 (11 %)*	*1.187 (14 %)*	*0.628 (29 %)*	*1.186 (14 %)*	*0.506 (34 %)*
18–20	*1.792 (6 %)*	0.971 (19 %)	*0.5386 (32 %)*	*1.165 (15 %)*	0.354 (38 %)
15–17	*1.544 (9 %)*	0.758 (24 %)	*0.639 (28 %)*	*1.262 (13 %)*	*0.565 (31 %)*
12–14	*1.377 (11 %)*	*1.110 (15 %)*	*0.682 (27 %)*	*1.457 (10 %)*	0.351 (38 %)
9–11	*1.305 (12 %)*	0.969 (19 %)	*0.642 (28 %)*	*1.257 (13 %)*	*0.519 (33 %)*
6–8	*2.006 (4 %)*	*1.138 (15 %)*	0.718 (26 %)	*1.446 (10 %)*	*0.500 (34 %)*
3–5	*2.503 (3 %)*	*1.394 (11 %)*	0.727 (25 %)	*1.564 (9 %)*	*0.613 (30 %)*
Obs.	**2.494 (3 %)**	**1.146 (15 %)**	**0.645 (28 %)**	**1.463 (10 %)**	**0.622 (30 %)**

Given in brackets is the probability that the respective value would occur based on a kernel normal probability density estimation of the N3.4 index with bandwidth of the kernel smoothing window h = 0.2704. Events are categorized as: strong/extreme (0–9 %), moderate (10–24 %), weak (25–35 %). The category of the event is predicted (bolditalic), or an adjacent category is predicted instead (italic). The observation values are in bold

Improving the long-lead predictability of El Niño using a novel forecasting scheme based···

Table 10 Correlation between forecasts and observations, and root mean square error (RMSE) as functions of lead month

Lead month	Correlation	RMSE
18	0.56	0.85
15	0.56	0.85
12	0.54	0.88
9	0.57	0.84
6	0.68	0.77
3	0.86	0.54

The period 1952–1982 is used for component estimation and forecasts are for the time series between 1983 and 2014

References

An SI, Choi J (2009) Seasonal locking of the ENSO asymmetry and its influence on the seasonal cycle of the tropical eastern Pacific sea surface temperature. Atmos Res 94:3–9

An SI, Jin FF (2004) Nonlinearity and asymmetry of ENSO. J Clim 17:2399–2412

Ballester J, Bordoni S, Petrova D, Rodo X (2015) On the dynamical mechanism explaining the western pacific subsurface temperature buildup leading to ENSO events. Geophys Res Lett 42:2961–2967

Ballester J, Bordoni S, Petrova D, Rodo X (2016) Heat advection processes leading to El Niño events as depicted by an ensemble of ocean assmiliation products. Submitted for review

Ballester J, Rodriguez-Arias MA, Rodo X (2011) A new extratropical tracer describing the role of the Western Pacific in the onset of El Nñio: Implications for ENSO understanding and forecasting. J Clim 24:1425–1437

Barnston A, Chelliah M, Goldenberg S (1997) Documentation of a highly ENSO-related SST region in the equatorial Pacific. Atmos Ocean 35:367–383

Barnston A, Ropelewski C (1992) Prediction of ENSO episodes using canonical correlation analysis. J Clim 5:1316–1345

Barnston A, Tippett M, L'Heureux M, Li S, DeWitt D (2012) Skill of real-time seasonal ENSO model predictions during 2002–11. Is our capability increasing? Am Meteorol Soc 93:631–651

Bjerknes J (1969) Atmospheric teleconnections from the equatorial Pacific. Mon Weather Rev 97:163–172

Brown J, Fedorov A (2010) How much energy is transferred from the winds to the thermocline on ENSO time scales? J Clim 23:1563–1580

Chen D, Cane M, Kaplan A, Zebiak S, Huang D (2004) Predictability of El Niño over the past 148 years. Nature 428:15

Chen D, Lian T, Fu C, Cane M, Tang Y, Murtugudde R, Song X, Wu Q, Zhou L (2015) Strong influence of westerly wind bursts on El Niño diversity. Nat Geosci 8:339–345

CPC (2015) Cold and warm episodes by season. http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml

de Jong P (1991) The diffuse Kalman Filter. Ann Stat 19:1073–1083

Doornik JA (2013) Object-Oriented Matrix Programming using Ox 7.0. Timberlake Consultants Ltd, London. See http://www.doornik.com

Durbin J, Koopman SJ (2012) Time series analysis by state space methods, 2nd edn. Oxford University Press, Oxford

Eisenman I, Yu L, Tziperman E (2005) Westerly wind bursts: ENSO's tail rather than the dog? J Clim 18:5224–5238

Fedorov A, Harper S, Philander SG, Winter B, Wittenberg A (2003) How predictable is El Niño? Bull Am Meteorol Soc 84:911–919

Fedorov A, Philander SG (2001) A stability analysis of tropical ocean-atmosphere interactions: bridging measurements and theory for El Niño. J Clim 14:3086–3101

Gebbie G, Tziperman E (2009) Predictability of SST-modulated westerly wind bursts. J Clim 22:3894–3909

Ghil M, Allen MR, Dettinger MD, Ide K, Kondrashov D, Mann ME, Robertson AW, Saunders A, Tian Y, Varadi F, Yiou P (2002) Advanced spectral methods for climatic time series. Rev Geophys 40:3.1–3.41

Gill A (1985) Elements of coupled ocean-atmosphere models for the tropics. Elsevier Oceanogr Ser 40:303–327

Glantz MH (2015) Shades of chaos: lessons learned about lessons learned about forecasting El Niño and its impacts. Int J Disaster Risk Sci 6:94–103

Good SA, Martin MJ, Rayner NA (2013) EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. J Geophys Res Oceans 118:6704–6716

Goswami BN, Shukla J (1991) Predictability of a coupled ocean-atmosphere model. J Clim 4:3–22

Hare S, Mantua N (2000) Empirical evidence for North Pacific regime shifts in 1977 and 1989. Prog Oceanogr 47:103–145

Harvey A (1989) Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge

Harvey A, Koopman SJ (2000) Signal extraction and the formulation of unobserved components models. Econom J 3:84–107

Harvey A, Koopman SJ, Penzer J (1998) Messy time series. In: Fomby T, Hill RC (eds) Advances in econometrics, vol 13. JAI Press, New York

Harvey A, Shephard N (1993) Structural time series models. Handbook of statistics, 11. Elsevier, Amsterdam

Ishii M, Shouji A, Sugimoto S, Matsumoto T (2005) Objective analyses of SST and marine meteorological variables for the 20th century using COADS and the Kobe collection. Int J Climatol 25:865–879

Izumo T, Lengaigne M, Vialard J, Luo J, Yamagata T, Madec G (2014) Influence of Indian Ocean Dipole and Pacific recharge on following year's El Niño: interdecadal robustness. J Clim Dyn 42:291–310

Jiang N, Neelin D, Ghill M (1993) Quasi-quadrennial and quasibiennial variability in COADS equatorial Pacific sea surface temperature and zonal wind. In: Proceedings of the 17th Climate Diagnostics Workshop. Climate Analysis Center, NOAA, pp 348–353

Jiang N, Neelin D, Ghill M (1995) Quasi-quadrennial and quasi-biennial variability in the equatorial Pacific. J Clim Dyn 12:291–310

Jin FF (1997a) An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model. J Atmos Sci 54:811–829

Jin FF (1997b) An equatorial ocean recharge paradigm for ENSO. Part II: a stripped down coupled model. J Atmos Sci 54:830–847

Jin FF, Kug JS, An S, Kang IS (2003) A near-annual coupled ocean-atmosphere mode in the equatorial Pacific. Geophys Res Lett 30

Jin FF, Neelin JD, Ghil M (1994) El Niño on the devil's staircase—annual subharmonic steps to chaos. Science 264:70–72

Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng Trans ASMA Ser D 82:35–45

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Leetmaa A, Reynolds R, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77:437–471

Kang IS, Kug JS (2002) El Niño and La Niña sea surface temperature anomalies: asymmetry characteristics associated with their wind stress anomalies. J Geophys Res 107:4372

Kleeman R, McCreary J, Klinger B (1999) A mechanism for generating ENSO decadal variability. Geophys Res Lett 26:1743–1746

Kondrashov D, Kravtsov S, Robertson A, Ghil M (2005) A hierarchy of data-based ENSO models. J Clim 18:4425–4444

Koopman SJ, Harvey AC, Doornik JA, Shephard N (2010) STAMP 8.3: structural time series analyser, modeller and predictor. Timberlake Consultants, London

Koopman SJ, Shephard N, Doornik JA (2008) Statistical algorithms for models in state space form: SsfPack 3.0. Timberlake Consultants Press, London

Krishnamurthy L, Vecchi G, Msadek R, Wittenberg A, Delworth T, Zeng F (2015) The seasonality of the great plains low-level jet and ENSO relationship. J Clim 28:4525–4544

Lau K, Shen P (1988) Annual cycle, quasi-biennial oscillation and Southern Oscillation in global precipitation. J Geophys Res 93:10975–10988

Mantua N, Battisti D (1995) Aperiodic variability in the Zebiak-Cane coupled ocean-atmosphere model: air-sea interactions in the western equatorial Pacific. J Clim 8:2897–2927

McGregor S, Timmermann A, Shneider N, Stuecker M, England M (2012) The effect of the South Pacific Convergence Zone on the termination of El Niño events and the meridional asymmetry of ENSO. J Clim 25:5566–5586

McPhaden M (2004) Evolution of the 2002/2003 El Niño. Am Meteorol Soc 85:677–695

McPhaden M, Timmermann A, Widlansky M, Balmaseda M, Stockdale T (2014) The curious case of the El Niño that never happened: a perspective from 40 years of progress in climate research and forecasting. Bull Am Meteorol Soc

McPhaden M, Yu X (1999) Equatorial waves and the 1997/98 El Niño. Geophys Res Lett 26:2961–2964

Moron V, Plaut G (2003) The impact of El Niño-Southern Oscillation upon weather regimes over Europe and the North Atlantic during boreal winter. Int J Climatol 23:363–379

Neelin J (1990) A hybrid coupled general circulation model for El Niño studies. J Atmos Sci 47:677–695

Neelin J (1991) The slow surface temperature mode and the fast-wave limit: analytic theory for tropical interannual oscillations and experiments in a hybrid coupled model. J Atmos Sci 48:584–606

Penland C (1996) A stochastic model of Indo-Pacific sea surface temperature anomalies. Phys D 98:534–558

Penland C, Magorian T (1993) Prediction of Niño3 sea surface temperatures using linear inverse modeling. J Clim 6:1067–1076

Philander S (1989) El Niño and La Niña. J Atmos Sci 77:451–459

Philander S, Yamagata T, Pacanowski R (1984) Unstable air-sea interactions in the tropics. J Atmos Sci 41:604–613

Ramesh N, Murtugudde R (2013) All flavours of El Niño have similar early subsurface origins. Nat Clim Change 3:42–46

Rasmusson E, Carpenter T (1982) Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. Mon Weather Rev 110:354–384

Rasmusson E, Wang X, Ropelewski C (1990) The biennial component of ENSO variability. J Mar Syst 1:71–96

Sarachik E, Cane M (2010) The El Niño Southern Oscillation Phenomenon. Cambridge University Press, Cambridge

Stein K, Timmermann A, Shcneider N (2011) Phase synchronization of El Niño–Southern Oscillation with the annual cycle. Phys Rev Lett 107

Thompson CJ, Battisti DS (2000) A linear stochastic dynamical model of ENSO. Part I: model development. J Clim 13:2818–2832

Thompson CJ, Battisti DS (2001) A linear stochastic dynamical model of ENSO. Part II: analysis. J Clim 14:445–466

Tong H (1990) Non-linear time series. Oxford University Press, Oxford

Torrence C, Webster PJ (1998) The annual cycle of persistence in the El Niño–Southern Oscillation. Q J R Meteorol Soc 124:1985–2004

Trenberth K (1976) Spatial and temporal variations in the Southern Oscillation. Q J R Meteorol Soc 102:639–653

Tziperman E, Stone L, Cane M, Jarosh H (1994) El Niño chaos: overlapping of resonances between the seasonal cycle and the Pacific ocean-atmosphere oscillator. Science 264:72–74

Tziperman E, Yu L (2007) Quantifying the dependence of westerly wind bursts on the large-scale tropical Pacific SST. J Clim 20:2760–2768

Tziperman E, Zebiak S, Cane M (1997) Mechanisms of seasonal—ENSO interaction. J Atmos Sci 54:61–71

Wyrtki K (1975) El Niño—the dynamic response of the equatorial Pacific Ocean to atmospheric forcing. J Phys Oceanogr 5:572–584

Wyrtki K (1985) Water displacements in the Pacific and the genesis of El Niño cycles. J Geophys Res 90:7129–7132

Xue Y, Cane M, Zebiak S, Blumenthal M (1994) On the prediction of ENSO: a study with a low-order Markov model. Tellus A 46:512–528

Yasunari T (1989) A possible link of the QBO's between the stratosphere, troposphere and the surface temperature in the tropics. J Meteorol Soc Jpn 67

Yeh S, Kirtman B (2004) The decadal ENSO variability in a hybrid coupled. J Am Meteorol Soc 17:1225–1238

Yeh S, Kirtman B (2005) Tropical Pacific decadal variability and ENSO amplitude modulation. Geophys Res Lett 32

Yu J, Kao H, Lee T (2011) Subsurface ocean temperature indices for central-Pacific and eastern-Pacific types of El Niño and La Niña events. Theor Appl Climatol 103:337–344

Yu J, Mechoso C (2001) A coupled atmosphere-ocean GCM study of the ENSO cycle. J Clim 14:2329–2350

Yu X, McPhaden M (1999) Seasonal variability in the equatorial Pacific. J Phys Oceanogr 29:925–947

Zebiak S (1985) Tropical atmosphere-ocean interaction and the El Niño/Southern Oscillation phenomenon. PhD Thesis, Massachusetts Institute of Technology

Zebiak S (1989) Oceanic heat content variability and El Niño cycles. J Phys Oceanogr 19:475–486

Zebiak SE, Cane MA (1987) A model El Niño-Southern Oscillation. Mon Weather Rev 115:2262–2278

3.3 Heat advection processes leading to El Niño events as depicted by an ensemble of ocean assimilation products

Ballester J (1), Bordoni S (1), Petrova D (2), Rodó X (2,3)

(1) California Institute of Technology, Pasadena, California, United States
(2) Institut Català de Ciències del Clima, Barcelona, Catalonia, Spain
(3) Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

Abstracto - *Summary in Spanish*

La naturaleza oscilatoria de El Niño-Oscilación del Sur resulta de una intrincada superposición de balances de cuasi-equilibrios y procesos de desequilibrio fuera de fase entre el océano y la atmósfera. El objetivo principal del presente trabajo es realizar un exhaustivo análisis espaciotemporal del presupuesto de calor oceánico superior en un conjunto de productos de asimilación oceánica de vanguardia. Ponemos énfasis en los mecanismos de advección del calor del océano, y su representación en conjuntos de miembros individuales y en las diferentes etapas de la oscilación ENSO que conduce a eventos EN. Nuestros análisis consistentemente muestran que el calentamiento subsuperficial inicial en el Pacífico ecuatorial occidental se advecta al Pacífico central por la subcorriente ecuatorial, la cual, junto con la advección ecuatorial asociada con anomalías tanto en el gradiente de temperatura meridional como en la circulación a nivel de la termoclina, explica la acumulación de calor en el Pacífico central durante la fase de recarga. También encontramos que la fase de recarga se caracteriza por un aumento de la inclinación meridional de la termoclina, así como un transporte de masa equatorial cruzada hacia el sur en el océano resultante del movimiento vertical anómalo inducido por Ekman en las zonas subecuatoriales. Aunque las diferencias entre conjuntos de datos son generalmente pequeñas, y las anomalías tienden a tener el mismo signo, las diferencias en la magnitud del término meridional se consideran clave para explicar los diferentes velocidades de propagación de la tendencia de calentamiento subsuperficial a lo largo de la termoclina. La única excepción es GECCO, que no produce los patrones de divergencia Ekman meridional superficial (Convergencia Sverdrup subsuperficial) en el Pacífico ecuatorial occidental y central.

AGUPUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

Heat advection processes leading to El Niño events as depicted by an ensemble of ocean assimilation products

Joan Ballester[1,2], Simona Bordoni[2], Desislava Petrova[1], and Xavier Rodó[1,3]

[1]Institut Català de Ciències del Clima (IC3), Barcelona, Catalonia, Spain, [2]California Institute of Technology (Caltech), Pasadena, California, USA, [3]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

Abstract The oscillatory nature of El Niño-Southern Oscillation results from an intricate superposition of near-equilibrium balances and out-of-phase disequilibrium processes between the ocean and the atmosphere. The main objective of the present work is to perform an exhaustive spatiotemporal analysis of the upper ocean heat budget in an ensemble of state-of-the-art ocean assimilation products. We put specific emphasis on the ocean heat advection mechanisms, and their representation in individual ensemble members and in the different stages of the ENSO oscillation leading to EN events. Our analyses consistently show that the initial subsurface warming in the western equatorial Pacific is advected to the central Pacific by the equatorial undercurrent, which, together with the equatorward advection associated with anomalies in both the meridional temperature gradient and circulation at the level of the thermocline, explains the heat buildup in the central Pacific during the recharge phase. We also find that the recharge phase is characterized by an increase of meridional tilting of the thermocline, as well as a southward upper-ocean cross-equatorial mass transport resulting from Ekman-induced anomalous vertical motion in the off-equatorial regions. Although differences between data sets are generally small, and anomalies tend to have the same sign, the differences in the magnitude of the meridional term are seen to be key for explaining the different propagation speed of the subsurface warming tendency along the thermocline. The only exception is GECCO, which does not produce the patterns of meridional surface Ekman divergence (subsurface Sverdrup convergence) in the western and central equatorial Pacific.

1. Introduction

El Niño-Southern Oscillation (ENSO) [*Ballester et al.*, 2016] is the dominant source of interannual variability worldwide and one of the most important modes of variability in the tropical Pacific, with far-reaching influences on the whole climate system [*Jin*, 1997a,1997b; *Meinen and McPhaden*, 2000; *Wang*, 2002; *Brown and Fedorov*, 2010; *Ballester et al.*, 2011, 2013; *Petrova et al.*, 2016]. The large amplitude of ENSO anomalies in the tropical Pacific is essentially explained by the strong coupling between the Walker circulation, the zonal gradient of sea surface temperature and the longitudinal tilt of the thermocline (i.e., the so-called Bjerknes feedback [*Bjerknes*, 1969; *Wyrtki*, 1975]). These interactions are however modulated by out-of-phase negative feedbacks that bound the amplitude and reverse the sign of interannual anomalies. According to the delayed oscillator theory, this reversal is explained by the differential propagation speed of wind-induced oceanic Kelvin and Rossby waves [*Battisti*, 1988; *Schopf and Suarez*, 1988]. While eastward-propagating Kelvin waves quickly deepen the warm ocean layer in the eastern Pacific [*Wang*, 2002], westward Rossby waves travel at lower speeds, and start to shallow the thermocline only after being reflected as Kelvin waves at the western boundary [*Fedorov and Brown*, 2009].

Among other models that have been proposed, the recharge oscillator emphasizes the time delay between anomalies in longitudinally averaged thermocline depth and eastern Pacific sea surface temperature [*Jin*, 1997a,1997b]. In this conceptual framework, a deeper-than-normal thermocline suppresses the active upwelling in the eastern Pacific and favors the growth of an El Niño (EN) event and the weakening of the trade winds, whose curl generates poleward Sverdrup transport that discharges the heat in the upper ocean and reverses the sign of ENSO [*Meinen and McPhaden*, 2000]. This theory, therefore, hypothesizes that the oscillatory nature of ENSO results from the balance between equatorial zonal winds and the pressure gradient associated with the equatorial thermocline tilt, as well as from the disequilibrium between the mean

basin-wide thermocline depth and the meridional convergence or divergence of Sverdrup transport due to tropical wind stress curl anomalies [*Jin*, 1997a,1997b; *Singh and Delcroix*, 2013].

Zonal and vertical currents are indeed intimately connected through the energy balance, because a significant fraction of the wind power is converted into buoyancy power [*Brown and Fedorov*, 2010]. This transfer explains how the energy supplied by enhanced trade winds to the westward South Equatorial Current (SEC) in the central Pacific is converted into downward (upward) mass fluxes in the western (eastern) Pacific that distort local ocean isopycnals and deepen (shoal) the thermocline [*Brown et al.*, 2011]. The increased (decreased) thermocline tilting in the equatorial Pacific associated with stronger (weaker) than normal trade winds induces large cold (warm) anomalies in sea surface temperature in the eastern Pacific, which are amplified by the ocean-atmosphere coupling and extended to the central Pacific by means of zonal advection.

The zonal advective, the Ekman pumping and the thermocline feedbacks have been described as the three major dynamical processes contributing to the amplification of temperature anomalies during the onset of ENSO events [*Jin and Neelin*, 1993]. Thus, assuming a small initial warm perturbation in the equatorial surface, the coupled system rapidly responds by weakening the trade winds and reducing the zonal tilting of the equatorial thermocline [*Jin et al.*, 2006], which in turn generates anomalous eastward geostrophic currents in the central and eastern Pacific [*Santoso et al.*, 2013]. The upper ocean response is characterized by the decrease of the depth of the thermocline and the generation of anomalous zonal currents in the central and eastern Pacific, which together, amplify the initial anomalies and bring the oscillation to a mature phase [*Jin and An*, 1999]. These mechanisms also play an important role in the dampening and reversal of ENSO conditions when Sverdrup mass divergence starts to discharge the heat content in the equatorial Pacific after the mature phase of EN conditions.

The main objective of the present work is to perform an exhaustive spatiotemporal analysis of the upper ocean heat budget, emphasizing similarities and differences between the individual members of an ensemble of state-of-the-art ocean assimilation products. We specifically focus on the ocean heat advection mechanisms that characterize the stages of the ENSO oscillation leading to EN events. With a similar approach, we have recently shown how the ensemble mean of these ocean assimilation products can provide important insights into mechanisms that contribute to the heat buildup in the western Pacific 18–24 months before EN events [*Ballester et al.*, 2015]. Here we extend these analyses, and we not only describe the similarities and differences between ensemble members, but also generalize them to the whole ENSO cycle and the entire Pacific ocean basin. The main results emerging from this study can hence be used to understand to what extent results derived from these assimilation products can be used to describe the dynamics of ENSO, such as those in *Ballester et al.* [2015]. Additionally, they can be taken as a reference for validation and assessment of numerical simulations. While still largely disagreeing in some key dynamical processes, because of the large differences in their underlying models, assimilation techniques and assimilated observations [*Ray et al.*, 2015], these products provide the best and most complete spatiotemporal picture of the ocean subsurface available to date. Indeed, all the data sets used in this study correctly capture all the EN and La Niña (LN) events, and they only differ in some of the processes leading to their mature phase, as analyzed and discussed below.

After discussing our methodology (section 2), we use the ensemble of assimilation products to describe the transitions that characterize the swing between phases of the oscillation, from a climatological neutral base state (section 3) to the generation of a subsurface warm buildup in the western Pacific (section 4), the recharged phase in basin-wide equatorial heat content (section 5) and the onset and mature phases of EN (section 6). Discussion and summary are provided in sections 7 and 8, respectively.

2. Methods

The onset of EN events is characterized by an initial subsurface heat buildup in the western Pacific, the subsequent eastward movement of the accumulated warm waters along the equatorial thermocline (i.e., recharge mode in the central Pacific) and the final rapid amplification of temperature anomalies in the eastern Pacific due to the coupled ocean-atmosphere Bjerknes feedback [*Ballester et al.*, 2015]. The present article describes the role of heat advection in each of these three stages of the oscillation before the mature phase of EN events. To this aim, we analyze the different terms of the temperature tendency equation,

which links the potential temperature (θ) tendency to the zonal (U_{adv}), meridional (V_{adv}) and vertical (W_{adv}) heat advection, thermal forcing (Q) and residual terms (R) through:

$$\frac{\partial \theta}{\partial t} = U_{adv} + V_{adv} + W_{adv} + Q + R. \tag{1}$$

We note that certain data assimilation techniques may result in this evolution equation being nonconservative. We do not explicitly compute the thermal forcing as our focus is on the equatorial subsurface below the mixed layer, with climatological depths ranging from 20 m in the eastern Pacific to 70 m in the western Pacific, where the effect of Q is small. The interannual anomalies of the heat advection components are expressed as

$$U'_{adv} = -\bar{u}\frac{\partial \theta'}{\partial x} - u'\frac{\partial \bar{\theta}}{\partial x} - u'\frac{\partial \theta'}{\partial x} + \overline{u'\frac{\partial \theta'}{\partial x}}, \tag{2}$$

$$V'_{adv} = -\bar{v}\frac{\partial \theta'}{\partial y} - v'\frac{\partial \bar{\theta}}{\partial y} - v'\frac{\partial \theta'}{\partial y} + \overline{v'\frac{\partial \theta'}{\partial y}} \tag{3}$$

and

$$W'_{adv} = -\bar{w}\frac{\partial \theta'}{\partial z} - w'\frac{\partial \bar{\theta}}{\partial z} - w'\frac{\partial \theta'}{\partial z} + \overline{w'\frac{\partial \theta'}{\partial z}}, \tag{4}$$

where the overbar and the prime denote the climatological and anomalous components, respectively, and u, v and w the zonal, meridional and vertical current velocities. A 13-term running average (1/24, 1/12, . . ., 1/12, . . ., 1/12, 1/24) is used to calculate the interannual anomaly component from detrended monthly variables. Other low-frequency filters were tested, such as a recursive Butterworth procedure [*Ballester et al.*, 2011], but similar results were found. Given that the contribution of the nonlinear advection terms (i.e., the last two terms in equations (2–4)) is generally small compared to the other components, they will not be explicitly described in this work, although they are implicitly included in the U'_{adv}, V'_{adv} and W'_{adv} terms throughout this article.

Ocean potential temperature and zonal and meridional current velocities are obtained from five assimilation products: NEMOVAR-COMBINE (model: NEMO v3.0) [*Balmaseda et al.*, 2010], GECCO (MITgcm) [*Köhl and Stammer*, 2008], SODA2.2.6 (POP2.x) [*Carton and Giese*, 2008], ORAS4 (NEMO v3.0) [*Balmaseda et al.*, 2013], and ORAS3 (HOPE) [*Balmaseda et al.*, 2008]. Vertical velocity is diagnosed by integrating horizontal divergence down from the surface, with surface values assumed to be equal to the time tendency of sea surface height.

EN events are chosen according to the classification of the Climate Prediction Center: December 1963, 1965, 1968, 1972, 1976, 1982, 1986, 1990, 1997, 2002, 2006 [*CPC*, 2015]. In those cases when EN conditions are observed in the tropical Pacific for two consecutive boreal winters (i.e., 1968/1969, 1976/1977, 1986/1987, 1990/1991), only the first year is considered for the calculation of the composite anomalies, given that the main objective of the article is the description of the onset of these events. The 1994 event was excluded from the analyses because it was the continuation of a previous warm event starting in 1990, with warm sea surface temperature anomalies persisting in the central and eastern tropical Pacific for almost 6 years [*Trenberth and Hoar*, 1996]. A similar criterion is used for the selection of LN years in Figure 2: December 1964, 1966, 1970, 1973, 1975, 1984, 1988, 1995, 1998, 2007 [*Ballester et al.*, 2015].

3. Climatological and ENSO Year Features

The main climatological features that characterize the circulation and the thermodynamic structure of the equatorial and off-equatorial Pacific Ocean in the assimilation products is consistent with previous observational studies and is shown in Figure 1. The westward SEC is simulated in the tropical south Pacific, extending from 20S to about 3–4N across the equator, where it is largely driven by the trade winds (Figures 1a and 1b). The circulation in the equatorial Pacific is also characterized by the eastward Equatorial Under Current (EUC), a subsurface current 200–400 km wide transporting 30–40 Sv along the tilted equatorial thermocline [*Izumo*, 2005, Figure 1b]. The zonal velocity of the EUC is strongest in the central Pacific at about 140W, and
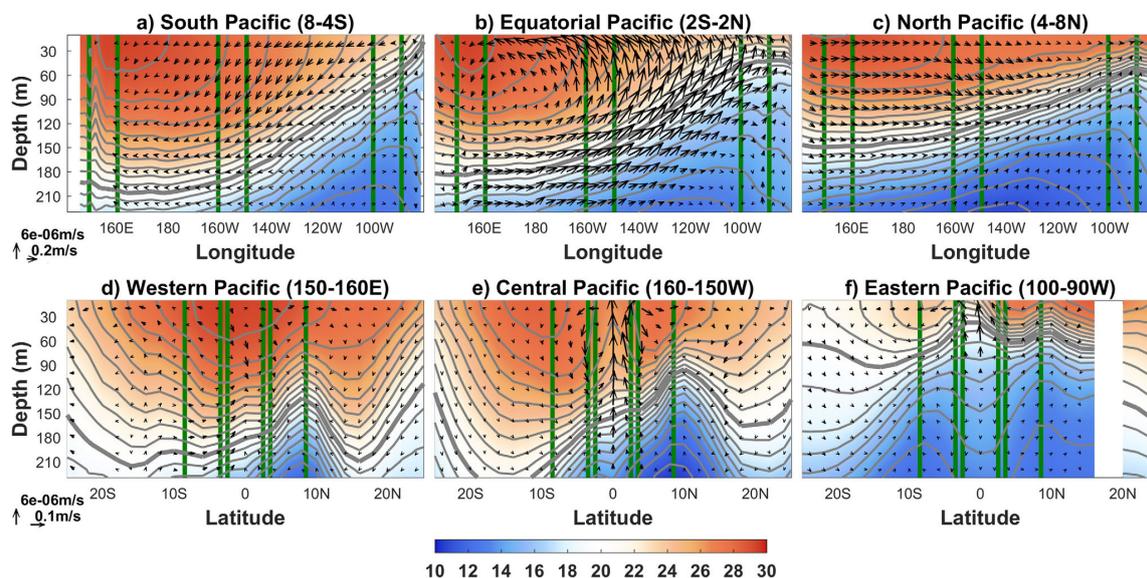
Figure 1. Climatological temperature and ocean currents. Multiproduct average for the common period 1961–2000 of potential temperature (°C, (a–f) shading and contours) and (a–c) zonal, (d–f) meridional, and (a–f) vertical ocean velocities (m/s, arrows) averaged in latitude over the south off-equatorial (8–4S), (b) equatorial (2S–2N), and (c) north off-equatorial (4–8N) regions, and averaged in longitude in the (d) western (150–160E), (e) central (160–150W) and (f) eastern (100–90W) Pacific (see the vertical green lines). The temperature contour interval is 1°C, and the thick contour corresponds to the 20°C isotherm.

is primarily driven by the east-west pressure gradient in the equatorial plane, in turn determined by the strength of the easterly zonal wind stress. In the northern off-equatorial Pacific, the assimilation products also consistently reproduce the North Equatorial Counter Current (NECC), whose eastward transport is determined by Sverdrup dynamics and whose spatial structure is constrained by near-equatorial zonal wind stress [*Yu et al.*, 2000, Figure 1c].

The role of the trade winds is also key for explaining the spatial distribution of temperature and vertical currents along the equatorial Pacific. The dynamical forcing associated with the easterly wind stress piles up warm waters to the western Pacific and deepens the local thermocline there (Figures 1b and 1d). In this area, horizontal current convergence in the ocean surface induces weak downwelling motion above the thermocline level, at the narrow westernmost edge of the longitudinal band with strong easterlies (i.e., 150–160E) [*Ballester et al.*, 2015]. In the eastern Pacific, the wind stress forcing shoals the thermocline, with a zonal contrast of about 100–120 m in thermocline depth between the western and eastern parts of the basin (cf. Figures 1b, 1d, and 1f). Ekman-driven equatorial upwelling in the central Pacific and coastal upwelling in the eastern Pacific bring to the surface cold water from below the thermocline level, which explains the equatorial minimum in surface temperatures relative to the off-equatorial bands (Figures 1e and 1f). The rising motion in the central equatorial Pacific is part of the shallow meridional overturning circulation, with surface poleward divergence, symmetric downwelling motion in the off-equatorial bands and meridional convergence at the pycnocline level [*Izumo*, 2005, Figures 1a–1c, and 1e].

The difference between the average thermodynamic structure in the equatorial and off-equatorial Pacific Ocean during the mature phase of EN and LN events is shown in Figure 2. During LN (EN) events, the strengthening (weakening or even reversal) of the trade winds increases (decreases) the tilting of the thermocline and enhances (reduces) most features of the oceanic circulation, including the SEC, the EUC, the downwelling motion in the western Pacific and the shallow meridional overturning cells in the central Pacific. Two prominent exceptions are found in the assimilation products. First, the eastward NECC is weakened (intensified) during the growing and mature phases of LN (EN) as a result of the decreased (increased) wind stress curl north of the equator [*Hsin and Qiu*, 2012, Figure 2c]; together with the enhancement (suppression) of the westward SEC, this contributes to the westward (eastward) displacement of the warm pool and the development of ENSO anomalies. Second, the equatorial upwelling in the far eastern Pacific starts
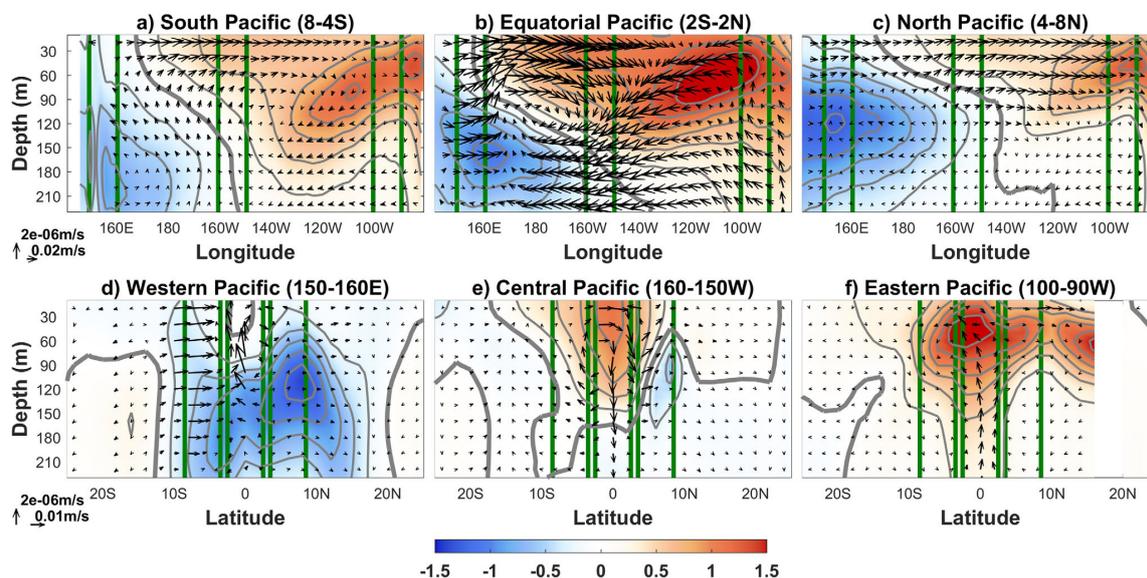
Figure 2. Same as Figure 1, but for the difference of El Niño minus La Niña years. The temperature contour interval is 0.25°C, and the thick contour corresponds to the 0°C isotherm.

to be suppressed (intensified) just a couple of months before the peak of LN (EN) events (Figure 2f), being a primary out-of-phase reversal mechanism for the oscillatory nature of ENSO [*Battisti*, 1988; *Jin*, 1997a,1997b].

4. Growth of the Warm Buildup in the Western and Central Pacific

Figures 3–5 depict the multiproduct average of the composite of EN events for the range of lags corresponding to the generation of the subsurface heat buildup in the western Pacific, between 36 and 25 months before El Niño events. The temperature tendency at these lags determines the anomalies that are observed later, during the peak of the subsurface heat buildup in the western Pacific at lag −21 months. The stippling highlights the inter-product similarities by showing the areas where anomalies have the same sign and magnitude larger than ±0.25°C/yr for all or all but one the members of the ensemble, an approach that we take throughout the manuscript. The inter-product differences corresponding to the vertical, meridional and zonal advection terms are additionally presented in Figures 6 and 7, respectively. Here we show the longitude-depth values along a narrow band in the equatorial Pacific (2S–2N, Figures 3 and 6, and 7), and two latitude-depth meridional transects representative of the processes that lead to the initial stages of the heat buildup in the warm pool (150–160E, Figure 4) and to the east of the dateline (160–150W, Figure 5).

Ballester et al. [2015] highlighted fundamental differences between the dynamical origin of the subsurface warming in these two regions. On the one hand, in the western Pacific (130–170E), anomalous downwelling motion from an upper layer (0–75 m) of horizontal convergence to a subsurface layer (75–190 m) of horizontal divergence deepens the thermocline and advects heat downward (cf. Figure 3a). Horizontal convergence near the surface is found to be an intricate combination of surface (0–60 m) zonal convergence and subsurface (40–75 m) meridional convergence, while the horizontal divergence just above the thermocline level is explained by the zonal component and the intensification of the EUC in the central Pacific. On the other hand, a completely different regime prevails in the central Pacific, where surface (0–60 m) zonal and meridional divergence and subsurface (60–190 m) zonal and meridional convergence generate strong upwelling motion (cf. Figure 3a).

Analysis of the heat budget reveals that most of the spatial structure of surface and subsurface heat anomalies in the equatorial Pacific is explained by zonal and vertical advective processes (cf. shading and contours in Figure 3b). To the west of 170E, vertical advection determines a large fraction of the subsurface warming (Figure 3h). In particular, this contribution is primarily explained by the vertical advection of the climatological
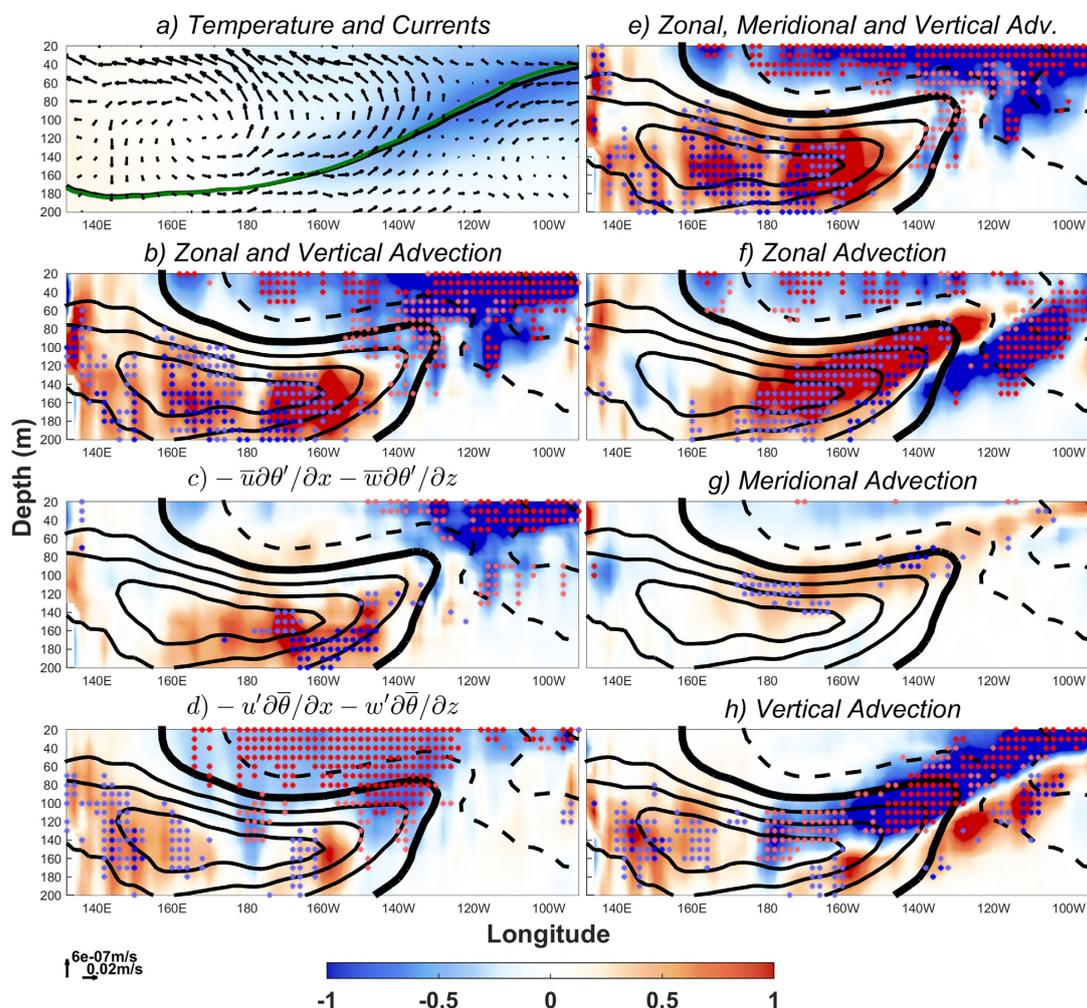
Figure 3. Multiproduct average of temperature tendency and heat advection during the growing phase of the heat buildup leading to El Niño events. Composite anomalies are averaged over the equatorial band (2S–2N), and for the range of lags between 36 and 25 months before the major El Niño events. (a) The anomalous zonal and vertical velocity (m/s, arrows), the anomalous potential temperature (°C, shading), the climatological 20°C isotherm (black contour) and the anomalous 20°C isotherm (green contour). (b–h) Time tendency of potential temperature (°C/yr, contours) and the heat advection terms specified in the plot titles (°C/yr, shading). The contour interval is 0.25°C/yr, with solid (dashed) lines depicting positive (negative) anomalies. The dark (light) stippling denotes areas where heat advection anomalies have the same sign and magnitude larger than ±0.25°C/yr for all (all but one) the members of the ensemble.

temperature by anomalous currents ($-w'\partial\overline{\theta}/\partial z$, not shown), which results from the combination of (rather weak) downwelling anomalies (Figure 3a) and the strong climatological vertical gradient of temperature (Figure 1b). Meridional cross-sections in the warm pool confirm the dominant role of vertical advection within the tropical band in this region (Figure 4). A tendency toward subsurface warming is present from 12S to 8N, which approximately corresponds to the latitudinal range with anomalous downwelling motion (Figure 4a). Nevertheless, the warming is clearly larger right at the equator near the thermocline, where both the vertical gradient of temperature and the anomaly in downward vertical velocity are largest (Figures 1d and 4a). Figure 4 also confirms the negligible contribution of zonal and meridional heat advection in this region at these very initial stages of the composite of EN events, regardless of the specific latitudes within the tropical band.

To the east of 170E, the subsurface warming is a complex combination of different mechanisms. The largest contribution to subsurface warming is associated with temperature and circulation changes along the equatorial thermocline from the zonal and vertical terms (Figure 3b). For the zonal component, the advection of
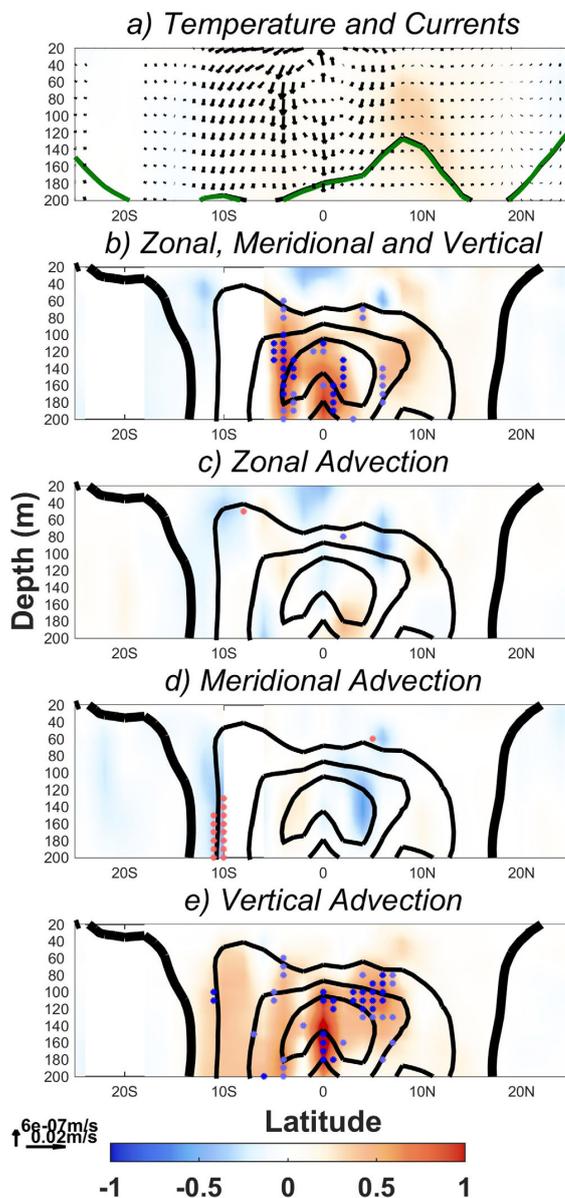
Figure 4. Same as Figures 3a and 3e–3h, but for the meridional transect of temperature, meridional and vertical currents and heat advection in the 150–160E sector.

anomalous heat by climatological currents (i.e., $-\bar{u}\,\partial\theta'/\partial x$) and the advection of climatological temperature by anomalous currents ($-u'\,\partial\bar{\theta}/\partial x$) have similar contributions, with anomalies reaching up to +1°C/yr in the central Pacific near the thermocline level (not shown). Note that nearly opposite anomalies are found for the vertical component (i.e., $-\bar{w}\,\partial\theta'/\partial z$ and $-w'\,\partial\bar{\theta}/\partial z$, not shown).

The tendency in subsurface warming in the central Pacific is to a large extent explained by $-(\bar{u}\,\partial/\partial x+\bar{w}\,\partial/\partial z)\theta'$ (cf. shading and contours in Figure 3c). This contribution is associated with the negative eastward and upward gradient of subsurface temperature along the equatorial thermocline due to increasing LN-like conditions (Figure 3a), which is advected to the central Pacific by the climatological EUC (Figure 1b). These factors generate positive zonal advection at the level of the thermocline and negative (positive) vertical advection above (below) the thermocline (Figures 3f and 3h), which together explain the warming tendency observed below 100–120 m (Figure 3c). The overall contribution of $-(u'\,\partial/\partial x+w'\,\partial/\partial z)\bar{\theta}$ is instead associated with the tilted stratification of the ocean (Figure 1b), which is advected by the intensification of the EUC (Figure 3a). The diapycnal component of these circulation anomalies is not negligible in this case, with eastward anomalies defining areas of warm advection at 160W and upward anomalies generating cold advection at 180 and 140W (Figure 3d).

Unlike what is seen in the warm pool, the tendency in subsurface temperatures in the central Pacific results from a combination of the three heat advection components (cf. shading and contours in Figure 5b). Notably, while smaller than the zonal and vertical components, the meridional advection plays an important role in determining the tendency of subsurface temperature in the off-equatorial bands (cf. Figures 5c and 5e). In fact, the zonal and vertical advection components tend to cancel each other in the latitude intervals 6–2S and 2–5N at 100 m (Figures 5c, 5d, and 5f), just above the thermocline (Figure 5a). Therefore, the warming tendency in this area is entirely explained by the meridional advection of anomalous heat by climatological currents ($-\bar{v}\,\partial\theta'/\partial y$, not shown), which advects relatively less cold off-equatorial temperature anomalies (Figure 5a) through the equatorward lower branch of the climatological shallow meridional overturning cells (Figure 1e).

Figure 5. Same as Figures 3a, b, and 3e–3h, but for the meridional transect of temperature, meridional and vertical currents and heat advection in the 160–150W sector.

Although the above-described advection processes are generally reproduced by the five assimilation products, there are still substantial differences in their magnitude and spatial extent at this early stage of the oscillation. For example, all products reproduce the subsurface warming due to vertical advection to the west of 170E, but its intensity and extent vary greatly among the data sets: the downward advection is weak in ORAS3, confined to a narrow band in SODA2.2.6, close in magnitude to the composite average in NEMOVAR-COMBINE and ORAS4, and strong and extending eastward to 160W in GECCO (Figure 6). These differences are essentially a reflection of differences in the ocean circulation, because of differing pattern and magnitude of the vertical velocity in each data set (not shown). In turn, the warming due to zonal advection in the central Pacific is directly related to the subsurface warming tendency to the west of 170E, which is zonally advected along the thermocline by the EUC (not shown). Differences are even larger for the meridional component, with no contribution to the subsurface warming in GECCO, around average values in SODA2.2.6 and ORAS3, and a strong positive contribution in NEMOVAR-COMBINE and ORAS4 (Figure 7). The large differences in the magnitude of the meridional term are seen to be key for explaining the different propagation speed of the subsurface warming tendency along the thermocline (cf. contours in Figure 7, find more details in the Discussion section below).

5. Transition to the Recharged Phase

The composites corresponding to the development of the basin-wide recharge mode in equatorial heat content are shown in Figures 8 and 9. These figures show averages for lags between 27 and 16 months before EN events, whose tendencies determine the subsequent peak of the recharge phase on average at lag-09 months. This range of lags is characterized by the peak in LN-like conditions (Figure 8a) and a tendency toward warming of the equatorial Pacific around the level of the thermocline (contours in Figure 8). In this section we show, as we did in the previous section, the longitude-
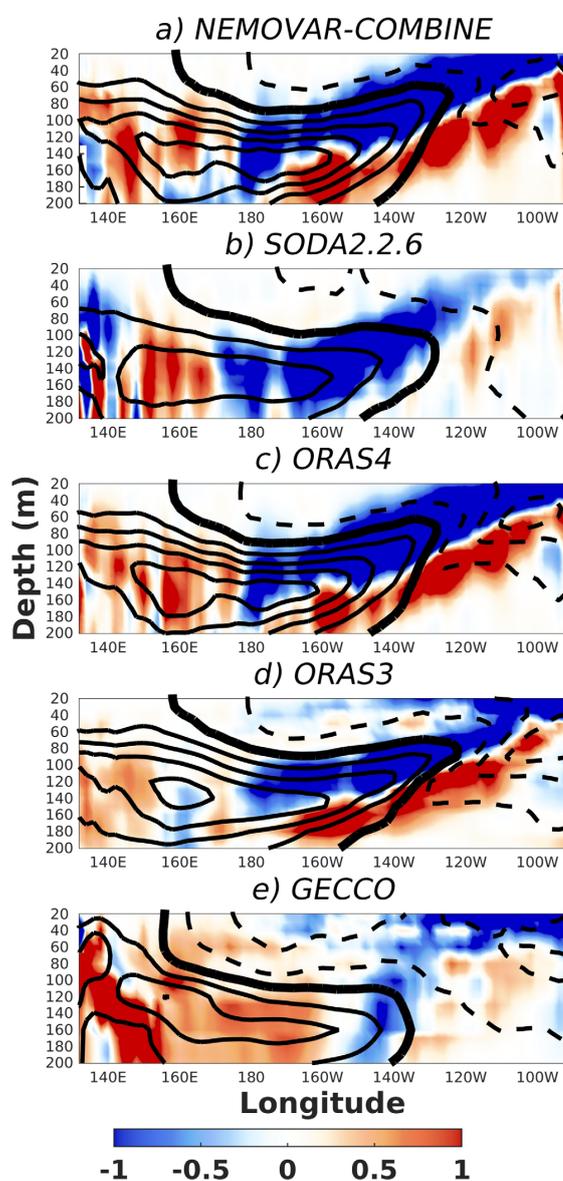
Figure 6. Same as Figure 3h, but for the individual members of the ensemble.

depth composite along the equatorial Pacific (2S–2N, Figure 8), as well as a latitude-depth meridional transect in the central Pacific (160–150W, Figure 9). This transect corresponds to the central part of the area of maximum warming rate (i.e., largest temperature tendency anomalies) and meridional mass exchange between the equatorial plane and the off-equatorial bands. Despite some minor residual, the sum of the zonal, meridional and vertical advection terms explains the general basin-wide subsurface warming near the thermocline (cf. shading and contours in Figure 8b).

Similarly to what seen at earlier lags, in this stage of the ENSO cycle the most important contribution to the subsurface warming in the equatorial central Pacific results once again from both zonal and vertical heat advection terms (Figure 8c). While the intensification of the EUC is confined to 160–120W and is weaker than in the previous phase, heat advection anomalies are largely explained by the strong eastward and upward gradient of anomalous temperature that characterizes the peak of the LN-like conditions (Figure 8a). Further decomposition of these terms reveals that $-(\bar{u}\partial/\partial x + \bar{w}\partial/\partial z)\theta'$ (Figure 8d) and $-(u'\partial/\partial x + w'\partial/\partial z)\bar{\theta}$ (Figure 8e) have opposite signs, showing how temperature and circulation anomalies have opposing tendencies in the subsurface temperatures in the central Pacific.

The mean advection of temperature anomalies $[-(\bar{u}\partial/\partial x + \bar{w}\partial/\partial z)\theta']$ is largely positive in the whole central Pacific, from 170E to 120W and between 40 and 180 m (Figure 8d). This warm anomaly can be further decomposed in its zonal and vertical contributions, which have large positive values at the level of the thermocline and at depths just above the thermocline, respectively, as a result of the corresponding advection of the subsurface anomalous heat buildup in the western Pacific (Figure 8a). The diapycnal transport is characterized by the intensification of the Ekman-induced upwelling motion in the central Pacific (Figure 8a), which drives cold waters to the surface and explains the negative anomalies in $-w'\partial\bar{\theta}/\partial z$. The zonal component $-u'\partial\bar{\theta}/\partial x$ is positive at the level of the thermocline, due to the weak intensification of the zonal component of the EUC (Figures 8a and 8f). Nevertheless, anomalies in the vertical component dominate, controlling the overall sign of $-(u'\partial/\partial x + w'\partial/\partial z)\bar{\theta}$ (Figure 8e).

Figure 7. Same as Figure 3g, but for the individual members of the ensemble.

While playing a relatively minor role in the central Pacific (cf. Figures 8c and 8g), meridional heat advection along the equatorial thermocline explains a large fraction of the subsurface warming in the eastern Pacific. Advection of mean temperatures by anomalous currents, $-v'\partial\bar{\theta}/\partial y$, and of anomalous temperatures by mean currents, $-\bar{v}\partial\theta'/\partial y$, both contribute to the overall meridional advection pattern. As shown in Figure 9a, in the central Pacific anomalous currents are characterized by a strengthening of the shallow meridional overturning cells, with anomalous surface Ekman divergence, off-equatorial downwelling and subsurface convergence. Thus, the subsurface convergence in meridional currents advects climatological off-equatorial warmer waters to the equator (Figure 1e). Similarly, temperature anomalies show weak warming in the off-equatorial regions near the level of the thermocline (Figure 9a), which are advected to the equator by the climatological shallow meridional overturning circulation (Figure 1e). When considering the near-equatorial band as a whole (e.g., 10S–10N), it is clear from the latitudinal transect that the combined contribution of the zonal (Figure 9c) and vertical (Figure 9e) components dominate over the meridional term in determining the subsurface warming in the recharge mode (cf. Figures 9b and 9d). Results also show that half of the contribution of the meridional advection is explained by the intensification of the ocean circulation, and the other half by the deepening of the off-equatorial thermocline (Figures 8i, 8j, and 9a).

The inter-product comparison reveals that very large differences exist between ensemble members at this intermediate phase of the oscillation. For example, the sum of the zonal and vertical advection terms contributes differently to the temperature tendency in each data set: This contribution is positive in 170E–150W at 100–200 m depth in NEMOVAR-COMBINE and ORAS4, weakly positive in 160E–140W at around 100 m in SODA2.2.6, and positive in the whole equatorial Pacific at 20–140 m in ORAS3 and at 80–160 m in GECCO (Figure 10). These differences primarily arise from the vertical component, given that the diapycnal upwelling of cold subsurface waters across the thermocline is subject to large uncertainties (not shown). The meridional term only differs in magnitude and not in sign or spatial extent (Figure 11). Interestingly, the larger the contribution of the meridional advection to the subsurface warming, the larger the warming tendency near the thermocline level

Figure 8. Multiproduct average of temperature tendency and heat advection during the growing phase of the recharge mode leading to El Niño events. Composite anomalies are averaged over the equatorial band (2S–2N), and for the range of lags between 27 and 16 months before the major El Niño events. (a) The anomalous zonal and vertical velocity (m/s, arrows), the anomalous potential temperature (°C, shading), the climatological 20°C isotherm (black contour) and the anomalous 20°C isotherm (green contour). (b–j) Time tendency of potential temperature (°C/yr, contours) and the heat advection terms specified in the plot titles (°C/yr, shading). The contour interval is 0.25°C/yr, with solid (dashed) lines depicting positive (negative) anomalies. The dark (light) stippling denotes areas where heat advection anomalies have the same sign and magnitude larger than ±0.25°C/yr for all (all but one) the members of the ensemble.

(cf. shading and contours in Figure 11), regardless of the contribution of the zonal and vertical components (Figure 10); this finding highlights the dynamical importance of the meridional term during this phase.
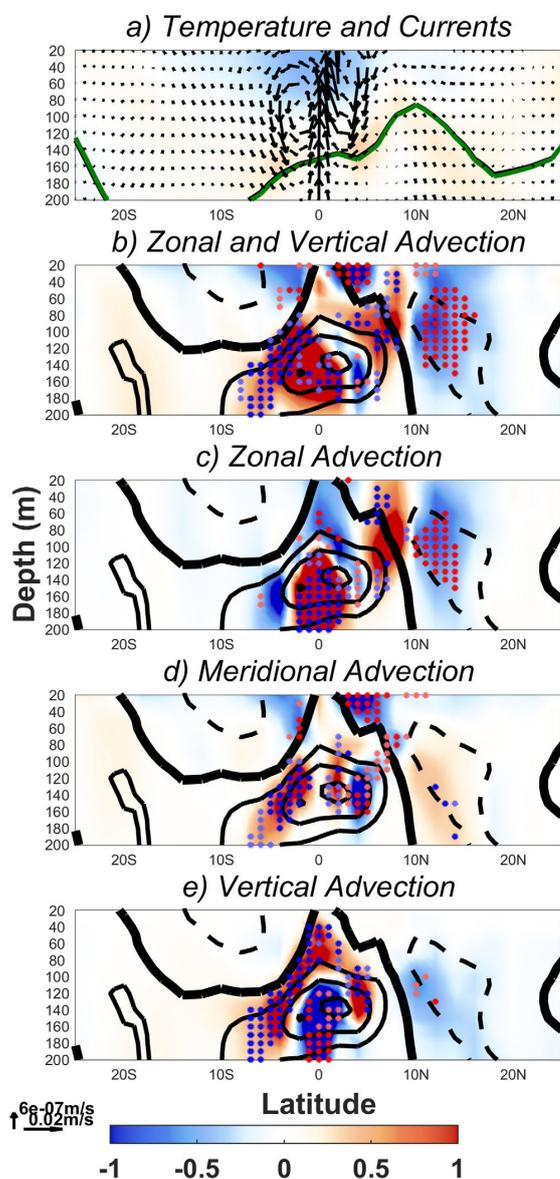
Figure 9. Same as Figure 8a, 8c, and 8f–8h, but for the meridional transect of temperature, meridional and vertical currents and heat advection in the 160–150W sector.

6. Onset of EN Events

Figures 12 and 13 depict composites corresponding to the onset and growing phase of EN events, with averages for lags between 15 and 4 months before EN events, whose tendencies determine the mature phase of EN. The oceanic conditions during this phase are characterized by warm subsurface anomalies along the equatorial thermocline and the weakening (enhancement) of the EUC in the western (eastern) Pacific (Figure 12a). The SEC, the downwelling motion in the warm pool and the coastal upwelling in the far eastern Pacific are also found to be weaker than in the climatology (Figure 12a). These anomalies are known to be associated with the tendency toward warm EN conditions in the central and eastern Pacific, including the beginning stage of weakened trade winds, the flattening of the equatorial thermocline and the development of a subsurface cold buildup in the western Pacific (contours in Figure 12). The role of advective processes is here briefly revisited through the longitude-depth composite along the equatorial Pacific (2S–2N, Figure 12) and the latitude-depth meridional transect in the central Pacific (160–150W, Figure 13).

Similarly to the previous phases, the combination of the three advection terms is in good agreement with the magnitude and spatial structure of the tendency in subsurface temperature (cf. shading and contours in Figure 12e). Note that this correspondence is again primarily explained by the combined contribution of the zonal and vertical advection (Figure 12b). On the one hand, the advection

of climatological temperature by anomalous zonal currents (i.e., $-u'\partial\bar{\theta}/\partial x$) is well known to be largely responsible for the warming tendency in the central and eastern upper ocean (zonal advection feedback [*An and Jin*, 2001], see Figure 12f as a reference). On the other hand, the advection of climatological temperature by anomalous vertical currents ($-w'\partial\bar{\theta}/\partial z$) has been described as a fundamental process for the warming in the far eastern Pacific (Ekman pumping feedback [*Jin et al.*, 2006, Figure 12h]).

The meridional transect in the central Pacific shows that the largest heat anomalies are confined to the latitudinal range 5S–5N, between 100 m and the thermocline level (Figure 13a). The vertical velocity anomalies are characterized by strong upwelling north of the equator (6N–9N) and strong downwelling south of it (4S–1S) (Figure 13a). Thus, anomalies in the northern hemisphere tend to restore the thermocline to its

Figure 10. Same as Figure 8c, but for the individual members of the ensemble.

climatological depth, while those in the southern hemisphere contribute to the deepening of the thermocline near the equator by intensifying the northernmost edge of the downwelling branch of the southern shallow meridional overturning cell (Figure 1e). Interestingly, inter-hemispheric differences in vertical velocity anomalies increase the meridional tilting of the thermocline and generate southward cross-equatorial mass transport in the upper 50 m of the ocean (Figure 13a).

The meridional heat advection is negative in the central and eastern Pacific near the level of the thermocline (Figure 12g). As a result, this component starts contributing to the weakening of the heat content in the equatorial Pacific subsurface already in the recharge phase (Figure 12a), before the onset of EN and the activation of the Bjerknes feedback. The decomposition of this term shows that $-\bar{v}\partial\theta'/\partial y$ is larger and has opposite sign relative to $-v'\partial\bar{\theta}/\partial y$ (cf. Figures 12c and 12d). Indeed, $-\bar{v}\partial\theta'/\partial y$ $(-v'\partial\bar{\theta}/\partial y)$ shows large negative (weak positive) anomalies in the off-equatorial regions (Figures 13c and 13d), near the areas of climatological (anomalous) subsurface equatorward convergence and strong meridional contrast in anomalous (climatological) temperature (Figures 1e and 13a).

7. Discussion

Figure 14 provides an integrated view of the ENSO oscillation by showing the contribution of the processes described throughout the manuscript to the equatorial temperature tendency at the thermocline level and as a function of the time lag. This comparison highlights the generally synchronous evolution of the temperature tendency and the advection along the tilted equatorial thermocline (cf. shading and contours in Figure 14c). In the central Pacific (170E–110W), this tendency is to a large extent explained by the advection of the initial subsurface warm buildup in the western Pacific by the climatological EUC (i.e., $-(\bar{u}\partial/\partial x + \bar{w}\partial/\partial z)\theta'$, Figure 14d). Instead, the anomalous downwelling motion of surface climatological warm waters $(-w'\partial\bar{\theta}/\partial z)$ is the key process explaining the evolution of the subsurface warming in the western (130–170E) and far eastern (110–90W) Pacific (Figure 14e). Note that the transition between advective processes near 170E is explained by a clear regime shift in the vertical structure of horizontal divergence [*Ballester et al.*, 2015].

Figure 11. Same as Figure 8g, but for the individual members of the ensemble.

The latitudinal heat advection structure is illustrated in the meridional transect in the central Pacific shown in Figure 15. The contribution of the meridional term is negligible right at the equator, where the Ekman-induced upwelling motion dominates, but it rapidly increases poleward, attaining anomalies that are already large at 2S and 2N (Figure 15g). Here we considered a narrow equatorial band (i.e., 2S–2N) in order to isolate the ascending branch of the shallow meridional overturning cells from the descending branches at around 8–3S and 3–8N (Figure 1e). Note that this circulation is associated with a cross-shaped anomalous pattern in both the zonal and vertical advection terms (Figures 15f and 15h), which is not found in their combined contribution (Figure 15c). The meridional advection remains qualitatively unmodified when the latitudinal range considered for the equatorial averages in the longitude-depth plane includes the descending branches of the cells (e.g., 6S–6N, not shown).

The peak in LN (EN) conditions is approximately in phase with the equatorward meridional warm (cold) advection in the central and eastern Pacific (Figure 14g). The phase of the oscillation in which the contribution of meridional advection reaches its peak is however the result of the combination of two different processes with different temporal evolution. On the one hand, the discharge (recharge) phase in basin-wide equatorial heat content leads to LN (EN) events by approximately 9 months [*Meinen and McPhaden*, 2000]. This phase is characterized by colder (warmer) temperature anomalies at the equator than in the off-equatorial regions (e.g., Figure 13a). Thus, the meridional circulation of the climatological shallow meridional overturning cells warms (cools) the equatorial and off-equatorial thermocline before the mature phase of LN (EN) events ($-\bar{v}\partial\theta'/\partial y$; Figures 14i and 15i). On the other hand, the strengthening (weakening) of the equatorial trade winds during LN (EN) conditions, as well as the associated changes in off-equatorial wind stress curl, induces anomalous subsurface equatorward (poleward) Sverdrup transport of mass [*Jin*, 1997a,1997b]. The disequilibrium balance between these processes generates a delayed warming (cooling) at the level of the thermocline ($-v'\partial\bar{\theta}/\partial y$; Figures 14j and 15j).

In this regard, the synchronous evolution of LN (EN) conditions and the equatorward warm (cold) advection in the central and eastern Pacific, as illustrated in Figure 14g, is shown to be compatible

Figure 12. Multiproduct average of temperature tendency and heat advection during the growing phase of El Niño events. Composite anomalies are averaged over the equatorial band (2S–2N), and for the range of lags between 15 and 4 months before the major El Niño events. (a) The anomalous zonal and vertical velocity (m/s, arrows), the anomalous potential temperature (°C, shading), the climatological 20°C isotherm (black contour) and the anomalous 20°C isotherm (green contour). (b–h) Time tendency of potential temperature (°C/yr, contours) and the heat advection terms specified in the plot titles (°C/yr, shading). The contour interval is 0.25°C/yr, with solid (dashed) lines depicting positive (negative) anomalies. The dark (light) stippling denotes areas where heat advection anomalies have the same sign and magnitude larger than ±0.25°C/yr for all (all but one) the members of the ensemble.

with the recharge theory formulated by Jin, which is mathematically described by the tilting mode and the recharge-discharge phase. The tilting mode characterizes the quick oceanic response to enhanced (weakened) easterly wind stress in the central tropical Pacific during LN (EN) conditions, which is proportional to the zonal tilting of the equatorial thermocline. The recharge (discharge) phase provides the required memory between opposite phases of the tilting mode. This transition period is characterized by the time tendency toward anomalous equatorward (poleward) Sverdrup convergence (divergence) of mass due to enhanced (weakened) easterly wind stress in the western and central tropical Pacific, and its associated change in the off-equatorial curl, which ultimately tends to deepen the thermocline.

The recharge theory is also found to be compatible with the longitudinal transition in the mechanisms explaining the initial subsurface heat buildup on either side of 170E, as well as the subsequent eastward propagation along the equatorial thermocline. Near and east of the dateline, anomalous easterly (westerly)
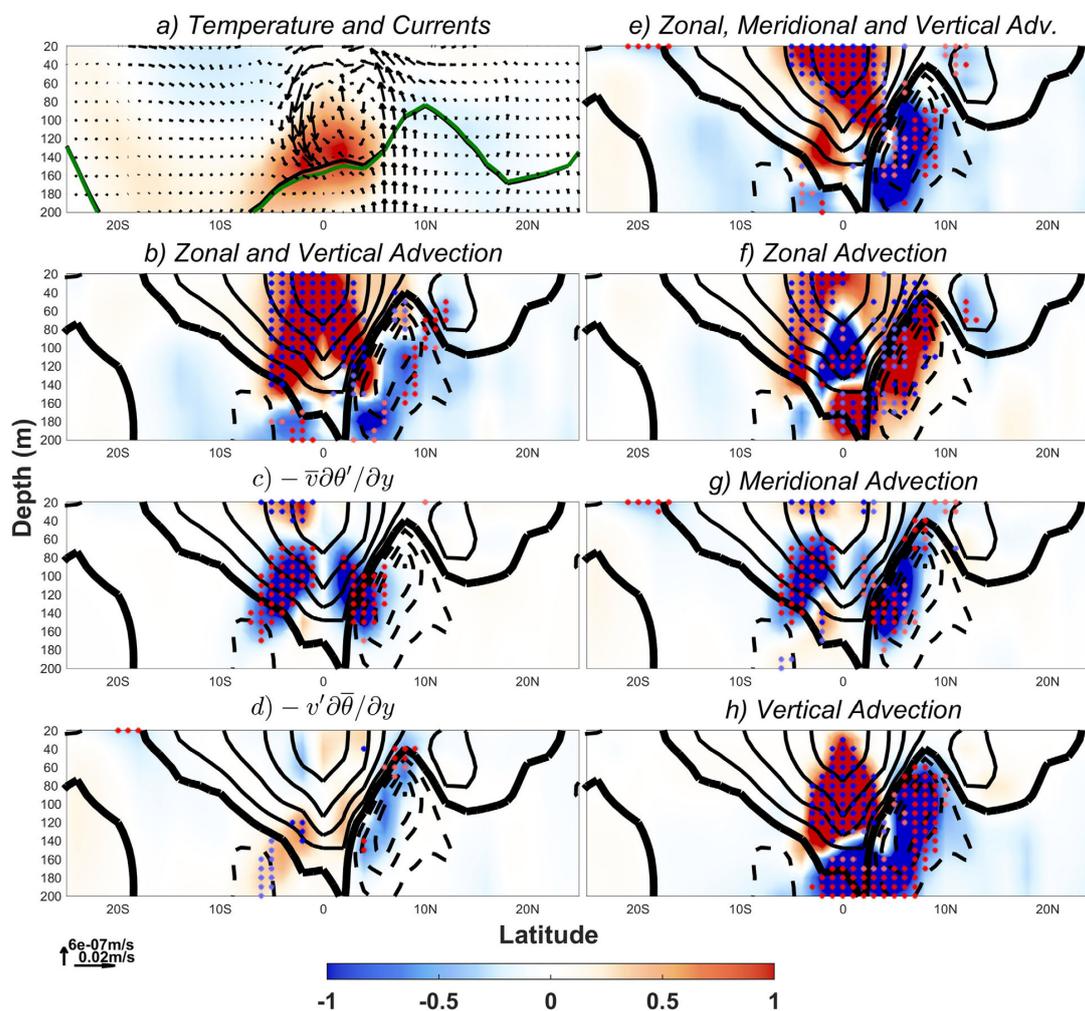
Figure 13. Same as Figure 12, but for the meridional transect of temperature, meridional, and vertical currents and heat advection in the 160–150W sector.

trade winds during LN (EN) events are associated with the tendency toward equatorward (poleward) Sverdrup mass convergence (divergence) and the deepening of the thermocline (e.g., Figure 8j). Near the edge of the warm pool, easterly (westerly) wind stress anomalies and the associated anticyclonic (cyclonic) curl anomalies are weaker, and therefore this delayed effect is smaller (Figure 8j). The oceanic response in this region appears to be more directly controlled by the zonal convergence (divergence) of the zonal wind stress along the equator, which favors anomalous surface ocean horizontal convergence (divergence) and downward (upward) motion during LN (EN) events [*Ballester et al.*, 2015; Figures 3h and 4e]. This process explains the much faster, albeit still somewhat delayed, response of subsurface temperatures in the warm pool (e.g., the zero contour in Figure 14e crosses longitude 160E at lag +03). The present article clarifies, within the context of the recharge oscillator theory, the relative contribution, spatial extent and delayed effect of each of the mechanisms involved in the subsurface buildup in the western and central Pacific, and its eastward propagation.

Finally, it is worth noting how relatively large agreement exists between assimilation products during the three phases of the oscillation. Differences between data sets are generally small, and anomalies tend to be similarly signed (see Figures 6–10, and 11). The only exception is GECCO, which does not produce the
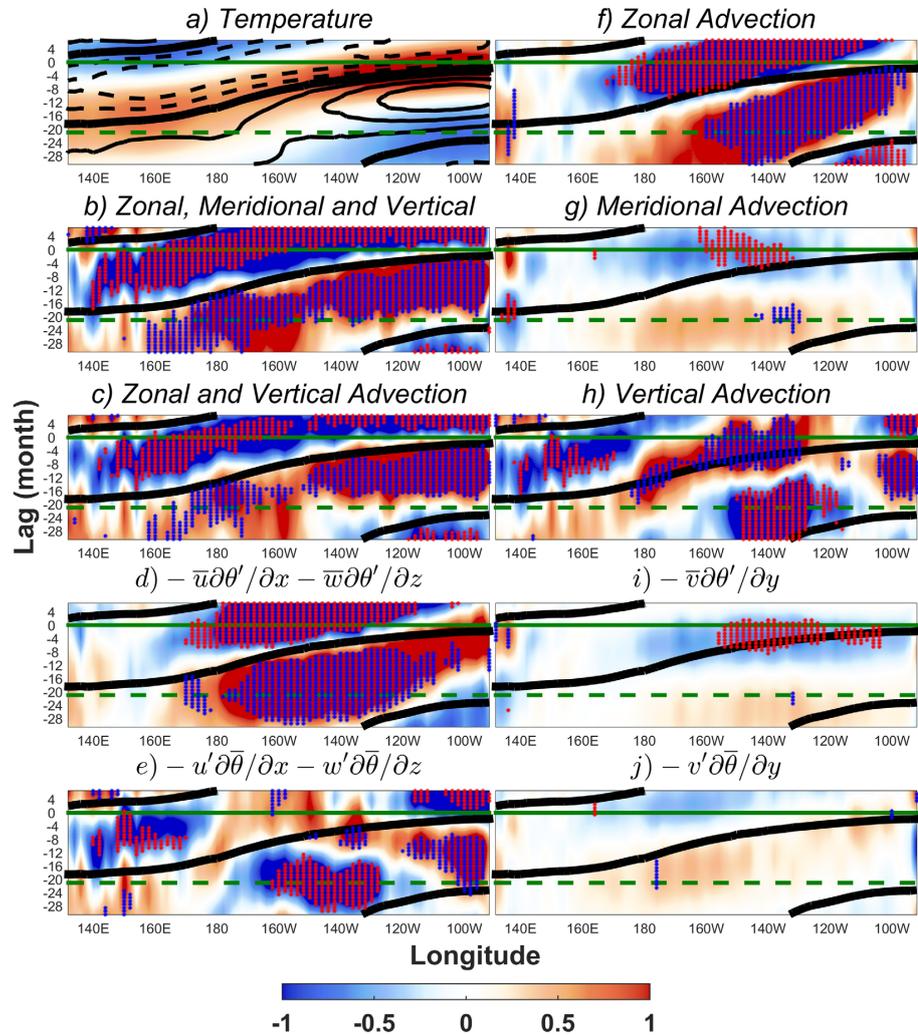
Figure 14. Multiproduct average of temperature tendency and heat advection before and after the peak of El Niño events. Composite anomalies are averaged over the equatorial band (2S–2N) at the depth of the tilted thermocline. (a) The anomalous potential temperature (°C, shading) and its time tendency (°C/yr, contours). The contour interval is 0.50°C/yr, with solid (dashed) lines depicting positive (negative) anomalies. (b–j) Zero contour of the time tendency of potential temperature, as well as the heat advection terms specified in the plot titles (°C/yr, shading). The solid (dashed) horizontal green line shows the phase corresponding to the peak of EN events (LN-like conditions). The stippling denotes anomalies in which heat advection has the same sign and magnitude larger than ±0.25°C/yr for all the members of the ensemble.

patterns of meridional surface Ekman divergence (subsurface Sverdrup convergence) in the western and central equatorial Pacific observed in the other assimilation data sets, therefore featuring the weakest subsurface heat buildup. The relatively coarser resolution of this product (1° × 1°) might partially explain some of these differences. More importantly, this assimilation data set is obtained by fitting the model simultaneously to all available data over the whole 50 year period, iteratively for 23 iterations by first running the forward model to calculate the model data misfit formulated as a cost function, followed by an adjoint model run to calculate the gradients of this cost function [*Köhl and Stammer*, 2008]. Given the relatively large computational demands of this approach, the optimization did not fully converge after this number of iterations. Although this solution was found to be good enough to investigate the underlying processes and causes of other phenomena such as the Meridional Overturning Circulation [*Köhl and Stammer*, 2008], it
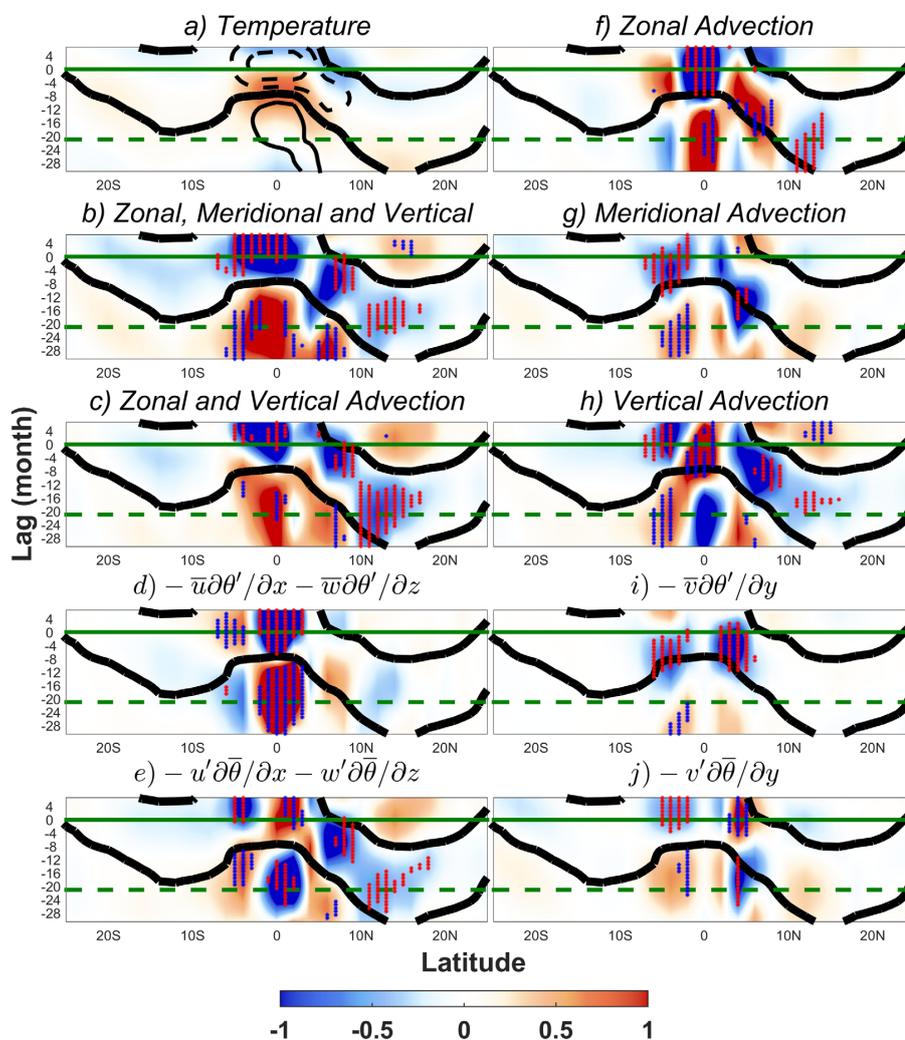
Figure 15. Same as Figure 14, but along the latitudinal axis and averaged over the longitude range 160–150W.

seems that the model was not brought into a good level of agreement with the data regarding those processes that are fundamental to the generation and growth of EN events.

Mechanisms described in the present work can be used as a reference for the validation of numerical simulations from intermediate and complex coupled climate models. For example, a key feature highlighted here is the asymmetry between the northern and the southern tropical hemispheres during the recharge mode. This phase is characterized by an increase of the meridional tilting of the thermocline and the southward cross-equatorial mass transport in the upper ocean as a result of the anomalous upwelling (downwelling) motion in 6–8N (3–1S). *Yu and Mechoso* [2001] showed that the anomalies in vertical velocity are due to the latitudinal distribution of zonal wind stress anomalies, which induces areas of convergence and divergence of meridional Ekman transport. Nonetheless, the climate model used by *Yu and Mechoso* [2001] simulated vertical anomalies of equal sign in both hemispheres and no cross-equatorial anomalies between the off-equatorial regions, in disagreement with our results. This highlights how the present work provides a description of dynamical processes that the climate modeling community might use as metrics to test the performance of the ENSO oscillation in state-of-the-art climate models.

AGU **Journal of Geophysical Research: Oceans**   10.1002/2016JC011718

8. Summary

The present work describes different processes that control subsurface temperatures and thermocline depth during the generation of El Niño events through a careful analysis of the subsurface heat budget. We emphasize the role that different processes play in the evolution of subsurface warm anomalies during the different stages of the oscillation. Main results include:

1. To the west of 170E, the vertical advection of climatological temperature by anomalous currents, induced by surface horizontal convergence, downwelling motion and subsurface divergence, was shown to explain alone the initial subsurface warming in the equatorial and off-equatorial Pacific during the warm buildup stage, between monthly lags -33 to −21 before the peak of EN events.
2. The role of horizontal advection was found to be confined to the east of 170E, explaining the tendency toward the return to climatological conditions of subsurface temperatures in the central Pacific, both through zonal and vertical advection along the equatorial thermocline and through meridional advection right above this level.
3. These two mechanisms were also shown to explain a large fraction of the subsurface warming associated with the recharge phase in basin-wide heat content. On the one hand, along the meridional axis, the equatorward advection of heat was shown to be explained to the same extent by anomalies in the meridional gradient of subsurface temperature and anomalies in the meridional ocean circulation. On the other hand, along the equatorial plane (i.e., combination of the zonal and vertical components), the anomalous heat accumulated in the western Pacific was seen to be advected to the central Pacific by the climatological currents. This contribution was found to be partially counterbalanced by the advection of climatological temperature by the anomalous currents, which is dominated by anomalous diapycnal upwelling of cold subsurface waters.
4. The large differences in the magnitude of the meridional term were seen to be key for explaining the different propagation speed of the subsurface warming tendency along the thermocline.

All terms in this analysis are inferred from an ensemble of state-of-the-art ocean assimilation products, focusing on those processes that are robustly produced by all the members of the ensemble, as well as those that are differently simulated by a subset of data sets. The combined use of multiple ocean analysis products provides a reference for a three-dimensional description of mechanisms leading to the generation of EN events. Additionally, it allows for a more detailed validation and assessment of mechanisms previously inferred from intermediate and complex coupled climate models, as well as for the determination of the limits in the use of assimilation products for the validation itself.

References

An, S. I., and F. F. Jin (2001), Collective role of thermocline and zonal advective feedbacks in the ENSO mode, *J. Clim.*, *14*, 3421–3432.
Ballester, J., M. A. Rodríguez-Arias, and X. Rodó (2011), A new extratropical tracer describing the role of the western Pacific in the onset of El Niño: Implications for ENSO understanding and forecasting, *J. Clim.*, *24*, 1425–1437.
Ballester, J., J. C. Burns, D. Cayan, Y. Nakamura, R. Uehara, and X. Rodó (2013), Kawasaki disease and ENSO-driven wind circulation, *Geophys. Res. Lett.*, *40*, 2284–2289, doi:10.1002/grl.50388.
Ballester, J., S. Bordoni, D. Petrova, and X. Rodó (2015), On the dynamical mechanisms explaining the western Pacific subsurface temperature buildup leading to ENSO events, *Geophys. Res. Lett.*, *42*, 2961–2967, doi:10.1002/2015GL063701.
Ballester, J., S. Bordoni, D. Petrova, and X. Rodó (2016),El Niño dynamics and long lead climate forecasts, *Climanosco*, in press.
Balmaseda, M. A., A. Vidard, and D. L. T. Anderson (2008), The ECMWF Ocean Analysis System: ORA-S3, *Mon. Weather Rev.*, *136*, 3018–3034.
Balmaseda, M. A., K. Mogensen, F. Molteni, and A. T. Weaver (2010), The NEMOVAR-COMBINE ocean re-analysis, *COMBINE Tech. Rep. 1.* [Available at http://www.combine-project.eu/fileadmin/user_upload/combine/tech_report/COMBINE_TECH_REP_n01.pdf.]
Balmaseda, M. A., K. Mogensen and, A. T. (2013), Weaver evaluation of the ECMWF ocean reanalysis system ORAS4. *Q. J. R. Meteorol. Soc.* *139*, 1132–1161.
Battisti, D. S. (1988), Dynamics and thermodynamics of a warming event in a coupled tropical atmosphere-ocean model, *J. Atmos. Sci.*, *45*, 2889–2919.
Bjerknes, J. (1969), Atmospheric teleconnections from the equatorial Pacific, *Mon. Weather Rev.*, *97*, 163–172.
Brown, J. N., and A. V. Fedorov (2010), How much energy is transferred from the winds to the thermocline on ENSO time scales?, *J. Clim.*, *23*, 1563–1580.
Brown, J. N., A. V. Fedorov, and E. Guilyardi (2011), How well do coupled models replicate ocean energetics relevant to ENSO?, *Clim. Dyn.*, *36*, 2147–2158.
Carton, J. A., and B. S. Giese (2008), A reanalysis of ocean climate using simple ocean data assimilation (SODA), *Mon. Weather Rev.*, *136*, 2999–3017.
CPC (2015), Cold and Warm Episodes by Season. [Available at http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml, last accessed Jan. 2015.]
Fedorov, A. V., and J. Brown (2009), Equatorial waves, in *Encyclopedia of Ocean Sciences*, 2nd ed., edited by J. Steele, pp. 3679–3695, Academic Press, Elsevier Science, U. K.

Hsin, Y. C., and B. Qiu (2012), The impact of Eastern-Pacific versus Central-Pacific El Niños on the North Equatorial Countercurrent in the Pacific Ocean, *J. Geophys. Res.*, *117*, C11017, doi:10.1029/2012JC008362.

Izumo, T. (2005), The equatorial undercurrent, meridional overturning circulation, and their roles in mass and heat exchanges during El Niño events in the tropical Pacific ocean, *Ocean Dyn.*, *55*, 110–123.

Jin, F. F. (1997a), An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model, *J. Atmos. Sci.*, *54*, 811–829.

Jin, F. F. (1997b), An equatorial ocean recharge paradigm for ENSO. Part II: A stripped-down coupled model, *J. Atmos. Sci.*, *54*, 830–847.

Jin, F. F., and S. I. An (1999), An thermocline and zonal advective feedbacks within the equatorial ocean recharge oscillator model for ENSO, *Geophys. Res. Lett.*, *26*, 2989–2992.

Jin, F. F., and D. J. Neelin (1993), Modes of Interannual Tropical Ocean–Atmosphere Interaction—A Unified View. Part I: Numerical Results, *J. Atmos. Sci.*, *50*, 3477–3503.

Jin, F. F., S. T. Kim, and L. Bejarano (2006), A coupled-stability index for ENSO, *Geophys. Res. Lett.*, *33*, L23708, doi:10.1029/2006GL027221.

Köhl, A., and D. Stammer (2008), Variability of the meridional overturning in the North Atlantic from the 50-year GECCO state estimation, *J. Phys. Oceanogr.*, *38*, 1913–1930.

Meinen, C. S., and M. J. McPhaden (2000), Observations of warm water volume changes in the equatorial Pacific and their relationship to El Niño and La Niña, *J. Clim.*, *13*, 3551–3559.

Petrova, D., S. J. Koopman, J. Ballester, and X. Rodó (2016), Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model, *Clim. Dyn.*, doi:10.1007/s00382-016-3139-y.

Ray, S., D. Swingedouw, J. Mignot, and E. Guilyardi (2015), Effect of surface restoring on subsurface variability in a climate model during 1949-2005, *Clim. Dyn.*, *44*, 2333–2349.

Santoso, A., S. McGregor, F. F. Jin, W. Cai, M. H. England, S. I. An, M. J. McPhaden, and E. Guilyardi (2013), Late-twentieth-century emergence of the El Niño propagation asymmetry and future projections, *Nature*, *504*, 126–130.

Schopf, P. S., and M. J. Suarez (1988), Vacillations in a coupled ocean-atmosphere model, *J. Atmos. Sci.*, *45*, 549–566.

Singh, A., and T. Delcroix (2013), Eastern and Central Pacific ENSO and their relationships to the recharge/discharge oscillator paradigm, *Deep Sea Res., Part I*, *82*, 32–43.

Trenberth, K. E., and T. J. Hoar (1996), The 1990-1995 El Niño-Southern Oscillation Event: Longest on Record, *Geophys. Res. Lett.*, *23*, 57–60.

Wang, B. (2002), Kelvin waves, in *Encyclopedia of Atmospheric Sciences*, edited by M. Shankar, pp. 1062–1068, Elsevier, Amsterdam.

Wyrtki, K. (1975), El Niño—The dynamic response of the equatorial Pacific Oceanto atmospheric forcing, *J. Phys. Oceanogr.*, *5*, 572–584.

Yu, J. Y., and C. R. Mechoso (2001), A coupled atmosphere–ocean GCM study of the ENSO cycle, *J. Clim.*, *14*, 2329–2350.

Yu, Z., J. P. McCreary, W. S. Kessler, and K. A. Kelly (2000), Influence of equatorial dynamics on the Pacific North Equatorial Countercurrent, *J. Phys. Oceanogr.*, *30*, 3179–3190.

3.4   Sensitivity of El Niño intensity and timing to preceding subsurface heat magnitude

Ballester J (1,2), Petrova D (1), Bordoni S (2), Cash B (3), García-Díez M (1), Rodó X (1,4)

(1) Institut Català de Ciències del Clima, Barcelona, Catalonia, Spain
(2) California Institute of Technology, Pasadena, California, United States
(3) George Mason University, Fairfax, Virginia, United States
(4) Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

Resumen - *Summary in Spanish*

A pesar de los extensos esfuerzos en curso para mejorar la predicción a largo plazo de El Niño-Oscilación Sur, la previsibilidad en los sistemas operativos de vanguardia sigue siendo limitada por factores tales como la barrera de primavera y la influencia de los vientos atmosféricos. Investigaciones recientes sugieren que el evento El Niño (EN) 2014/15 se estancó como resultado de una explosión de viento del este en toda la cuenca en junio, lo que provocó la descarga de una gran parte del calor oceánico subsuperficial. Aquí utilizamos registros observacionales y experimentos numéricos para explorar la sensibilidad de EN a la magnitud de la acumulación de calor que ocurre en el subsuelo del océano 21 meses antes. Nuestras simulaciones sugieren que un gran aumento en el contenido de calor durante esta fase puede conducir a condiciones cálidas uniformes en la cuenca en el Pacífico ecuatorial en el invierno antes de la ocurrencia de un evento EN muy fuerte. En nuestra configuración de modelo, el sistema compensa cualquier disminución inicial en el contenido de calor y naturalmente evoluciona hacia una nueva recarga, dando lugar a un retraso de hasta un año en la ocurrencia de un evento EN. Ambos escenarios justifican la dependencia no lineal entre la intensidad de la acumulación de calor subsuperficial y la magnitud y el momento de los episodios posteriores de EN.

SCIENTIFIC REP⚙RTS

Sensitivity of El Niño intensity and timing to preceding subsurface heat magnitude

Joan Ballester[1,2,†], Desislava Petrova[1,†], Simona Bordoni[2], Ben Cash[3], Markel García-Díez[1] & Xavier Rodó[1,4,†]

Despite extensive ongoing efforts on improving the long-term prediction of El Niño-Southern Oscillation, the predictability in state-of-the-art operational schemes remains limited by factors such as the spring barrier and the influence of atmospheric winds. Recent research suggests that the 2014/15 El Niño (EN) event was stalled as a result of an unusually strong basin-wide easterly wind burst in June, which led to the discharge of a large fraction of the subsurface ocean heat. Here we use observational records and numerical experiments to explore the sensitivity of EN to the magnitude of the heat buildup occurring in the ocean subsurface 21 months in advance. Our simulations suggest that a large increase in heat content during this phase can lead to basin-wide uniform warm conditions in the equatorial Pacific the winter before the occurrence of a very strong EN event. In our model configuration, the system compensates any initial decrease in heat content and naturally evolves towards a new recharge, resulting in a delay of up to one year in the occurrence of an EN event. Both scenarios substantiate the non-linear dependency between the intensity of the subsurface heat buildup and the magnitude and timing of subsequent EN episodes.

El Niño-Southern Oscillation (ENSO) is the dominant mode of interannual variability in the tropical Pacific[1–5], and a major source of climate predictability, large-scale teleconnections, and impacts worldwide[6–9]. After several decades of intensive research, the main mechanisms explaining the dynamics of the phenomenon and the onset of El Niño (EN) and La Niña (LN) events are nowadays thought to be relatively well understood[3,10–12]. Since the eighties, when the first successful prediction of EN was issued[13], steady improvements in the forecast of the phenomenon have led to a plateau at moderate skill at short and medium lead times[14]. Forecasts issued in spring or before are still generally unable to foresee whether an EN or a LN event will occur at the end of the year, the so-called spring barrier[15,16]. Although recent advances suggest that it is indeed possible to overcome this predictability limit, either with intermediate[6], purely dynamical[7] or purely statistical[17] models, the value of a real-time operational scheme derived from these results remains to be fully accomplished.

Mounting evidence has favored the view of ENSO as a slightly damped periodic oscillation modulated by stochastic noise[12,14,18–20], with its long-range potential predictability arising from the dynamics of the ocean subsurface[3,21]. The recharge oscillator[10–11] is currently the leading paradigm for the link between the Tropical Heat Content (THC) in the ocean subsurface and ENSO variability. This theory is based on the delay between anomalies in the longitudinally-averaged depth of the thermocline, and therefore the THC of the basin, and eastern Pacific Sea Surface Temperatures (SST). This paradigm, which has been thoroughly validated in the literature[22,23], describes a recharge (discharge) phase of Warm Water Volume (WWV) along the equator that leads EN (LN) events by about 2 to 3 seasons[21,24]. This stage is in turn preceded by the tilting mode[2,25,26], which is characterized by an anomalous zonal gradient of the depth of the thermocline and is significantly associated with ENSO at long-lead times (e.g. the correlation between ENSO and WWV anomalies to the west of 155W is around 0.57 at lag 15 months, see ref. 3).

This description of the oscillation emphasizes the importance of the THC at different lead times for the growth and magnitude of subsequent EN events. Reference 27 found that the duration of a complete EN cycle is determined by the time required for the slow accumulation of warm water in the western Pacific. Reference 28 showed

[1]Institut Català de Ciències del Clima (IC3), Barcelona, Catalonia, Spain. [2]California Institute of Technology (Caltech), Pasadena, California, United States. [3]George Mason University, Fairfax, Virginia, United States. [4]ICREA, Barcelona, Catalonia, Spain. [†]Present address: ISGlobal, Barcelona Institute for Global Health, Barcelona, Catalonia, Spain. Correspondence and requests for materials should be addressed to J.B. (email: joanballester@ic3.cat)
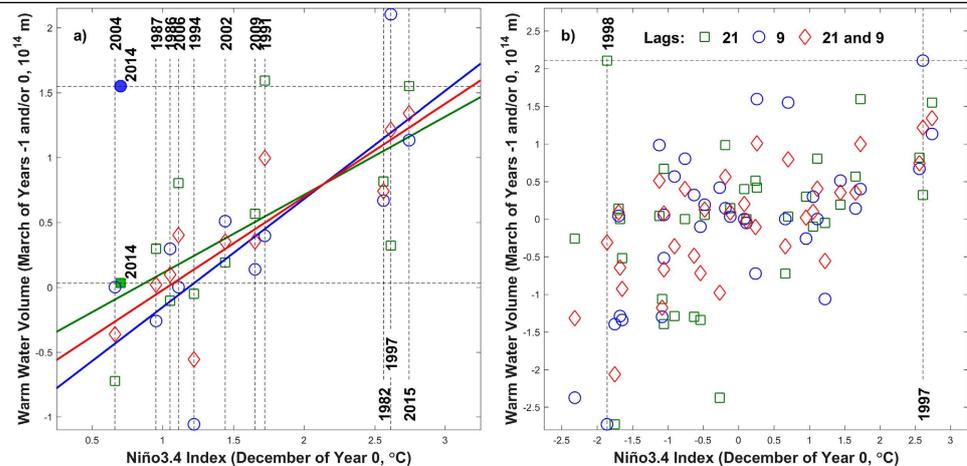
Figure 1. Observed relationship between the December anomaly of the Niño3.4 Index (in °C) and leading Warm Water Volume anomalies (10^{14} m) in March of the same year (blue circles), March of the preceding year (green squares), and the average of both years (red diamonds). Episodes in panel a correspond to all El Niño events since 1980 according to the NOAA's Climate Prediction Center (vertical dashed lines, *33*), with colored straight lines depicting the linear fitting, while panel b shows values for all years in the 1980–2015 period. The 21-month (9-month) lag relationship corresponding to the neutral year of 2014 is additionally shown as a solid square (circle) in panel a. Warm Water Volume was computed as the volume of water masses above the 20 °C isotherm within 120E-80W and 5S-5N.

that the onset of EN usually begins during the second half of the year before the event, being identified as a fundamental process independent of the flavor of the episodes. A buildup in heat content along the equator has indeed preceded all the major EN events since 1980, and the magnitude of EN usually scales in proportion to the magnitude of the heat content buildup 2 to 3 seasons in advance[3,29]. Reference 30 used ocean-atmosphere coupled simulations to show that the prescription of a warm heat content anomaly immediately before the spring barrier can lead to the generation of a moderate EN event. These same simulations also showed that prescribed westerly wind bursts alone do not lead to significant EN anomalies, but they can instead greatly amplify heat content anomalies and generate a strong EN event when they are superimposed to an initially recharged ocean state[30]. Nonetheless, results also suggest that the heat content buildup may be a necessary but not sufficient condition for EN to occur[3]. For example, ref. 31 have recently shown that the 2014/15 EN event was stalled as a result of an unusually strong basin-wide easterly wind burst in June 2014, which discharged the basin, suppressed the Bjerknes feedback and impeded the growth of the strong episode that was expected for the end of the year. At that time, basin-wide uniform warm conditions were instead observed in the equatorial Pacific, which were followed by the record-breaking EN episode in boreal winter 2015/16.

To illustrate the importance of the THC at different lead times, Fig. 1a depicts the relationship between the December anomaly of the Niño3.4 (N34) Index and leading WWV anomalies[32] in March of the same and/or the preceding year (here referred to as years 0 and −1, respectively) for all the EN episodes since 1980 (see ref. 33 for the classification of events). WWV is defined as the volume of water masses above the 20 °C isotherm within 120E-80W and 5S-5N. This figure confirms the widely accepted and well understood 9-month lead association between high WWV anomalies and EN events (blue circles), but importantly, it also shows that the same relationship holds for WWV anomalies one year earlier (i.e. 21 months before EN, green squares). Both relationships are strong, with correlations of 0.75 and 0.63 respectively, which increase to 0.85 ($r^2 = 72\%$) when March WWV anomalies are averaged for both years (red diamonds). Note that this double relationship is in general also valid for all years since 1980 (Fig. 1b), with only one major exception: WWV was largely positive 21 months before the 1998 LN event, given that it was preceded by the very strong 1997 EN episode and its associated 9-month leading recharge phase in spring 1997 (dashed lines in Fig. 1b). In this way, when the 1998 LN event is not taken into account, the correlation between the December N34 Index and 21-month leading WWV anomalies is equal to 0.55 for all the years since 1980 (green squares in Fig. 1b).

Given the strong relationship observed between this very long-lead heat buildup and subsequent EN events here we use a state-of-the-art Earth System Model to explore the response to a decrease or increase in the magnitude of the heat content stored in the ocean subsurface (see Methods). We performed 11 sets of ensemble experiments, with initial conditions corresponding to an early phase of the onset of an EN episode in March of year −1 (i.e. lead time of 21 months). Previous studies have explored this relationship at shorter lead times of up to one year, through statistical analyses of observational data (e.g. ref. 3) or through numerical experiments (e.g. ref. 30). As we wanted to specifically study the dynamics of very strong EN events, we prescribed anomalous conditions to mimic as close as possible an episode of magnitude similar to the recent 2015/16 event (i.e. N34 = +2.8 °C). In each of these sets of ensemble experiments, the intensity of the subsurface warm anomaly was decreased (negative sign representing a discharge in heat content) or increased (positive sign or recharge) by ±20%, ±40%, ±60%,
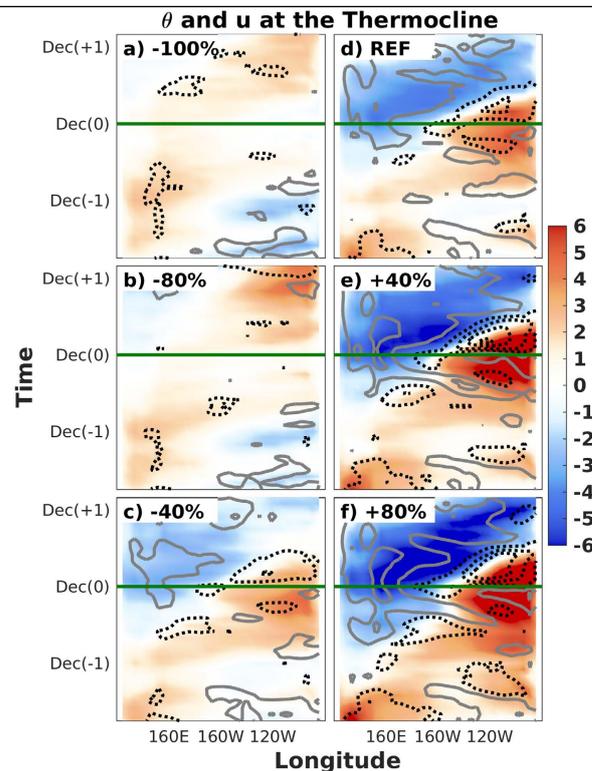
Figure 2. Longitude-time Hovmöller diagram of equatorial temperature (shading, in °C) and zonal current (contour, in m/s) anomalies at the level of the thermocline. Panels correspond to the −100% (**a**), −80% (**b**), −40% (**c**), REF (**d**), +40% (**e**) and +80% (**f**) experiments. The minimum contour is ±0.1 m/s and the contour interval is 0.2 m/s, with grey solid (black dashed) lines depicting positive (negative) anomalies. The horizontal green lines indicate the December month in which El Niño peaks in the reference simulation.

±80% and ±100% (Supplementary Figure 1a,c,d,f) relative to the unmodified reference (REF) simulation ensemble (Supplementary Figure 1b,e). As such, differences among experiments are explained by both the magnitude of the initial subsurface heat content and the strong coupling between the ocean and the atmosphere that characterizes the dynamics of ENSO. We note that unlike in ref. 30 we only prescribe anomalies in the ocean subsurface, and therefore the atmosphere is only indirectly modified when the readjustment of the ocean affects the ocean surface and the interaction between the ocean and the atmosphere (see Supplementary Figures 2–4).

Figure 2 shows the longitude-time Hovmöller diagram of equatorial potential temperature and zonal current anomalies at the level of the thermocline for experiments representative of the different types of ocean responses to the prescribed anomalies, Fig. 3 shows the corresponding anomalies of equatorial SST and zonal wind stress, and Fig. 4 the anomalies of equatorial Sea Surface Height (SSH) and surface zonal currents. As additional information, the Supplementary Material includes the vertical profiles and horizontal maps of these variables for the time steps in which the recharge (Supplementary Figures 5–7), EN (Supplementary Figures 8–10), discharge (Supplementary Figures 11–13) and LN (Supplementary Figures 14–16) phases of the oscillation are observed in the REF ensemble.

The REF ensemble is found to correctly reproduce the main features of a canonical ENSO oscillation (Figs 2–4d). It is initially characterized by easterly wind and cold SST anomalies in the central Pacific and the generation of the subsurface heat buildup in the western Pacific, which peaks in spring of year −1 (i.e. beginning of the simulations). Reference 26 showed that meridional and eastward heat advection due to equatorward subsurface mass convergence and transport along the equatorial undercurrent contribute to this long-leading subsurface warming at 170E-150W, while surface horizontal convergence and downwelling motion have a leading role in subsurface warming in the warm pool. Westerly wind anomalies appear at the beginning of the following year, when the warm waters start to propagate to the eastern Pacific along the equatorial thermocline as downwelling Kelvin waves. Some few months later, in spring of year 0, the warm anomalies reach the eastern Pacific subsurface during the basin-wide recharge phase of the ENSO oscillation, which is immediately followed by the beginning of the warming of the ocean surface. Equatorial SST anomalies exceed the +1 °C threshold in the central and eastern equatorial Pacific between the summers of this and the following year (i.e. year +1). The eastward surface current anomalies rapidly become westward just after the peak, favoring the decaying phase of EN. The warm phase of ENSO is associated with the shoaling of the thermocline and the accumulation of subsurface cold waters in the
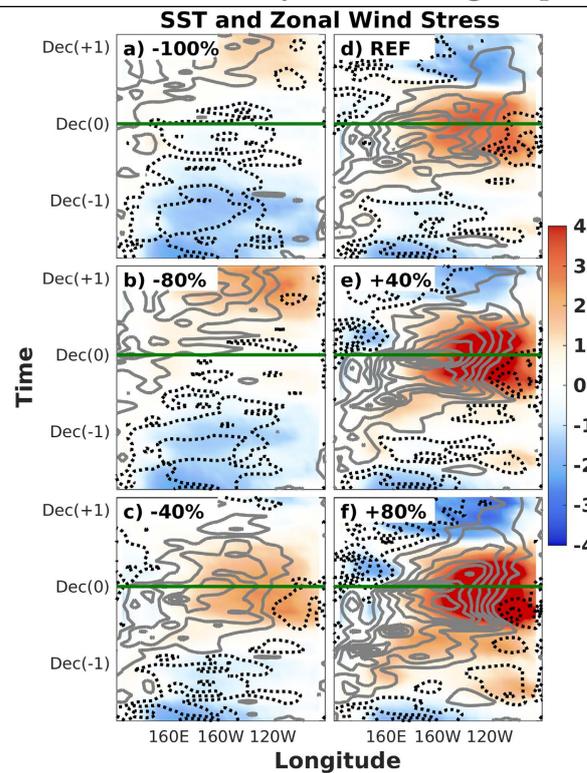
Figure 3. Longitude-time Hovmöller diagram of sea surface temperature (shading, in °C) and zonal wind stress (contour, in N/m²) anomalies. Panels correspond to the −100% (**a**), −80% (**b**), −40% (**c**), REF (**d**), +40% (**e**) and +80% (**f**) experiments. The minimum contour is ±0.005 N/m² and the contour interval is 0.01 N/m², with grey solid (black dashed) lines depicting positive (negative) anomalies. The horizontal green lines indicate the December month in which El Niño peaks in the reference simulation.

western Pacific, which propagate to the east as upwelling Kelvin waves once the EN event starts to decay and the zonal wind anomalies become easterly in the western Pacific. A LN event develops as soon as the subsurface cold anomalies reach the eastern Pacific, although the magnitude of its peak is found to be approximately one half of the magnitude of the preceding EN episode.

When the initial heat buildup is modified by ±40%, the growth and propagation of the subsurface warm anomaly, as well as the subsequent onset of the EN event, remain very similar in timing and approximately proportional in magnitude to that shown in the REF ensemble (Figs 2–4c,e). This result is found to be valid both for surface and subsurface temperatures, as well as for surface winds and ocean currents, showing that the same dynamical mechanisms operate during the phases of the oscillation that precede and follow the peak of EN events. In the particular case of the +40% (−40%) ensemble, the magnitude of the event is nearly proportionally larger (weaker), with the anomaly of the N34 index reaching +3.5 °C (+2 °C) and representing an increase (decrease) of about 30% relative to the REF ensemble.

Further increasing the initial heat buildup up to 80% of the REF simulation induces some interesting differences (Figs 2–4f). The excess heat in the western Pacific is released and quickly starts to warm the eastern part of the equatorial Pacific subsurface. Nevertheless, the surface warming at the end of year −1 is found to be weak and uniformly distributed along the equatorial Pacific. This configuration does not favor the activation of the Bjerknes feedback and therefore the Walker circulation remains in a neutral phase (i.e. westerlies in the west but easterlies in the east), resulting in weak EN conditions (N34 < +1 °C). As a result, the accumulated heat is not discharged towards higher latitudes, and therefore this initial EN-like event only represents a step in the slow but steady warming of the basin, characterized by an initial warm base state of the equatorial Pacific, and enhanced by strong westerly wind anomalies progressing to the east throughout year 0. The subsequent EN event is found to be very strong (N34 = +4 °C) and followed by a strong LN event one year later.

The picture is however completely different in the ensemble of simulations in which the initial heat buildup is reduced by 80% (Figs 2–4b). The equatorial easterly wind anomalies (or, more precisely, the off-equatorial wind stress curl) observed before and after the beginning of the simulations are associated with the positive change rate in subsurface meridional convergence[10–11]. This tendency towards equatorward mass convergence is associated with upwelling of subsurface cold waters that favors the persistence of cold SST and easterly wind anomalies in the central Pacific, which in turn deepen the thermocline and accumulate subsurface warm waters in the western
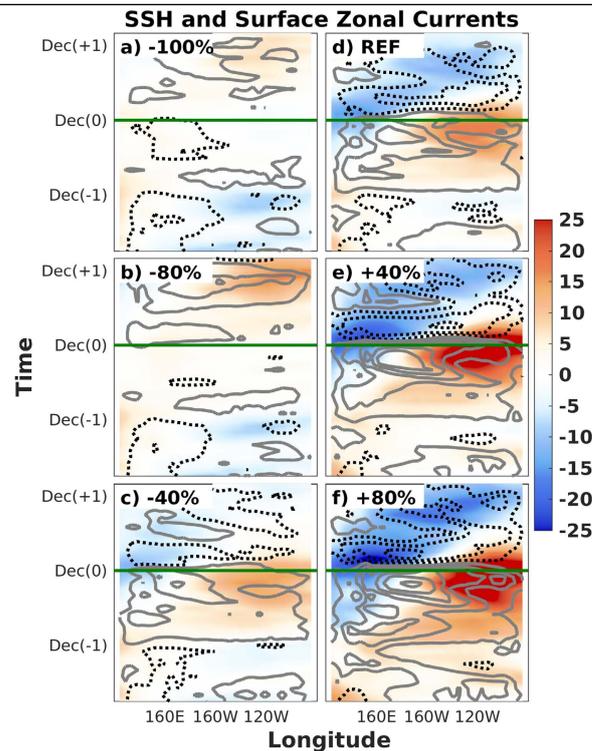
Figure 4. Longitude-time Hovmöller diagram of sea surface height (shading, in cm) and surface zonal current (contour, in m/s) anomalies. Panels correspond to the −100% (**a**), −80% (**b**), −40% (**c**), REF (**d**), +40% (**e**) and +80% (**f**) experiments. The minimum contour is ±0.1 m/s and the contour interval is 0.2 m/s, with grey solid (black dashed) lines depicting positive (negative) anomalies. The horizontal green lines indicate the December month in which El Niño peaks in the reference simulation.

Pacific[26]. A weak LN event therefore develops at the end of year −1, which re-activates the generation of the subsurface heat buildup in the western Pacific. From this point onward, the evolution of the accumulated heat mimics that of the REF ensemble in terms of magnitude, timing, propagation and mechanisms, but with a one-year delay, leading to the growth of a strong EN episode that peaks in December of year +1 (N34 = +2 °C, i.e. same magnitude as in the −40% experiment, but one year later).

A similar evolution is found when the accumulated heat in the subsurface is completely suppressed in the −100% experiment (Figs 2–4a). In this case, however, given that the initial THC is weaker and the heat buildup is completely removed, the renewed recharge of the tropical Pacific results in a delayed EN event of smaller magnitude (N34 = +1 °C).

The relationship between initial THC and ENSO variability is summarized in Fig. 5. THC is defined here as the average heat content within 120E-80W, 5S-5N and the upper 300 m. Results show that the greater the initial THC, the greater its magnitude up to the spring of year 0, when the peak in THC is observed in the initially-recharged simulations (Fig. 5a). This dependency is strong during this initial period (correlation ≈ 1, pink line in Fig. 5b), but the recharge does not occur at the same pace (regression = 0.61 J/J in January of year 0, red line in Fig. 5b), as this process is faster in the initially-discharged simulations. The relationship between the initial and the time-varying THC becomes negative in autumn of year 0 (red and pink lines equal to zero in Fig. 5b), when all the simulations exhibit similar THC values (Fig. 5a). This includes the few initially-discharged experiments in which the heat content is still increasing at the end of year 0 (i.e. −100% and −80%), as well as all the other experiments (i.e. from −60% to +100%), in which the heat content is already being discharged at this point in time (Fig. 5a). The relationship then becomes negative, but it is still strong (correlation = −0.92 and regression = −1.35 J/J in June of year +1, Fig. 5b).

Results also show that the greater the initial THC, the greater the magnitude of the subsequent EN event (see Fig. 5c at the end of year 0). The relationship is again strong (correlation = 0.92 and regression = 1.56 °C/10^16 J in December of year 0, blue and cyan lines in Fig. 5b), but not completely linear. On the one hand, in the initially-recharged simulations, negative feedbacks in the central and eastern Pacific limit the growth and magnitude of EN during its mature phase[10,11,34–36], which explains why the N34 Index in December of year 0 is only +1 °C warmer in the +100% ensemble than in REF (Fig. 5c). On the other hand, the spread in the initially-discharged simulations is large, given that the Bjerknes feedback is not activated in some experiments (i.e. the N34 Index is equal to +2.8 °C in REF and negative in −100%, Fig. 5c). The relationship becomes negative
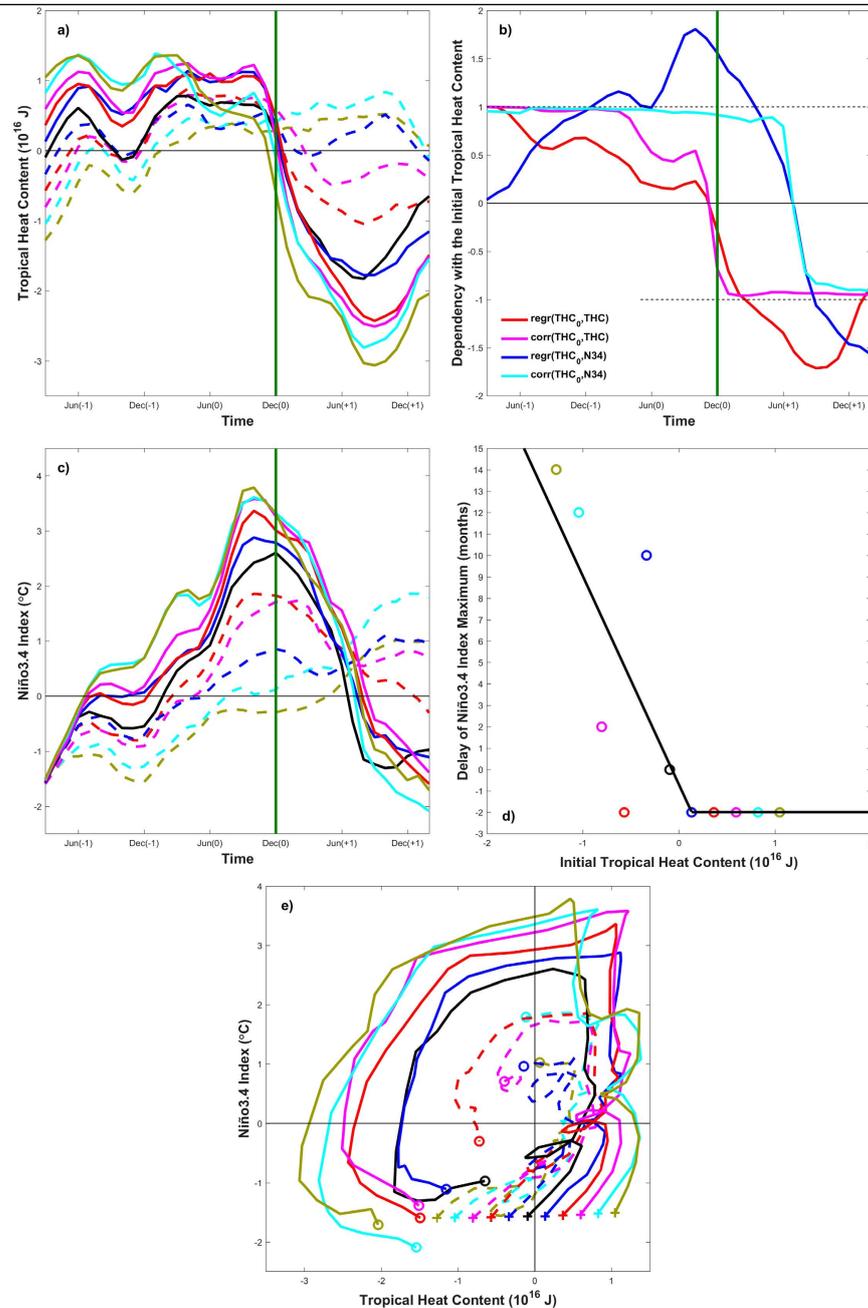
Figure 5. Relationship between Tropical Heat Content (10^{16}J) and the Niño3.4 Index (°C). Panels (a,c,e) depict the evolution of Tropical Heat Content (THC) and the Niño3.4 (N34) Index. Crosses (circles) in panel (e) correspond to the initial (final) month of the model runs. Panel (d) shows the delay (in months) in the maximum of N34 relative to the December month in which El Niño peaks in the reference simulation (see vertical green lines in panels a–c). The black line in panel d shows the piecewise linear fitting between x- and y-axis variables. Curves and circles in panels (a,c–e) correspond to the −100% (dashed lime), −80% (dashed cyan), −60% (dashed magenta), −40% (dashed red), −20% (dashed purple), REF (solid black), +20% (solid purple), +40% (solid red), +60% (solid magenta), +80% (solid cyan) and +100% (solid lime) experiments. Panel (b) shows the regression coefficient (red and blue lines) and the Pearson correlation (pink and cyan) among model experiments between the Initial Tropical Heat Content (THC_0) and THC (red and pink) and N34 (blue and cyan). Regression values in panel (b) are unitless for THC, and °C/10^{16}J for N34. THC was computed as the temperature average within 120E-80W, 5S-5N and the upper 300 m.

in early summer of year $+1$ (blue and cyan lines equal to zero in Fig. 5b), when the transition between warm and cold conditions is found in most, albeit not all, simulations (Fig. 5c). Indeed, in the initially-discharged simulations, the timing of the peak of EN depends on the prescribed THC, with a clear phase-locking to the seasonal cycle that characterizes the jump between the winters of years 0 and $+1$[37–40]; Fig. 5d). Instead, all of the initially-recharged simulations are found to peak in October of year 0, defining a stepwise relationship between the initial THC and the timing of EN maxima (Fig. 5d).

The phase evolution of the system shows the traditional counterclockwise trajectory, in which the change rate of the N34 Index is approximately proportional to the THC, and the radius of the trajectories monotonically increases as a function of the initial THC (Fig. 5e, see also ref. 30 for a similar approach based on ocean energetics). Nonetheless, the trajectories of all the initially-discharged simulations tend to evolve towards the diagram values that correspond to the recharge phase of the REF ensemble (i.e. $+0.5\cdot10^{16}$ J $\leq$ THC $\leq +1\cdot10^{16}$ J and $-0.5\,°C \leq$ N34 $\leq +0.5\,°C$), and only then do they diverge to reach weaker N34 values than in REF. This indicates that, in our model configuration, the system compensates for the initially-prescribed reduction in heat content, and evolves towards a new recharge in THC and the generation of EN events through the memory of the system, regardless of the magnitude of the initial THC, and even when it is completely removed.

This general increasing trend of THC in the initial period of all the simulations is explained by the recharge theory, in which the off-equatorial wind stress curl is associated with the positive change rate in subsurface meridional convergence. In the initially-recharged simulations, this tendency towards the deepening of the thermocline in the central Pacific contributes to the transition towards the recharge phase that leads to EN by 2 to 3 seasons[10,11]. Instead, in the simulations in which the initial heat content has been completely or largely suppressed (i.e. -100% and -80%), the equatorward mass convergence is associated with the upwelling of subsurface cold waters that favors the persistence of cold SST and easterly wind anomalies in the central Pacific, which in turn deepen the thermocline and generate a new heat buildup in the western Pacific[26,27]. We must however note that, in a more general framework, the initial THC could increase at a different rate or even decrease if wind stress anomalies were also prescribed or a different time frame was chosen as initial conditions for the experiments.

The new recharge process in the -100% and -80% experiments can result in a delay in the occurrence of the EN event, which highlights the non-linear dependency between the intensity of the subsurface heat buildup and both the magnitude and timing of subsequent EN episodes. The numerical simulations reported here show that the accumulation of warm waters in the western Pacific determines the timing of the transition between LN and EN conditions, which is here seen to increase by one year when the initial subsurface heat is largely reduced. We found that the stepwise relationship between the initial THC and the timing of EN maxima also affects the magnitude of the events. For example, the EN episode at the end of year 0 in the -40% ensemble has similar magnitude to the event occurring one year later in the -80% experiment, because the longer timescale of the recharge process compensates the magnitude of the initial discharge. In this respect, our results provide new insight into the fundamental role of the ocean heat content, in this case at longer lead times than traditionally described, and therefore they have important implications for the understanding of the genesis of EN events, their dynamics and their predictability.

Methods

Estimates of WWV for the 1980–2015 period were derived from the TAO Project Office of NOAA/PMEL, which are based on temperature analyses of the Bureau National Operations Centre at the Australian Bureau of Meteorology and profiles from TAO moorings, Argo floats and XBTs, with anomalies calculated after removing the 1980–2002 mean seasonal cycle[32].

The model used in the present work, the Community Earth System Model (CESM) v1.2, is a global coupled model extending its predecessor, the Community Climate System Model (CCSM) v4, by incorporating new Earth system simulation capabilities[41]. The model configuration used here couples the latest version of the Community Atmosphere Model v5[42] with the Parallel Ocean Program v2[43], the Community Land Model v4[44], the Community Ice Code v4[45] and the River Transport Model[46]. The atmospheric model has a resolution of 2.5° in longitude and 1.875° in latitude, with 30 vertical levels. The ocean model uses a displaced pole grid with approximately 1° resolution in longitude and 0.5° in latitude, which is refined within the tropical band up to 0.25° at the equator. There are 60 vertical levels with the highest resolution (10 m) in the upper 150 m, and the lowest (250 m) in the deep ocean.

CESM was chosen to conduct the experiments reported here because CCSM was found to be one of the three best models in the simulation of the dynamic warm pool edge among 19 coupled ocean-atmosphere general circulation models from the Coupled Model Intercomparison Project Phase 5[47]. This feature corresponds to the easternmost edge of the western Pacific warm pool, the maximum in zonal salinity gradient and the area of surface ocean convergence, downward motion and advection and subsurface divergence[26,48–50]. The correct simulation of these dynamical features is seen to be key for the reproduction of the ENSO oscillation and the transition between events of opposite sign, including the generation of the heat buildup in the western Pacific and the eastward propagation of the accumulated heat[21,47].

We performed 11 sets of ensemble experiments, with initial conditions corresponding to a very early phase of the onset of an EN episode, in March of the year preceding the winter peak of a strong warm event (i.e. a lead time of 21 months with regard to the December maximum), chosen from a reference 100 year spin-up simulation. This stage of the ENSO oscillation is characterized by cold LN-like conditions in the tropical Pacific and the generation of a subsurface heat buildup in the western tropical Pacific (Supplementary Figure 1b,e;[21,26]). In each of these sets of ensemble experiments, the intensity of the subsurface warm anomaly was decreased (negative sign representing a discharge in heat content) or increased (positive sign or recharge) by $\pm20\%$, $\pm40\%$, $\pm60\%$, $\pm80\%$ and $\pm100\%$. The word "anomaly" in the experiments refers to the difference of a monthly value with regard to the long-term mean annual cycle computed from the REF simulation. We note that warm temperature anomalies were fully modified only in the inner three-dimensional box [120E-80W] $\times$ [10S-10N] $\times$ [50–200 m],

and that this modification was linearly decreased to zero from the border of this box to the frontier of the outer box [100E-60W] × [15S-15N] × [20–300 m]. For example, Supplementary Figure 1a,d (Supplementary Figure 1c,f) show the initial condition in the −100% (+100%) experiment, in which the subsurface warm anomaly was suppressed (doubled). For the sake of clarity, we use the terminology "initially discharged" ("initially recharged") simulations or ensembles to refer to the experiments with initially reduced (increased) subsurface warming, and we also compare the magnitude of these prescriptions by saying that the −100% (+100%) ensemble is "more initially discharged" ("more initially recharged") than for example the −20% (+20%) experiment. Each set of experiments in turn consists of 10 simulations with slightly perturbed initial conditions, from which only the ensemble average is shown here.

References

1. Bjerknes, J. Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review* **97,** 163–172 (1969).
2. Wyrtki, K. El Niño - The Dynamic Response of the Equatorial Pacific Ocean to Atmospheric Forcing. *Journal of Physical Oceanography* **5,** 572–584 (1975).
3. Meinen, C. S. & McPhaden, M. J. Observations of Warm Water Volume Changes in the Equatorial Pacific and Their Relationship to El Niño and La Niña. *Journal of Climate* **13,** 3551–3559 (2000).
4. McPhaden, M. J., Zebiak, S. E. & Glantz, M. H. ENSO as an Integrating Concept in Earth Science. *Science* **314,** 1740 (2006).
5. Ballester, J., Rodríguez-Arias, M. A. & Rodó, X. A new extratropical tracer describing the role of the western Pacific in the onset of El Niño: Implications for ENSO understanding and forecasting. *Journal of Climate* **24,** 1425–1437 (2011).
6. Chen, D., Cane, M. A., Kaplan, A., Zebiak, S. E. & Huang, D. Predictability of El Niño over the past 148 years. *Nature* **428,** 733–736 (2004).
7. Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S. & DeWitt, D. G. Skill of Real-Time Seasonal ENSO Model Predictions during 2002–11: Is Our Capability Increasing? *Bulletin of the American Meteorological Society* **93,** 631–651 (2012).
8. Ballester, J. *et al.* Kawasaki disease and ENSO-driven wind circulation. *Geophysical Research Letters* **40,** 2284–2289 (2013).
9. Ballester, J., Lowe, R., Diggle, P. & Rodó, X. Modelling and prediction of climate and health impacts: challenges and opportunities. *Annals of the New York Academy of Sciences*, doi: 10.1111/nyas.13129 (2016a).
10. Jin, F. F. An Equatorial Ocean Recharge Paradigm for ENSO. Part I: Conceptual Model. *Journal of the Atmospheric Sciences* **54,** 811–829, doi: 10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2 (1997a).
11. Jin, F. F. An Equatorial Ocean Recharge Paradigm for ENSO. Part II: A Stripped-Down Coupled Model. *Journal of the Atmospheric Sciences* **54,** 830–847, doi: 10.1175/1520-0469(1997)054<0830:AEORPF>2.0.CO;2 (1997b).
12. Wang, C. & Picaut, J. *Understanding ENSO physics - A review* (eds Wang, C., Xie, S. P. & Carton, J. A.). 21–48 (American Geophysical Union, 2004).
13. Zebiak, S. E. & Cane, M. A. A Model El Niño–Southern Oscillation. *Monthly Weather Review* **115,** 2262–2278 (1987).
14. Chen, D. & Cane, M. A. El Niño prediction and predictability. *Journal of Computational Physics* **227,** 3625–3640 (2008).
15. Izumo, T. *et al.* Influence of the state of the Indian Ocean Dipole on the following year's El Niño. *Nature Geoscience* **3,** 168–172 (2010).
16. Duan, W. & Wei, C. The 'spring predictability barrier' for ENSO predictions and its possible mechanism: results from a fully coupled model. *International Journal of Climatology* **33,** 1280–1292 (2013).
17. Petrova, D., Koopman, S. J., Ballester, J. & Rodó, X. Improving the Long-Lead Predictability of El Niño Using a Novel Forecasting Scheme Based on a Dynamic Components Model. *Climate Dynamics*, doi: 10.1007/s00382-016-3139-y (2016).
18. Kessler, W. S. Is ENSO a cycle or a series of events? *Geophysical Research Letters* **29,** doi: 10.1029/2002GL015924 (2002).
19. Fedorov, A. V., Harper, S. L., Philander, S. G., Winter, B. & Wittenberg, A. How predictable is El Niño? *Bulletin of the American Meteorological Society* **84,** 911–919 (2003).
20. Fedorov, A. V., Hu, S., Lengaigne, M. & Guilyardi, E. The impact of westerly wind bursts and ocean initial state on the development, and diversity of El Niño events. *Climate Dynamics* **44,** 1381–1401 (2015).
21. Ballester, J., Bordoni, S., Petrova, D. & Rodó, X. Heat advection processes leading to El Niño events as depicted by an ensemble of ocean assimilation products. *Journal of Geophysical Research: Oceans* **121,** 3710–3729 (2016b).
22. Ren, H. L. & Jin, F. F. Recharge Oscillator Mechanisms in Two Types of ENSO. *Journal of Climate* **26,** 6506–6523 (2013).
23. McPhaden, M. J. & Nagura, M. Indian Ocean dipole interpreted in terms of recharge oscillator theory. *Climate Dynamics* **42,** 1569–1586 (2014).
24. McPhaden, M. J. Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophysical Research Letters* **30,** doi: 10.1029/2003GL016872 (2003).
25. Wang, B., Wu, R. & Lukas, R. Roles of the western north Pacific wind variation in thermocline adjustment and ENSO phase transition. *Journal of The Meteorological Society of Japan* **77,** 1–16 (1999).
26. Ballester, J., Bordoni, S., Petrova, D. & Rodó, X. On the dynamical mechanisms explaining the western Pacific subsurface temperature buildup leading to ENSO events. *Geophysical Research Letters* **42,** 2961–2967 (2015).
27. Wyrtki, K. Water displacements in the Pacific and the genesis of El Nino cycles. *Journal of Geophysical Research: Oceans* **90,** 7129–7132 (1985).
28. Ramesh, N. & Murtugudde, R. All flavours of El Niño have similar early subsurface origins. *Nature Climate Change* **3,** 42–46 (2013).
29. McPhaden, M. J. Evolution of the 2002/03 El Niño. *Bulletin of the American Meteorological Society* **85,** 677 (2004).
30. Hu, S., Fedorov, A. V., Lengaigne, M. & Guilyardi, E. The impact of westerly wind bursts on the diversity and predictability of El Niño events: an ocean energetics perspective. *Geophysical Research Letters* **41,** 4654–4663 (2014).
31. Hu, S. & Fedorov, A. V. Exceptionally strong easterly wind burst stalling El Niño of 2014. Proceedings of the National Academy of Sciences, doi: 10.1073/pnas.1514182113 (2016).
32. TAO. Upper Ocean Heat Content and ENSO. http://www.pmel.noaa.gov/elnino/upper-ocean-heat-content-and-enso. Last accessed in September (2016).
33. CPC. Cold and Warm Episodes by Season. http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml. Last accessed in August (2016).
34. Suarez, M. J. & Schopf, P. S. A Delayed Action Oscillator for ENSO. *Journal of the Atmospheric Sciences* **45,** 3283–3287 (1988).
35. Weisberg, R. H. & Wang, C. A western Pacific oscillator paradigm for the El Niño-Southern Oscillation. *Geophysical Research Letters* **27,** 779–782 (1997).
36. Picaut, J., Masia, F. & du Penhoat, Y. An advective-reflective conceptual model for the oscillatory nature of ENSO. *Science* **277,** 663–666 (1997).
37. Tziperman, E., Cane, M. A., Zebiak, S. E., Xue, Y. & Blumenthal, B. Locking of El Niño's Peak Time to the End of the Calendar Year in the Delayed Oscillator Picture of ENSO. *Journal of Climate* **11,** 2191–2199 (1998).
38. Harrison, D. E. & Vecchi, G. A. On the termination of El Niño. *Geophysical Research Letters* **26,** 1593–1596 (1999).
39. Jin, F. F., Kim, S. T. & Bejarano, L. A coupled-stability index for ENSO. *Geophysical Research Letters* **33,** L23708, doi: 10.1029/2006GL027221 (2006).

40. Stuecker, M. F., Timmermann, A., Jin, F. F., McGregor, S. & Ren, H. L. A combination mode of the annual cycle and the El Niño/Southern Oscillation. *Nature Geoscience* **6,** 540–544 (2013).
41. Hurrell, J. W. *et al.* The Community Earth System Model: A Framework for Collaborative Research. *Bulletin of the American Meteorological Society* **94,** 1339–1360 (2013).
42. Neale, R. B. *et al.* The mean climate of the Community Atmosphere Model (CAM4) in forced SST and fully coupled experiments. *Journal of Climate* **26,** 5150–5168 (2013).
43. Smith, R. *et al.* The Parallel Ocean Program (POP) reference manual: Ocean component of the Community Climate System Model (CCSM) and Community Earth System Model (CESM). Los Alamos National *Laboratory Technical Report LAUR-10-01853.* 141 pp (2010).
44. Lawrence, D. M. *et al.* Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems* **3,** M03001, doi: 10.1029/2011MS000045 (2011).
45. Hunke, E. C. & Lipscomb, W. H. CICE: The Los Alamos sea ice model, documentation and software, version 4.0. Los Alamos National *Laboratory Technical Report LACC-06-012.* 76 pp (2008).
46. Branstetter, M. L. Development of a parallel river transport algorithm and applications to climate studies. *University of Texas (*2001).
47. Brown, J. N., Langlais, C. & Maes, C. Zonal structure and variability of the Western Pacific dynamic warm pool edge in CMIP5. *Climate Dynamics* **42,** 3061–3076 (2014).
48. Picaut, J., Ioualalen, M., Menkes, C., Delcroix, T. & McPhaden, M. J. Mechanism of the Zonal Displacements of the Pacific Warm Pool: Implications for ENSO. *Science* **274,** 1486–1489 (1996).
49. Bosc, C., Delcroix, T. & Maes, C. Barrier layer variability in the western Pacific warm pool from 2000 to 2007. *Journal of Geophysical Research* **114,** C06023, doi: 10.1029/2008JC005187 (2009).
50. Singh, A., Delcroix, T. & Cravatte, S. Contrasting the flavors of El Niño-Southern Oscillation using sea surface salinity observations. *Journal of Geophysical Research* **116,** C06016, doi: 10.1029/2010JC006862 (2011).

Acknowledgements

Author Contributions

J.B. took the lead role in the design of the study, performed the simulations, analyzed the data and wrote the manuscript. D.P., S.B., B.C., M.G.-D. and X.R. contributed to the discussion of results and to the revision of the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ballester, J. *et al.* Sensitivity of El Niño intensity and timing to preceding subsurface heat magnitude. *Sci. Rep.* **6**, 36344; doi: 10.1038/srep36344 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.5 Multi-year statistical prediction of ENSO enhanced by the tropical Pacific observing system

Petrova D (1,2), Ballester J (1,2), Koopman SM (3), Rodó X (1,2,4)

(1) Institut Català de Ciències del Clima, Barcelona, Catalonia, Spain

(2) Barcelona Institute for Global Health, Barcelona, Catalonia, Spain

(3) Vrije Universiteit Amsterdam, Amsterdam, Netherlands

(4) Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

Resumen - *Summary in Spanish*

Se ha demostrado que el límite teórico de predictibilidad de El Niño-Oscilación Sur (ENSO) es del orden de los años, sin embargo, predicciones de largo plazo de El Niño (EN) y La Niña (LN) son esencialmente no disponibles, y los esquemas de pronóstico más avanzados son en gran parte incapaces de predecir correctamente más allá de la barrera de primavera. Los esfuerzos se han dedicado principalmente a la mejora de los modelos dinámicos, mientras que los esquemas estadísticos se han mantenido subdesarrollados. En consecuencia, no se han aprovechado plenamente la disponibilidad de variables oceánicas subsuperficiales que se han proporcionado regularmente durante las últimas décadas como resultado del Programa de Océanos Tropicales Atmósfera Global (TOGA). Aquí utilizamos varias variables predictoras, incluyendo la temperatura subsuperficial a diferentes profundidades y regiones del océano ecuatorial, en un modelo estadístico flexible de componentes dinámicos para hacer previsiones retrospectivas hábiles de largo plazo del Índice Niño3.4 (N3.4) en el período 1970-2016. El modelo predice exitosamente hasta dos años y medio de antelación todos los episodios más importantes de EN, incluyendo el reciente extremo EN 2015/16. El análisis demuestra que los eventos se predicen con mucha mayor precisión después de la finalización del sistema de observación en el Pacífico tropical en 1994, como resultado de la mejora de la calidad de los datos y la cobertura alcanzada por TOGA. Por lo tanto, ahora es posible emitir predicciones de largo plazo de este importante fenómeno climático a un bajo costo computacional.

Multi-year statistical prediction of ENSO enhanced by the tropical Pacific observing system

Desislava Petrova[a,b], **Joan Ballester**[a,b], **Siem Jan Koopman**[c], and **Xavier Rodó**[a,b,d]

[a]Climate Dynamics and Impacts Unit, Catalan Institute for Climate Sciences (IC3), Carrer del Dr. Trueta, 203, 08005 Barcelona, Catalonia, Spain; [b]Climate and Health Program, Barcelona Institute for Global Health (ISGLOBAL), Carrer del Dr. Aiguader 88, 08003 Barcelona, Catalonia, Spain; [c]Department of Econometrics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands; [d]Institució Catalana de Recerca i Estudis Avancats (ICREA), Barcelona, Catalonia, Spain

The theoretical predictability limit of El Niño-Southern Oscillation (ENSO) has been shown to be on the order of years (1–5), however, long-lead predictions of El Niño (EN) and La Niña (LN) events are essentially lacking, and state-of-the-art forecasting schemes are largely unable to predict correctly beyond the spring barrier (6). Efforts have mostly been dedicated to the improvement of dynamical models, while statistical schemes have remained underdeveloped. Consequently, they have not taken full advantage of the availability of subsurface ocean variables (6), which have been provided on a regular basis for the last couple of decades as a result of the Tropical Ocean Global Atmosphere Program (TOGA) (7). Here we use several predictor variables, including subsurface temperature at different depths and regions of the equatorial ocean, in a flexible statistical dynamic components model (8) to make skilful long-lead retrospective forecasts of the Niño3.4 Index (N3.4) in the period 1970-2016. The model successfully predicts up to two and a half years in advance all the major EN episodes, including the recent extreme 2015/16 EN. The analysis demonstrates that events are predicted much more accurately after the completion of the observational array in the tropical Pacific in 1994 (9), as a result of the improved data quality and coverage achieved by TOGA. Therefore, there is now potential to issue long-lead predictions of this important climatic phenomenon at a low computational cost.

ENSO | long-lead prediction | subsurface ocean dynamics| observing system

Skilful long-range forecasts of EN are still in high demand. After decades of extensive efforts, dynamical models nowadays represent the best available tools to issue ENSO forecasts at lead times of up to two seasons, although they are still largely constrained by the lack of complete understanding of the physics of the phenomenon, by problems arising from the initialization of the components of the climate system or by the need for accurate parametrization of important physical processes (6). Statistical models, on the other hand, largely depend on the availability of ocean and atmosphere historical data, so that the longer the length of the data, the more robust is the predictor-predictand relationship identified by the model (6). In addition to these factors, the low signal-to-noise ratio in boreal spring (10), the influence of high-frequency atmospheric winds (11, 12), as well as the natural irregularity of the climate system (13) all limit the long-term dynamical and statistical forecasting of the phenomenon. Some of the classical ENSO theories view the oscillation as self-sustained (14–16), and support the claim that it is potentially predictable several years in advance (1, 3, 5, 17, 18), but only a few studies document such long-lead forecasts of past events, and most of them use dynamical models (4, 5, 8). Statistical models are assumed to be less skilful at long lead times, and comparable in performance to dynamical schemes at shorter lead times of about half a year (18, 19). To some extent this is explained by the fact that a new generation of statistical models has not been developed and added to the ENSO forecasting plume, while the majority of the old models have not been substantially revised since they were created in the 1980s and early 1990s (6).

One of the strongest events on record - the 1982/83 EN - surprised the scientific community (1, 7) as it was neither predicted, nor identified until very late in its development. This triggered a decade-long effort to put in place a monitoring system in the tropical

---

Significance Statement

The El Niño Southern Oscillation is the most energetic climatic mode in the tropical Pacific and has profound impact on the weather patterns worldwide. Long-lead predictions of the phenomenon more than one year ahead could be extremely useful for many people around the globe, but they are still a challenge for existing models. Very few multi-year ahead experimental forecasts performed with dynamical models have been documented in the literature. The potential of statistical models has been overlooked, and they have not been upgraded to use effectively the available subsurface ocean information from the equatorial Pacific Ocean, which is of utmost importance for the evolution of El Niño and La Niña events. Here we use a new statistical model to predict ENSO at least 2 years in advance. We discuss our forecasts in the context of the tropical Pacific observing system, and highlight the dramatic improvement of our simulations of the period after the installation of a high-resolution array network of moorings (TAO-TRITON) measuring regularly subsurface and surface temperature and wind stress among other variables.

[2]Desislava Petrova. E-mail: desislava.petrova@isglobal.org

Pacific with the aim of studying ENSO better and putting emphasis on the improvement of the predictive capacity of models (7), which led to the inauguration of the TOGA research program in 1985 (7). It deployed a three-dimensional array in the tropical Pacific that since then regularly samples the subsurface temperature down to 500 metres depth. The system was completed in 1994, just in time to track the stronger-than-normal trade winds in 1995/96, which generated a buildup of warm waters in the western tropical Pacific more than one year before the peak of the record-breaking 1997/98 EN event (7). This was the first time when the scientific community and the public could see the benefits of TOGA. Although a number of studies now fully recognize the fundamental role that the intensification of the trade winds and the subsurface heat buildup in the western equatorial Pacific play in the onset of EN events (1, 8, 16, 20–23), many of the operational statistical models do not account in detail for these processes that occur early on in the generation of the events (24–27).

In the present study we use the flexible statistical dynamic components model described in ref. 8, which at long lead times incorporates explanatory variables designed to capture the three-dimensional shape of the warm pool subsurface heat buildup at different depth levels, as well as zonal wind stress anomalies in the central and western equatorial Pacific (see Methods). The model consists of several stochastic cycle components with frequencies corresponding to the main peaks in the spectrum of N3.4 (see ref. 8), as well as explanatory regression variables such as sea surface and subsurface temperature and zonal wind stress. These variables enter the model equations in the form of lagged time series with respect to the December value of N3.4, and are selected to be consistent with the EN dynamical evolution. In this way, different covariates are used for forecasting at different lead times, depending on the average temporal progression of EN events. For instance, at a lead time of 24 months, the selected predictors are subsurface temperatures in the far western equatorial Pacific Ocean at depths of 250, 300 and 400 metres, as the heat buildup is generally intense there at this early stage in the generation of a warm event (8, 23).

Model Simulations and Forecasts

The observed and forecast monthly N3.4 anomalies at 6 and 24 months lead time are presented in Fig. 1. The 6-month lead forecast correctly predicts the timing and magnitude of all EN and LN events, and no false alarms are generated ($r^2 = 0.72$, $p < 0.001$, $RMSE = 0.54$; Fig. 1A). Since an ENSO event is typically already under-way half a year before its peak in December-January-February (DJF), the majority of the operational forecasting schemes are able to produce predictions at this lead time that are highly correlated with the observations (6) ($r^2 > 0.40$). The 24-month lead forecast, and in general any lead time forecast beyond the spring barrier (i.e. from 8 months onward; not shown), properly reproduces the crests and troughs in the time series, which is also evident in the high correlation coefficient between the observations and the forecasts ($r^2 = 0.66$, $p < 0.001$, $RMSE = 0.62$; Fig. 1B). However, for the period before the prominent 1997/98 EN, we find that the predicted amplitudes of the larger events are notably smaller than the observed. We highlight that this cannot be explained by a change in the interannual ENSO activity in the different time periods, as three sizeable EN (i.e. 1972/73, 1982/83, 1986/87 and 1997/98, 2009/10, 2015/16) and LN (i.e. 1973/74, 1975/76, 1987/88 and 1998/00, 2007/08, 2010/11) episodes have occurred before and after 1994 (28).

To characterize better the difference between periods, Fig. 2 displays the regressions between the observations and forecasts for two consecutive 22-year sub-periods (1972-1993 in blue and 1994-2015 in red) at 6- and 24-month lead. No substantial difference is observed between the slopes of the regression lines for the two periods at the shorter lead time ($regr_{1972-1993} = 0.65$, $t = 23.88$, $regr_{1994-2015} = 0.74$, $t = 27.34$, $p < 0.001$; Fig. 2A), indicating that the model performance is comparable. Conversely, the regression coefficients significantly increase for the long-range forecasts made after 1994 ($regr_{1972-1993} = 0.35$, $t = 17.12$, $regr_{1994-2015} = 0.65$, $t = 30.93$, $p < 0.001$; Fig. 2B), which represents a major improvement in the capacity of the model. The change in the overall similarity between the observations and the forecasts at 24-month lead time is also assessed by the fifteen-year moving correlation shown in Fig. 1C. The correlation increases monotonically with time until the early 1990s and then stays relatively constant afterwards. At the same time, data availability was constantly improving during TOGA, until the tropical Pacific network array of moorings was fully into place at the end of the program in 1994 (9).

To further explore the difference in the model performance over the two periods, Fig. 3 shows correlations and root mean square errors (RMSE) for the whole range of lead times up to 24 months. For lead times of about 2 seasons both the correlations and RMSE are similar among periods, while for lead times beyond the spring predictability barrier they start to diverge. We also observe that correlations and RMSE stay relatively constant beyond spring. Previous studies (2, 4, 18, 29) have already concluded that the spring predictability barrier is not an intrinsic barrier to the system itself, but it rather depends on model skill and data availability. Here we confirm this result, as we also find that the drop in forecast skill is minimal beyond the spring barrier (Fig. 3). The statistical model we use is linear, and while its stochastic cyclical components are mainly responsible for capturing the correct phase of the oscillation, the lagged predictor explanatory variables are expected to contribute to the correct forecast of the amplitudes of the events, especially at longer lead times (see Methods and ref. 8 for details). Therefore, below we analyse if the predictor variables add significantly to the EN forecasts of the earlier period, which also coincides with a time when no regular subsurface temperature and wind stress data were being provided yet (9).

The forecasts at several lead times of the strongest EN events in the study period (see ref. 28) are displayed in Fig. 4. In all cases the model is capable of predicting the occurrence of a warm event 29 months in advance (magenta curve), although there are evident errors in the amplitude and timing in some cases. A much better representation of the amplitudes in the long-lead forecasts of the events in the second period (1997/98, 2009/10 and 2015/16), as compared to those occurring in the first period (1972/73, 1982/83 and 1986/87), is also clearly distinguished in the figure. The estimated coefficients and the corresponding t- and p-values for the explanatory variables used in the 24-month lead forecasts of all the warm events in the study period are listed in Table 1. Remarkably, none of the three predictor variables is found to be significant at the 90% level for the forecasts of any of the events before 1994, while there is at least one significant variable for each forecast of the episodes that occurred afterwards. Similar results hold for the other

long-lead forecasts shown in Figure 4 (Supplementary Tables 1 and 2).

Implications for Long-Lead Statistical El Niño Predictability: the Contribution of TOGA

As seen above, there is a well-defined shift between the lack of significance of the predictor variables for the forecasts of the warm events before the end of TOGA and their significance afterwards. Our results strongly support the view that the improved forecasts are due to the availability of regular and higher resolution subsurface data ensured by the implementation of the observational network array (30, 31). The correct and relevant subsurface information also has consequential implications for the proper forecasting of the magnitudes of the warm events (32). In the linear framework of the model that we use, at the longer lead times the explanatory predictor variables have more forecast weight than they do at the shorter lead times (see Methods). Noticeably, the forecast amplitudes of the three earlier events shown in Fig. 4A-C do not exceed 1.5°C at the long lead times of 21 and 29 months (green and magenta curves). At the same time, the forecast amplitudes of the three events that took place in the later period, when the explanatory variables are shown to have an impact (Table 1), are consistent with the occurrence of a strong EN even at these very long lead times (green and magenta curves in Fig. 4D-F). This is remarkable, given the fact that the two most extreme events on record are among them, with anomalies above 2.5°C.

Although there is a marked difference in the predictive capacity of the model during the earlier and later sub-periods, it still exhibits outstanding skill (i.e. correlations and RMSE in Fig. 3) for a statistical model, and even in comparison to the ensemble of dynamical models (6). We have demonstrated that the equatorial Pacific observing system, and especially the provision of subsurface temperature data on a regular basis, has a vital contributing role for its long-lead forecasting capabilities. With the end of TOGA in 1994 nearly the whole equatorial band between 10°N-10°S was covered with moorings (9), and this is also the start of altimetry data (29). As seen in the Figure of the Tropical Atmosphere Ocean-Triangle Trans-Ocean Buoy Network (TAO-TRITON) array development (30), some subsurface data from the central Pacific was already streamed at the end of 1987, while at the end of 1991 data was also coming in from the western Pacific, which represents a key region for the forecast of the phenomenon at lead times of two years or more. Thus, almost three decades have passed since the three-dimensional observations began in the tropical Pacific. As a result, the limited span of the data is now much less of a problem for the robust definition of statistical predictive schemes (6), as we have shown here.

In 1997 simulations with the dynamical Lamont model (33), which at the time was considered to be the benchmark, did not predict the prominent EN that was developing, but forecast a cold boreal winter instead. Only after sea level data (a proxy for upper ocean heat content) from the observing system was included to its initialization scheme, was it able to forecast the event in retrospect (7). In this way, the input of TOGA significantly improved the forecast skill of this and many other dynamical models (7). Moreover, nowadays the most skilful forecasts with such models are those initialized by subsurface observations (3). This confirms the importance of relevant subsurface data and the usefulness of TOGA not just in the case of statistical models discussed here, but also of the more complex dynamical models. In essence, the same conclusion as the one reached here has been made by (29), where a large reduction of the errors in N3.4 SST forecasts made after 1994 is detected with the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecast system 3. The results are also in agreement with an earlier study with the same system (34), in which the effect of ARGO floats are removed from the observations, and it is established that improvements in the forecast are clearly explained by the improved observing system.

Some of the existing statistical systems already include measures of integrated equatorial heat content (6). However, our model uses temperature data from a selection of dynamically relevant regions and depths to maximize its predictive power. These values are not well-represented by spatially-integrated measures of heat content, and our analysis suggests that the integration masks the intensity of the heat buildup in specific regions in the subsurface at long lead times, and more importantly, does not allow the systems to properly track the eastward propagation of heat along the equatorial thermocline (8, 23). Therefore, statistical models should be improved in the direction of using the available subsurface information that is fundamental for ENSO in a more discrete and targeted way. Here we have established that this type of models, which are computationally less expensive than dynamical ones, are able to make forecasts multiple seasons and even years in advance. Hence, we argue that there is no inherent limitation in statistical predictions at long lead times as suggested by ref. 3 and ref. 6. Thus, statistical models can also provide useful information about EN to decision makers around the world, which could prevent threats to human lives and reduce economic costs.

Materials and Methods

The model used in the present study is a more advanced version of the statistical dynamic components model proposed by ref. 8 and developed specifically for prediction of the Niño3.4 Index defined as the average sea surface temperature in the box [5°N-5°S, 170°W-120°W]. The difference with the previous version is the addition of three stochastic cycle components, and the replacement of the fixed seasonal component. Instead, two of the new cyclical components have now been assigned with period lengths close to the annual (12 months) and semi-annual (6 months) frequencies. They are allowed to vary slowly over time in order to address the finding in our previous study that the annual periodicity of the seasonal component was not sufficiently well-simulated, because the annual periodicity is not strictly fixed at 12 months, and especially because during El Niño events the amplitude of the seasonal cycle is suppressed (35). The third additional cycle component in the model is associated with ENSO variability on decadal time scales, as in our original study we also established that this low-frequency variability is important for the simulation of some El Niño events, and this feature was not explicitly resolved in the previous model version (8). The three additional cycle components are modelled as linear dynamic stochastic processes, which are generated by independent series of serially uncorrelated disturbances, (36, 37) as

143 further described in ref. 8.

144   The model belongs to the class of linear Gaussian state space models. Its statistical treatment, which includes parameter estimation
145 and forecasting, is based on the Kalman filter (37). The filter delivers the prediction errors and their variances, from which the
146 Gaussian loglikelihood function is computed. Parameter estimation reduces to numerical maximisation of the loglikelihood function.
147 Given the parameter estimates, the forecast function is constructed from the Kalman filter, and is regarded as a linear function of past
148 and current observations of the dependent and explanatory regression variables. It evaluates the one-step ahead prediction of the state
149 vector conditional on the past and current observations, $y_1, \ldots, y_t$ and $Z_1, \ldots, Z_t$, as denoted by $a_{t+1|t}$, together with its conditional
150 error variance matrix $P_{t+1|t}$, in a recursive updating scheme, for $t = 1, \ldots, n$, (see ref. 8 and ref. 37 for details and derivations). The
151 prediction error at time t is then given by $v_t = y_t - Z_t a_{t|t-1}$ and its error variance is given by $F_t = Z_t P_{t|t-1} Z_t' + H_t$, for $t = 1, \ldots, n$.
152 The system matrices Z_t, H_t, T_t and Q_t can be time-varying but in a deterministic way; they can be functions of the time index t.

153   The ability of the Kalman filter to treat missing observations is a powerful one, and the forecasting problem can be viewed as a
154 missing data problem by treating future observations as missing. Assume that at a certain time index τ, observation y_τ is missing.
155 In the Kalman filter we can treat the one-step ahead prediction error v_τ as unknown with its error variance $F_\tau \to \infty$. This setting
156 reflects that we have no information about v_τ. The updating steps for the state vector estimate and its error variance matrix simply
157 reduce to the prediction steps

$$a_{t+1|t} = T_t a_{t|t-1}, \qquad P_{t+1|t} = T_t P_{t|t-1} T_t' + H_t H_t', \tag{1}$$

158 for $t = \tau$. Note that $a_{t|t} = a_{t|t-1}$ and $P_{t|t} = P_{t|t-1}$ for $t = \tau$. The implementation of a Kalman filter with missing data entries is
159 straightforward and relies on a conditional statement: if y_t is observed, carry out the Kalman filter as usual; if y_t is missing, carry out
160 the prediction step Eq. (1).

161   The treatment of missing values can be adapted to the forecast and forecast error computations. Assume that after the last
162 available observation with time index m, the forecasts of $y_{m+1}, \ldots, y_{m+h}$ are requested, for some forecast horizon h. We then add
163 a series of h missing values to the data and the Kalman filter treats them in the way described above. In this way we obtain
164 the state prediction estimates $a_{m+h|m}$ and its prediction error variance matrix $P_{m+h|m}$ for $h = 1, 2, \ldots$. The observation forecasts
165 $\hat{y}_{m+h|m} = \mathrm{E}(y_{m+h}|y_1, \ldots, y_m)$ and the error variance matrix $V_{m+h|m} = \mathrm{Var}(y_{m+h} - \hat{y}_{m+h}|y_1, \ldots, y_m)$ are then computed by

$$\hat{y}_{m+h|m} = Z_{m+h} a_{m+h|m}, \qquad V_{m+h|m} = Z_{m+h} P_{m+h|m} Z_{m+h}' + H_{m+h},$$

166 This simple treatment of missing observations and forecasting is one of the advantageous features of state space analysis.

167   Observations are also weighted when forecasting. The recursions above implicitly provide the optimal weights to the available
168 observations for forecasting, in terms of minimum mean squared error. For a linear time series model with time-varying components
169 as in our case, the forecasting weights gradually decline when observations are further distanced from the forecast origin point, since
170 they become less relevant. In this way, at the longer lead times the forecast weights are smaller for the dynamic cyclical components,
171 and the weight of the explanatory regression variable component is bigger, which makes it then more significant than it is at the
172 shorter lead times.

173   The presented results in Fig. 3 are based on the parameter estimates from the period 1952-1970 for the prediction of the time
174 series between 1972-1993, and from the period 1974-1992 for the prediction of the time series between 1994-2015. Importantly, the
175 forecasts in Figs. 1, 2 and 4 are performed using all available observations before each prediction point for parameter estimation,
176 component estimation and forecasting. This explains the difference between the reported correlation coefficients between the forecasts
177 and observations presented in Fig. 1 and those presented in Fig. 3. We would like to stress that the skill of the model is most
178 accurately reflected by the results in Fig. 1. However, to avoid heavy computations, we have produced the forecasts in Fig. 3 based
179 on a pre-fixed period for calibration purposes, instead of updating the information every time a new observation becomes available.
180 Also note that every point in the forecasts in Fig. 4 is predicted at the respective lead time indicated in the figure. This is different
181 from the forecasts shown in Fig. 9 of ref. 8, where only the target month of December was predicted at the respective lead time. All
182 presented results here are based on the same data sets for SST, subsurface temperature and wind stress as described in ref. 8.

References.

188

189   1. M.A. Cane, S.E. Zebiak, and S.C. Dolan. Experimental forecasts of El Niño. *Nature*, 321:827–832, 1986.
190   2. D. Chen, S. Zebiak, A.J. Busalacchi, and M. Cane. An improved procedure for El Niño forecasting: implications for predictability. *Science*, 269:1699–1702, 1995.
191   3. M. Latif, D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, and E. Schneider. A review of the predictability and prediction of ENSO. *Journal of Geophysical Research*, 103:
192   14375–14393, 1999.
193   4. D. Chen, M. Cane, A. Kaplan, S. Zebiak, and D. Huang. Predictability of El Niño over the past 148 years. *Nature*, 428:15, 2004.
194   5. P. Gonzalez and L. Goddard. Long-lead ENSO predictability from CMIP5 decadal hindcasts. *Climate Dynamics*, 46:3127–3147, 2016.
195   6. A. Barnston, M. Tippett, M. L′Heureux, Sh. Li, and D. DeWitt. Skill of real-time seasonal ENSO model predictions during 2002-11. Is our capability increasing? *American Meteorological Society*, 93:631–651,
196   2012.
197   7. M. McPhaden and X. Yu. Equatorial waves and the 1997/98 El Niño. *Geophysical Research Letters*, 26:2961–2964, 1999.
198   8. D. Petrova, S.J. Koopman, J. Ballester, and X. Rodó. Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model. *Climate Dynamics*, 48:
199   1249–1276, 2017.
200   9. M. McPhaden, A. Busalacchi, R. Cheney, J-R. Donguy, K. Gage, D. Halpern, M. Ji, P. Julian, G. Meyers, G. Mitchum, et al. The Tropical Ocean-Global Atmosphere observing system: A decade of progress.
201   *Journal of Geophysical Research: Oceans*, 103:14169–14240, 1998.
202   10. E. Sarachik and M. Cane. *The El Niño Southern Oscillation Phenomenon*. Cambridge University Press, 2010.
203   11. A. Fedorov, S. Harper, S. G. Philander, B. Winter, and A. Wittenberg. How predictable is El Niño? *Bulletin of the American Meteorological Society*, 84:911–919, 2003.

204 12. A. V. Fedorov, Sh. Hu, M. Lengaigne, and E. Guilyardi. The impact of westerly wind bursts and ocean initial state on the development, and diversity of El Niño events. *Climate Dynamics*, 44:1381–1401, 2015.

205 13. A. Wittenberg. Are historical records sufficient to constrain ENSO simulations. *Geophysical Research Letters*, 36:L12702, 2009.

206 14. M.A. Cane, M. Munnich, and S.E. Zebiak. A study of self-excited oscillations of the tropical ocean-atmosphere system. Part 1: linear analysis. *Journal of Atmospheric Sciences*, 47:1562–1577, 1990.

207 15. F. F. Jin, J. D. Neelin, and M. Ghil. El Niño on the devil's staircase - annual subharmonic steps to chaos. *Science*, 264:70–72, 1994.

208 16. F. F. Jin. An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model. *Journal of Atmospheric Sciences*, 54:811–829, 1997.

209 17. B. N. Goswami and J. Shukla. Predictability of a coupled ocean-atmosphere model. *Journal of Climate*, 4:3–22, 1991.

210 18. D. Chen and M. A. Cane. El Niño prediction and predictability. *Journal of Computational Physics*, 227:3625–3640, 2008.

211 19. A. Barnston. Long-lead seasonal forecasts: where do we stand? *Bulletin of the American Meteorological Society*, 75:2097–2114, 1994.

212 20. K. Wyrtki. Water displacements in the Pacific and the genesis of El Niño cycles. *Journal of Geophysical Research*, 90:7129–7132, 1985.

213 21. M. McPhaden. Evolution of the 2002/2003 El Niño. *American Meteorological Society*, 85:677–695, 2004.

214 22. N. Ramesh and R. Murtugudde. All flavours of El Niño have similar early subsurface origins. *Nature Climate Change*, 3:42–46, 2013.

215 23. J. Ballester, S. Bordoni, D. Petrova, and X. Rodó. On the dynamical mechanism explaining the western Pacific subsurface temperature buildup leading to ENSO events. *Geophysical Research Letter*, 42:
216 2961–2967, 2015.

217 24. A. Barnston and Ch. Ropelewski. Prediction of ENSO episodes using Canonical Correlation Analysis. *Journal of Climate*, 5:1316–1345, 1992.

218 25. C. Penland and T. Magorian. Prediction of Niño3 sea surface temperatures using linear inverse modeling. *Journal of Climate*, 6:1067–1076, 1993.

219 26. Y. Xue, M. Cane, S. Zebiak, and M. Blumenthal. On the prediction of ENSO: a study with a low-order Markov model. *Tellus A*, 46:512–528, 1994.

220 27. D. Kondrashov, S. Kravtsov, A. Robertson, and M. Ghil. A hierarchy of data-based ENSO models. *Journal of Climate*, 18:4425–4444, 2005.

221 28. CPC. Cold and warm episodes by season. *http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml*, 2016.

222 29. T.N. Stockdale, D. Anderson, M. Balmaseda, F. Doblas-Reyes, L. Ferranti, K. Mogensen, T. Palmer, F. Molteni, and F. Vitart. ECMWF seasonal forecast system 3 and its prediction of sea surface temperature.
223 *Climate Dynamics*, 37:455–471, 2011.

224 30. NOAA. TAO Array Figure. *http://www.pmel.noaa.gov/tao/images/87_94.gif*, 1994.

225 31. NOAA. TAO-TRITON Array. *http://www.pmel.noaa.gov/tao/*, 2016.

226 32. J. Ballester, D. Petrova, S. Bordoni, B. Cash, M. García-Díez, and X. Rodó. Sensitivity of El Niño intensity and timing to preceding subsurface heat magnitude. *Scientific Reports*, 2016.

227 33. S. E. Zebiak and M. A. Cane. A model El Niño-Southern Oscillation. *Monthly Weather Review*, 115:2262–2278, 1987.

228 34. M. Balmaseda and D. Anderson. Impact of initialization strategies and observations on seasonal forecast skill. *Geophysical Research Letters*, 36, 2009.

229 35. E. Guilyardi. El Niño-mean state-seasonal cycle interactions in multi-model ensemble. *Climate Dynamics*, 26:329–348, 2006.

230 36. A. Harvey. *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press, 1989.

231 37. J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods.* Oxford University Press, 2 edition, 2012.
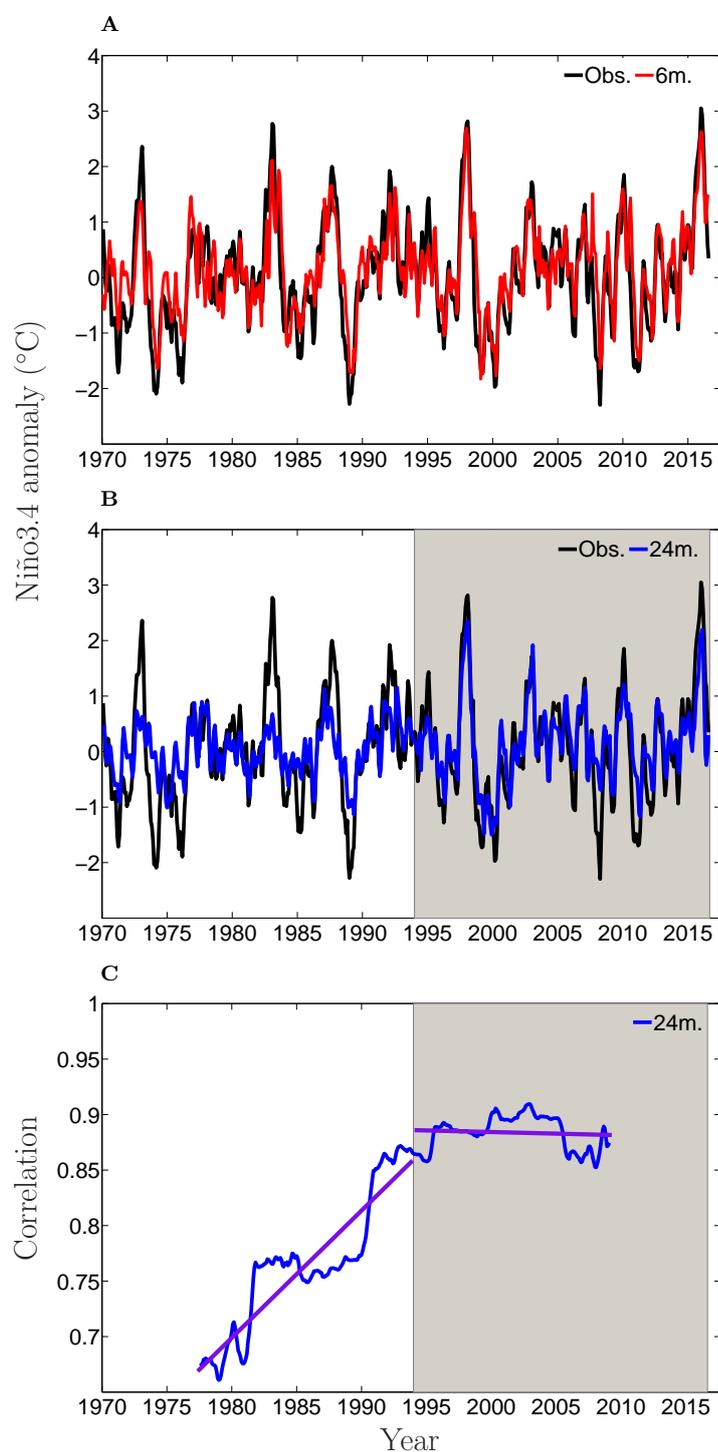
Fig. 1. Retrospective prediction of the Niño3.4 Index. Monthly observations (black curve) and model prediction at (**A**) 6-month lead (red curve) and (**B**) 24-month lead (blue curve). (**C**) 15-year moving correlation between the observations and the prediction in (B) (blue curve), and the piecewise linear fitting (purple lines) before and after (shading) the completion of the observing system in 1994.
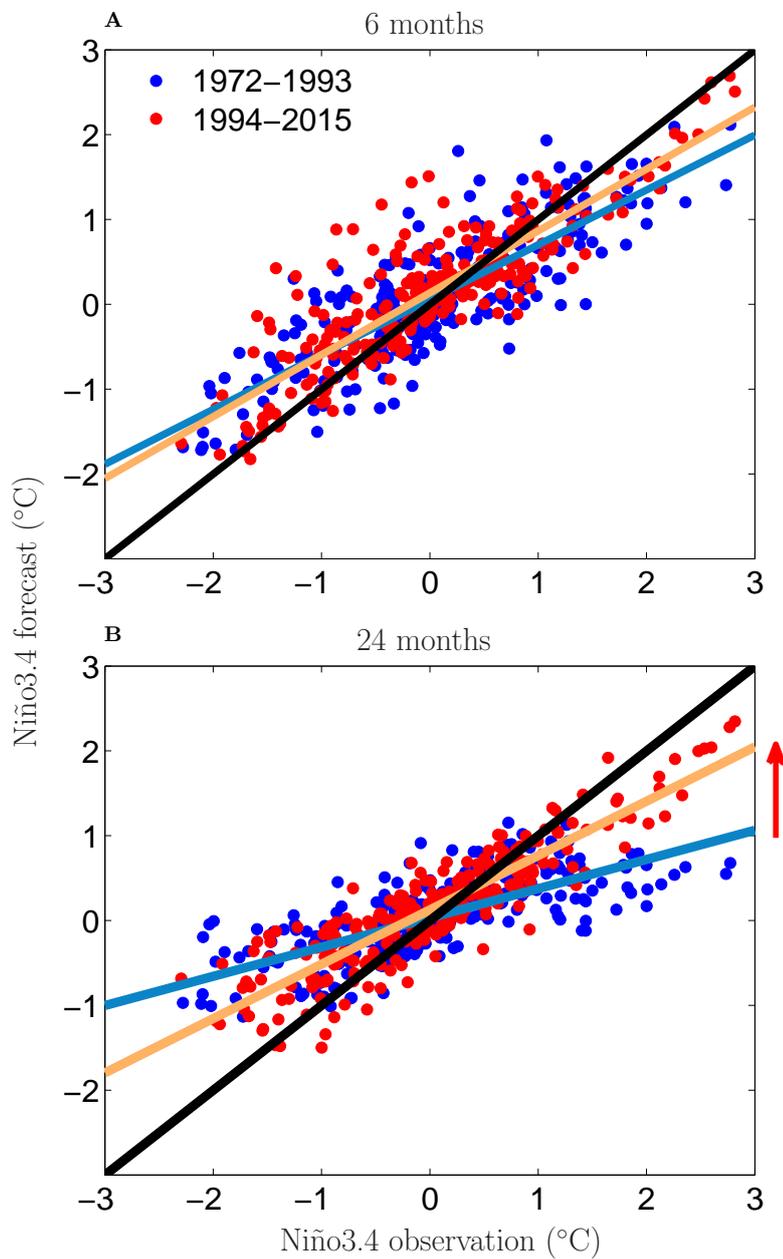
Fig. 2. Relationship between observations and model predictions. Scatter plots of the Niño3.4 Index observations and the model predictions at (**A**) 6-month lead and (**B**) 24-month lead. The blue dots correspond to the period 1972-1993 with a linear regression line in light blue, and the red dots correspond to the period 1994-2015 with a linear regression line in beige. The red arrow indicates the improvement in the slope of the regression line for the period 1994-2015 with respect to the slope of the regression line for the period 1972-1993.
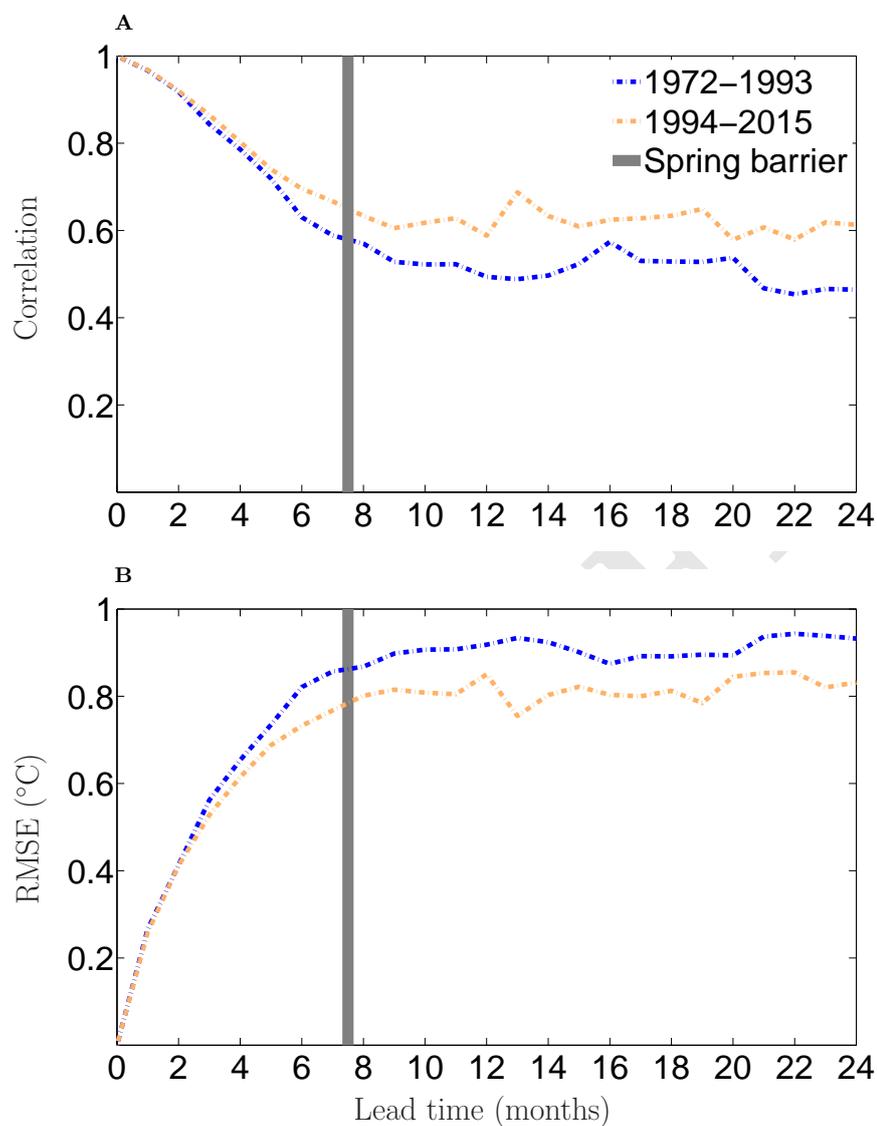
Fig. 3. General forecast skill of the model. (A) Correlations between the Niño3.4 Index observations and model predictions and **(B)** root mean square errors (RMSE) as functions of lead time for two consecutive 22-year periods, 1972-1993 (blue) and 1994-2015 (beige).
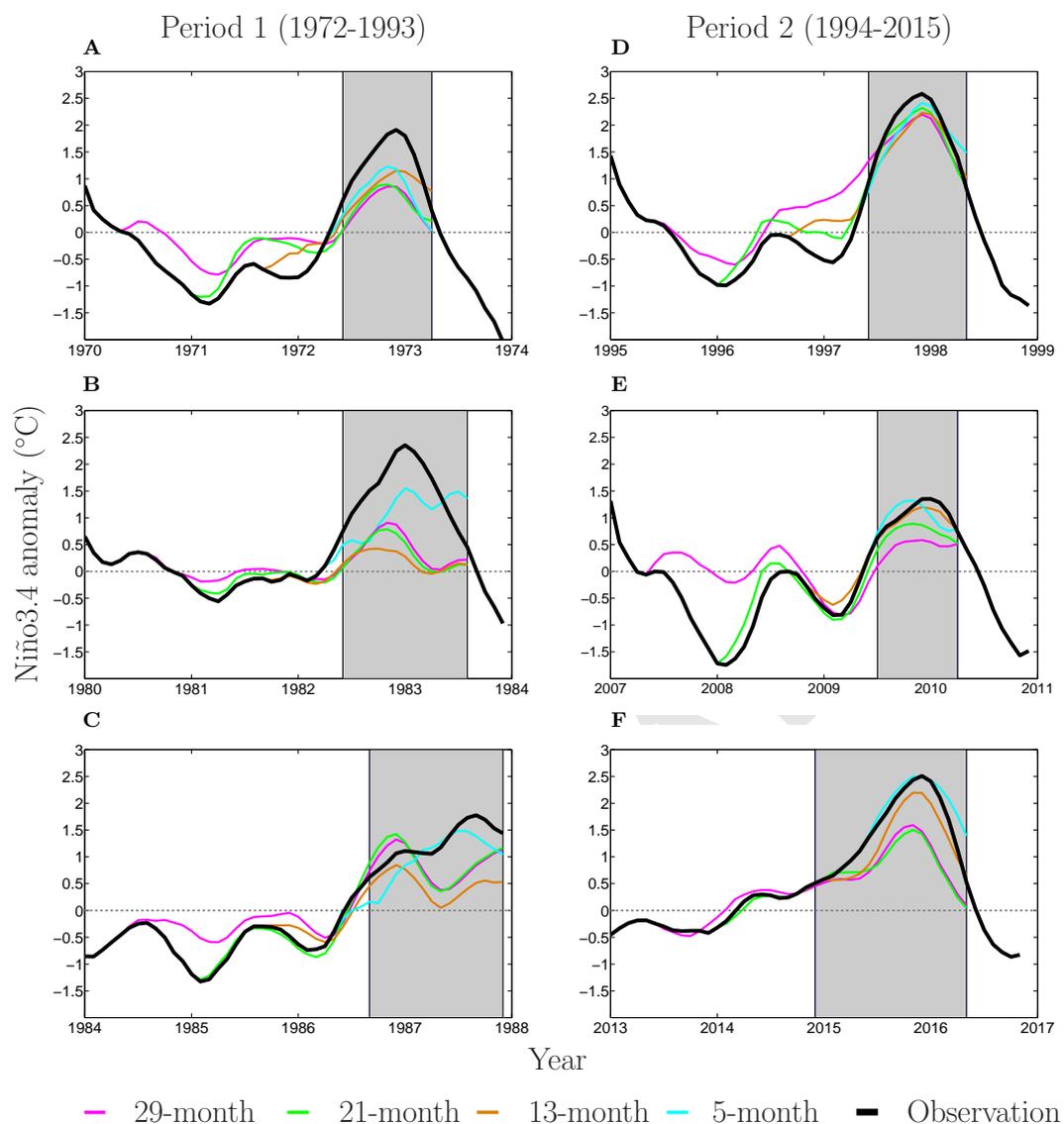
Fig. 4. Forecasts of the major El Niño events since 1970. (A)-(C) El Niño events in the period 1972-1993 and **(D)-(F)** 1994-2015. The thick black curves are the observed Niño3.4 Index anomalies, and the thin magenta, green, beige and cyan curves are predictions started 29, 21, 13 and 5 months in advance.

Table 1: Coefficients, t-values and p-values for subsurface temperature predictor regression variables at 24-month lead. Values significant at the 90% level are bold.

El Niño event	250m. RI	300m. RI	400m. RI
1972/73			
Coefficient	0.12	−0.17	−0.29
t	0.78	−0.82	−0.86
p	0.43	0.41	0.39
1982/83			
Coefficient	0.09	0.01	0.20
t	0.78	0.03	0.90
p	0.43	0.97	0.36
1986/87			
Coefficient	−0.03	−0.12	−0.02
t	−0.30	−0.88	−0.09
p	0.76	0.37	0.92
1991/92			
Coefficient	0.07	−0.09	0.09
t	0.64	−0.56	0.38
p	0.52	0.57	0.70
1997/98			
Coefficient	**0.24**	0.35	0.46
t	**1.61**	1.52	1.46
p	**0.10**	0.12	0.14
2002/03			
Coefficient	**0.21**	0.31	0.38
t	**1.67**	1.57	1.44
p	**0.09**	0.11	0.15
2006/07			
Coefficient	**0.23**	**0.32**	**0.43**
t	**2.07**	**1.80**	**1.75**
p	**0.04**	**0.07**	**0.08**
2009/10			
Coefficient	**0.17**	0.24	**0.46**
t	**1.68**	1.46	**1.95**
p	**0.09**	0.14	**0.05**
2014/15			
Coefficient	**0.15**	**0.25**	0.34
t	**1.61**	**1.63**	1.59
p	**0.10**	**0.10**	0.11
2015/16			
Coefficient	0.14	**0.28**	0.32
t	1.55	**1.85**	1.56
p	0.12	**0.06**	0.12

Supplementary Table 1: Coefficients, *t*-values and *p*-values for subsurface temperature predictor regression variables at 21-month lead. Values significant at the 90% level are bold.

El Niño event	250m. RII	wnd RI
1972/73		
Coefficient	−0.08	−0.70
t	−0.62	−0.26
p	0.53	0.79
1982/83		
Coefficient	−0.07	−0.88
t	−0.70	−0.40
p	0.48	0.68
1986/87		
Coefficient	−0.05	−0.32
t	−0.50	−0.16
p	0.61	0.87
1991/92		
Coefficient	−0.14	0.48
t	−1.27	0.23
p	0.20	0.81
1997/98		
Coefficient	**-0.33**	3.96
t	**-1.96**	1.34
p	**0.05**	0.18
2002/03		
Coefficient	**-0.34**	**4.67**
t	**-2.08**	**1.83**
p	**0.03**	**0.06**
2006/07		
Coefficient	−0.24	**4.41**
t	−1.60	**1.92**
p	0.11	**0.05**
2009/10		
Coefficient	**-0.30**	**4.07**
t	**-2.10**	**2.02**
p	**0.03**	**0.04**
2014/15		
Coefficient	**-0.20**	**4.05**
t	**-1.67**	**2.53**
p	**0.09**	**0.01**
2015/16		
Coefficient	−0.14	**3.59**
t	−1.19	**2.25**
p	0.23	**0.02**

Supplementary Table 2: Coefficients, t-values and p-values for subsurface temperature predictor regression variables at 29-month lead. Coefficients significant at the 90% level are bold.

El Niño event	250m. RI	300m. RI	400m. RI
1972/73			
Coefficient	0.06	0.04	−0.51
t	0.38	0.18	−1.25
p	0.70	0.85	0.21
1982/83			
Coefficient	−0.07	−0.01	0.07
t	−0.55	−0.07	0.30
p	0.58	0.94	0.76
1986/87			
Coefficient	0.04	0.07	0.18
t	0.37	0.50	0.81
p	0.71	0.61	0.41
1991/92			
Coefficient	−0.04	0.13	0.22
t	−0.36	0.72	0.91
p	0.71	0.47	0.36
1997/98			
Coefficient	**0.25**	**0.37**	**0.57**
t	**1.73**	**1.68**	**1.88**
p	**0.08**	**0.09**	**0.06**
2002/03			
Coefficient	0.15	0.28	**0.50**
t	1.15	1.35	**1.78**
p	0.25	0.17	**0.07**
2006/07			
Coefficient	0.18	**0.38**	**0.40**
t	1.60	**2.09**	**1.64**
p	0.11	**0.03**	**0.10**
2009/10			
Coefficient	**0.18**	**0.29**	0.32
t	**1.68**	**1.71**	1.38
p	**0.09**	**0.08**	0.16
2014/15			
Coefficient	**0.16**	**0.29**	**0.41**
t	**1.64**	**1.87**	**1.89**
p	**0.10**	**0.06**	**0.06**
2015/16			
Coefficient	0.11	0.25	**0.38**
t	1.23	1.60	**1.79**
p	0.22	0.11	**0.07**

Chapter 4

Appendices

4.1 Long-lead El Niño forecasts coupled to a dengue model vastly extend the predictability of epidemics in southern coastal Ecuador

Petrova D (1,2), Lowe R (2,3), Stewart-Ibarra A (4), Ballester J (2), Koopman S J (5), Rodó X (1,2,6)

(1) Institut Català de Ciències del Clima, Barcelona, Catalonia, Spain

(2) Barcelona Institute for Global Health, Barcelona, Catalonia, Spain

(3) Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.

(4) SUNY Upstate Medical University, New York, USA

(5) Vrije Universiteit Amsterdam, Amsterdam, Netherlands

(6) Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

Long-lead El Niño forecasts coupled to a dengue model vastly extend the

predictability of epidemics in southern coastal Ecuador

Desislava Petrova[*]

Catalan Institute for Climate Sciences, Barcelona, Spain

Rachel Lowe

Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical

Medicine, London, UK.

A. Stewart-Ibarra

SUNY Upstate Medical University, New York, USA.

Joan Ballester

Climate and Health Programme, Barcelona Institute of Global Health, Barcelona, Spain

S.J. Koopman

VU Amsterdam, Amsterdam, Netherlands

Xavier Rodó

Climate and Health Program, Barcelona Institute of Global Health, Barcelona, Spain.

[*]*Corresponding author address:* Climate Dynamics and Impacts Unit, Catalan Institute for Climate

Sciences (IC3), Carrer del Dr. Trueta, 203, 08005, Barcelona, Spain.

[18] E-mail: desislava.petrova@isglobal.org

ABSTRACT

El Niño-Southern Oscillation (ENSO) is a high-impact climatic phenomenon that causes extreme weather events worldwide. It leads to floods or droughts in certain regions, damages agriculture and the economy, and increases the risk of infectious diseases. Therefore, ENSO forecasts could help authorities to plan in advance of imminent disasters, to mitigate the risk, and to protect vulnerable communities. Certain diseases are particularly sensitive to climate extremes. For example, a previous study found that the timing and magnitude of dengue outbreaks in El Oro province in Ecuador were associated with El Niño. In this study, long-lead forecasts of equatorial Pacific sea surface temperatures (SST) are used within a dengue prediction model, to assess the extent to which epidemics can be predicted well in advance.

A structural time series model, which uses a state space approach and predictors relevant to the El Niño evolution, was developed to predict SST in the Niño 3.4 region. The model configuration is specifically tailored to forecast the events at long lead times of 2 years or more, well beyond the traditional spring barrier of ENSO prediction. The forecasting scheme provides information about the amplitude of the events, their duration, and the peak time of the SST. The model output was then used within a dengue prediction model to estimate the incidence during the 2010 epidemic in El Oro, Ecuador.

The ENSO forecasting model correctly forecast the 2009-2010 El Niño, which could have helped to predict the dengue outbreak as early as 30 months ahead. Thus, long-lead ENSO forecasts could be incorporated into dengue prediction models, to enhance the development of a dengue warning system for Ecuador. This framework could be extended to other tropical and subtropical countries, which are directly and severely affected by the anomalous

4

temperature and precipitation rates during and after El Niño.

1. Introduction

ENSO is a climate phenomenon characterized by coupled ocean-atmosphere anomalies in the tropical Pacific. It is commonly defined and assessed on the basis of SST deviations in the Niño 3.4 region (N3.4; [120-170°W, 5°S-5°N]) and with the help of the Oceanic Niño Index (ONI), which represents the three-month running-mean of SST anomalies in this region. El Niño (EN; the warm phase) is associated with a positive ONI $\geq$ +0.5 °C for a period of at least 5 consecutive overlapping three-month seasons, while La Niña (LN; the cold phase) is associated with a negative ONI $\leq$ -0.5 °C for the same period of time (CPC 2017). The abnormally high or low temperatures of the ocean affect the atmosphere directly above as a result of the strong ocean-atmosphere coupling in the equatorial region (Bjerknes 1969), and thus trigger atmospheric teleconnections that change the usual weather patterns globally (Ropelewski and Halpert 1987; Kiladis and Diaz 1989; Rodó et al. 2006; Sarachik and Cane 2010), but especially in the vicinity of the equator. For example, in Ecuador the warm phase of the oscillation induces extreme events such as severe and prolonged deluges that affect a sizeable proportion of the population in the country (Larkin and Harrison 2002; Rossel and Cadier 2009; Recalde-Coronel et al. 2014).

ENSO is a main driver of tropical climate and triggers atmospheric teleconnections that directly impact the health, resources and livelihoods of millions of people worldwide. Therefore, predicting particular ENSO events well in advance is of high relevance to decision makers in charge of planning and preparing for climate-related disasters. At the same time, few studies exist on long-lead predictions of ENSO on the order of more than 1 year in advance (Chen et al. 2004; Ludescher et al. 2014; Gonzalez and Goddard 2016). A recent study, Petrova et al. (2016), documented unprecedented forecasts of past El Niño events started more than two years ahead of peak SST anomalies. All the EN events that occurred in the period 1996-2015 were predicted

5

at long lead times in a retrospective forecasting experiment. The forecasting scheme is based on structural time series modelling and analysis by state space methods (Durbin and Koopman 2012). Additionally, predictors capturing the state of the atmosphere and ocean at different stages of the development of an El Niño are incorporated in a manner so that specific variables such as wind stress, surface and subsurface ocean temperature are only used at their relative time of importance (see Petrova et al. (2016) for details). The long-lead capabilities of the model and the ENSO information that it provides - estimates of the timing, magnitude and duration of the events - could be used as input in other climate and tropical disease prediction models, which would help to anticipate and assess the risks for the nations in the tropics, but also in other parts of the world where ENSO is known to have high impact.

The climate of Ecuador is heavily affected by the inter-annual ENSO variability. The country does not have well-distinguished seasons in regards to temperature, the mean ranges between 21 - 28°C, and thus is only slightly higher during boreal winter (Figure 1(a), (b)). There is a marked wet season, however, which lasts from January to May (Figure 1(e),(f); Bendix and Lauer 1992; Moran-Tejeda et al. 2016). Rainfall is highly related to the meridional movement of the Inter-Tropical Convergence Zone (ITCZ). When it is in its southernmost position it brings warm and moist air to coastal regions, resulting in increased precipitation rates and higher temperatures. Conversely, when the ITCZ is in its northernmost position, the coast is affected by the upwelling processes in the equatorial Pacific, and by drier and cooler air masses (Figure 1(c),(d),(g),(h); Rossel and Cadier 2009).

During El Niño strong convection triggered by warmer SST results in higher temperatures in southern coastal Ecuador (Figure 2(a)-(d); Aceituno 1988; Bendix and Bendix 2006; Rossel and Cadier 2009; Moran-Tejeda et al. 2016), and much heavier precipitation (Figure 2(e)-(h)) than observed during neutral or LN years (at least double the amount). Moreover, the peak of ENSO,

⁹² which is normally at the end of the year (November-February), coincides with the rainy season in

⁹³ Ecuador (Figure 1(e)), and is immediately followed by the peak of precipitation anomalies associ-

⁹⁴ ated with ENSO in southern Ecuador in February to March (Figure 2(g)-(h); Bendix and Bendix

⁹⁵ 2006). Thus, the normal precipitation patterns in the area are significantly enhanced during EN.

⁹⁶ It was also shown that the ENSO-associated SST anomalies and coastal precipitation are highly

⁹⁷ correlated (with correlation coefficients up to 0.8 for Dec.-May period during EN; Coelho et al.

⁹⁸ 2002). Therefore, ENSO has predictive capacity for precipitation rates in the area (Rossel and

⁹⁹ Cadier 2009). In addition, a positive relationship, albeit lower in magnitude, was also found be-

¹⁰⁰ tween EN and maximum and minimum temperature in the region (Moran-Tejeda et al. 2016).

¹⁰¹ The effect of El Niño events on local precipitation and temperature, could lead to an increase

¹⁰² in the frequency of vector-borne diseases, such as dengue (Hsieh and Chen 2009). In a previous

¹⁰³ study (Stewart-Ibarra and Lowe 2013), a statistical mixed model was developed to assess the im-

¹⁰⁴ portance of climatic and non-climatic drivers of dengue interannual variability in southern coastal

¹⁰⁵ Ecuador (in El Oro province). The authors found that the ONI, rainfall, and minimum temperature

¹⁰⁶ were positively associated with dengue, with more cases of dengue expected during El Niño events

¹⁰⁷ (Figure 3). Field studies in the same region also found that rainfall and minimum temperature were

¹⁰⁸ key drivers of *Aedes aegypti* dynamics (Stewart Ibarra et al. 2013), providing mechanistic evidence

¹⁰⁹ for the influence of local climate on dengue transmission. Moreover, dengue transmission in El

¹¹⁰ Oro has a well-defined seasonal pattern, and most cases are reported from February to May, which

¹¹¹ is the period of the year characterized by warmer and rainy weather (see Figure 1 in Stewart-Ibarra

¹¹² and Lowe (2013)).

¹¹³ The aim of this study is to test the application of this recently developed ENSO forecasts to epi-

¹¹⁴ demic dengue forecasts in an endemic region. For this reason, we use forecasts of the 2009/10 El

¹¹⁵ Niño event and the calculated ONI at multiple lead times, derived from the structural time series

model, to produce predicted dengue distributions for El Oro province in southern coastal Ecuador for March 2010. In 2010 the province experienced the most severe dengue epidemic on record, with approximately 4008 suspected cases. People under 20 years of age bore the greatest burden of disease (Stewart Ibarra et al. 2014a). Statistical and wavelet analyses revealed that the epidemic may have been triggered by above normal minimum temperature and above normal rainfall during the 2009/10 EN event (Stewart-Ibarra and Lowe 2013; Stewart Ibarra et al. 2014b). Rainfall in February, preceding the peak of transmission, was almost double the long-term average. Social risk factors during the epidemic included household demographics (i.e., age and gender of heads of household) and housing conditions (i.e., access to piped water, poor housing construction) (Stewart Ibarra et al. 2014b). Other regions also experienced major dengue epidemics in the same year, including Puerto Rico, which reported the largest historical outbreak (21,000 cases) (CDC 2010). Globally, the greatest number of deaths due to dengue were reported in 2010, in the period from 1990 to 2013 (Stanaway et al. 2016).

We describe the ENSO and dengue models and how they were combined (Section 2), present the forecasts (Section 3), and discuss what are the implications for climate services for health (Section 4).

2. Modelling and Data

a. ENSO Forecasting Model

A structural time series model described in Petrova et al. (2016) was used to predict the 2009/10 El Niño event. The model is built in terms of unobserved components - a trend, a seasonal, three cycle components, an irregular term and some explanatory regression variables. It is given by:

$$y_t = \mu_t + \gamma_t + \psi_{1t} + \psi_{2t} + \psi_{3t} + x_t'\delta + \varepsilon_t \tag{1}$$

137 where y_t represents the monthly Niño 3.4 index at time t; μ_t is the trend component specified as

138 a random walk process; γ_t is the seasonal component; ψ_{1t}, ψ_{2t} and ψ_{3t} are three cycles with dif-

139 ferent frequencies λ_j, $(j = 1, 2, 3)$, persistences φ_{ψ_j} and variances $\sigma^2_{\kappa_j}$; $x'_t\delta$ represents a predictor

140 regression variable; and ε_t represents the irregular term. The trend, seasonal, and cycle compo-

141 nents are modeled as linear dynamic stochastic functions of time (Harvey and Koopman 2000).

142 More information about the components can be found in Harvey (1989) and Durbin and Koopman

143 (2012).

144 An important feature of the structural time series model is that it can be embedded in a state

145 space framework (Durbin and Koopman 2012), in which all unknown parameters associated with

146 the model components are put in state and disturbance vectors, and estimated together in a dy-

147 namic way using the Kalman Filter (Kalman 1960). Then forecasting is performed through signal

148 extraction of the components, and likelihood evaluation. STAMP, SsfPack and OxMetrics (Koop-

149 man et al. 2008, 2010; Doornik 2013) are used for the estimations and forecasts.

150 As there is a long-term variability in the mean of the Niño 3.4 time series, it is incorporated in the

151 model through the trend component. ENSO is also associated with a phase-locking to the annual

152 cycle (Rasmusson and Carpenter 1982; An and Choi 2009; Stein et al. 2011), as its two phases

153 usually peak in the Northern Hemisphere winter months of November, December, January (NDJ;

154 Sarachik and Cane 2010) and decay by the following summer. Thus, seasonality in the tropical

155 Pacific is of prime importance for the timely growth and decline of ENSO (Tziperman et al. 1997;

156 Krishnamurthy et al. 2015), and hence, for its accurate prediction. In this model it is accounted

157 for by the seasonal component. It has also been previously shown that other cyclical processes

158 on inter-annual scales that occur in the equatorial Pacific ocean and atmosphere play a role in the

159 formation of the ENSO phenomenon (Jin et al. 2003; Rasmusson et al. 1990; Jiang et al. 1995 and

160 others). These variability modes have been found to have periods close to 1.5 years (near-annual

161 mode), 2-2.5 years (quasi-biannual mode), and 4-6 years (quasi-quadrennial mode). They are in-

162 cluded in the model in the form of stochastic cycle components. For more information regarding

163 the dynamical interpretation of these cycles the reader is referred to Petrova et al. (2016).

164 In addition to the components discussed above, a set of predictor regression variables are added

165 to account for more specific dynamical variation of the Niño 3.4 time series. These include surface

166 and subsurface ocean temperature and wind stress in the equatorial western and central Pacific.

167 The variables were selected on the basis of the dynamical processes in the equatorial Pacific that

168 occur up to three years prior to the peak of an El Niño event.

169 ENSO is a mechanism through which, as a result of the prevailing easterly zonal winds, heat is

170 accumulated in the western to central equatorial Pacific subsurface ocean, and then released to the

171 atmosphere or to higher latitudes after a saturation point is reached (Wyrtki 1985; Zebiak 1989;

172 Jin 1997). In this way, the zonal wind in the tropics directly affects the heat content of the ocean

173 via Kelvin and Rossby downwelling and upwelling waves, which in turn influence SSTs (Ballester

174 et al. 2015). The change in SSTs then affects the zonal wind patterns, and the so-called Bjerknes

175 feedback between the ocean and the atmosphere is thus activated (Bjerknes 1969). Considering

176 these processes, we have selected regions in the tropical Pacific from which to obtain information

177 about the coupled system in the form of time series, which are used as explanatory covariates in

178 the ENSO forecasting model. For more details about the relevant dynamics, how these regions are

179 defined, and about the predictor variables the reader is referred to Petrova et al. (2016).

180 The structural time series ENSO model described above has successfully predicted all the major

181 El Niño and La Niña events in the study period (Figure 4). For the purposes of the current analysis,

182 it was then used to predict the 2009/10 moderate-to-strong El Niño event at long lead times. The

183 aim was to use this early information and try to increase the prediction window for the dengue

184 epidemics that is known to have occurred in southern coastal Ecuador after this particular warm

event (Stewart-Ibarra and Lowe 2013). The data sets used for the predictions are wind stress from

the NCEP/NCAR reanalysis (Kalnay et al. 1996), sea surface temperature from the NOAA-OI-

SST-V2 data provided by the NOAA/OAR/ESRL PSD (www.esrl.noaa.gov/psd/), and subsurface

temperature data from the Subsurface Temperature And Salinity Analyses by Ishii et al. (2005),

archived at the National Center for Atmospheric Research, Computational and Information Sys-

tems Laboratory (www.rda.ucar.edu/datasets/ds285.3/).

b. Dengue Forecasting Model

A statistical mixed model, described in Stewart-Ibarra and Lowe (2013), was used to predict

dengue incidence in March 2010 for the province of El Oro, in southern coastal Ecuador. Briefly,

a negative binomial generalised linear mixed model was formulated using monthly dengue cases

data as the response variable and the expected number of cases, based on the underlying popu-

lation, as the model offset (Lowe et al. 2011, 2013). Dengue is a mandatory notifiable disease,

and cases included all suspected (clinically diagnosed) cases reported to the Ministry of Health

surveillance system. Explanatory variables included the ONI, local monthly rainfall and minimum

temperature anomalies from the Granja Santa Ines weather station located in Machala, Ecuador

($3°17''26''$ S, $79°54'5''$ W, 5 m.a.s.l.) for the period 1995-2010, the mean monthly proportion of

households with *Aedes aegypti* immatures (House Index) for the province of El Oro provided by

the National Service for the Control of Vector-Borne Diseases, and the number of serotypes circu-

lating in the country each month in the period 2001-2010. Serotypes are reported by the National

Reference Center (NRC) for Dengue and other Arboviruses, the official diagnostic laboratory of

the Ministry of Health. Temporally autocorrelated random effects for each calendar month were

included to account for confounding factors, which might influence the annual cycle of dengue as

well as climate, introducing an extra source of variability into the model. Yearly random effects

208 were also included to capture non-climate related inter-annual variability, such as changes in vec-

209 tor control or reporting practices.

210 The model, comprising a combination of lagged climate variables, non-climate covariates and

211 temporally structured and unstructured random effects, is formulated as follows:

$$y_t \sim NegBin(\mu_t, k) \tag{2}$$

$$log(\mu_t) = log(e_t) + \alpha + \beta_{t'(t)} + \sum y_j x_{jt} + \delta_{T'(t)} + \sum \varepsilon_j z_{jt} \tag{3}$$

212 where y_t is the monthly dengue cases, k is the overdispersion parameter, μ_t is the mean dengue

213 cases and e_t is the model offset. The relative risk then includes a model intercept, α, temporally

214 autocorrelated random effects for each calendar month, $\beta_{t'(t)}$, and exchangeable non-structured

215 random effects for each year, $\delta_{T'(t)}$. The variables x_{jt} represent climate variables: anomalies of

216 precipitation ($j = 1$) and minimum temperature ($j = 2$), and the ONI index ($j = 3$). The variables

217 z_{jt} represent non-climate factors: the House Index and the number of serotypes circulating in the

218 country.

219 Model parameters were estimated in a Bayesian framework using Markov Chain Monte Carlo

220 (MCMC; Gilks et al. 1996). This approach accounts for parameter uncertainty by assigning prior

221 distributions to the parameters, with associated MCMC sampling yielding samples from the full

222 posterior predictive distribution of dengue in any given month.

223 The structured temporal random effects were assigned a first-order autoregressive month effect

224 for each month ($t'(t) = 2, ..., 12$) with month 1 (January) aliased to the model intercept, α, and

225 subsequent months following a random walk or first difference prior in which each effect is de-

226 rived from the immediately preceding effect. The temporally unstructured random year effects

227 were assigned independent diffuse Gaussian exchangeable prior distributions.

228 The most significant variables associated with dengue variations were found to be the ONI

lagged by three months, and the Aedes aegypti larval House Index in the previous month. It was shown that a 1°C increase in SST anomaly in Niño3.4 would result in 28% increase in dengue cases 3 months later, while 1% increase in the number of households with Aedes aegypti immatures would result in 48% increase in dengue cases one month later. Anomalies in minimum temperature (Tmin) with a time lag of two months, and anomalies in precipitation with a time lag of one month were also important climate predictors (see Stewart-Ibarra and Lowe (2013) for further details). Interestingly, ONI was found to be a more important climate variable than both Tmin and precipitation, which could imply that the impact of ENSO is not limited to these two variables. Finally, ENSO was shown to have a strong positive association Tmin, while the association with precipitation was only weak. In Figure 3(a) it can be seen that all El Niño events in the study period are related to positive anomalies in the minimum temperature, and similarly to dengue anomalies in the region (Figure 3(b) and (c)). However, only the stronger EN are related to positive precipitation anomalies (Supplementary figure A?).

3. Results

a. Forecast of the 2009/10 El Niño

Shown in Figure 5(a) are forecasts of the N3.4 sea surface temperature anomalies at lead times between 29 and 4 months prior to the peak of the 2009/10 El Niño together with the observed anomalies (black curve). The explanatory covariates used for forecasting are subsurface ocean temperature and zonal wind stress in the western and central tropical Pacific at their significant lag times as discussed in Section 2a. A warm event is foreseen at all lead times, and as early as 29 months (magenta curve in Figure 5(a)) in advance the forecasting model already indicates a positive anomaly of approximately +1 °C for the peak month of December 2009. As lead time

decreases, forecasts and observations become more convergent, and the model prediction skill increases. An exception is the 8 month lead forecast (green curve in Figure 5(a)), which is generally less accurate than the 12 month lead one. We attribute this to the "spring barrier" in ENSO prediction, and the fact that forecasts for the end of the year started in the months March-May tend to be less accurate due to the presence of more noise in the atmosphere-ocean system (Sarachik and Cane 2010; Barnston et al. 2012). One full year before the peak of the EN (red curve in Figure 5(a)), the forecast predicts the correct value of the warm peak in the N3.4 region of +1.46 °C at the correct time (December 2009).

Depicted in Figure 5(b) is the ONI for NDJ, calculated from the forecasts in Figure 5(a). The 0 lead time corresponds to the observed values for the ONI at the end of 2009/beginning of 2010 (+1.3 °C for NDJ (CPC 2017)). It is evident that the forecasting scheme is capable of providing highly accurate information about the amplitude of the event one year in advance of the mature phase of El Niño (magenta curve in Figure 5(b)). Moreover, a medium-sized warm event (ONI≈+1 °C) is predicted for the eastern equatorial Pacific as early as 21 months ahead.

b. Forecast of the 2010 dengue outbreak in El Oro

Figure 6 shows the posterior predictive distribution of dengue cases for March 2010 in El Oro province. In this figure we show the skill of the model with observed data in order to be able to compare it later with forecasts from the ENSO model. The explanatory variables used to produce the dengue simulations here included the House Index in the previous month (February 2010), anomalies of precipitation in the previous month (February 2010) and minimum temperature two months previous (January 2010), and the ONI for the NDJ period (all other variables are held constant). The posterior predictive mean (dashed line) and observed dengue cases (arrow) are coincidental. In Figure 7, the posterior predictive distribution of dengue cases for March 2010 is

simulated several times, varying the ONI value each time. In Figure 7(a), the observed NDJ ONI value is used to produce the dengue simulation (i.e. a dengue forecast with a 3 month lead time). In Figure 7(b)-(f), forecast NDJ ONI values are used to simulate dengue, providing lead times of 7, 11, 15, 20 and 24 months with respect to the March 2010 dengue peak. Note, all other explanatory variables were held constant. Climatological values of precipitation and temperature anomalies over previous years (i.e. climatology) and the House Index for the previous year (February 2009) were used (at relevant time lags).

The dengue simulation in Figure 6 provides a baseline reference forecast for which to compare dengue simulations produced using forecasts of ONI from the ENSO model (see Figure 7). The skill of the dengue predictions gradually decreases with increasing lead time. However, even at long lead times, the posterior mean prediction is contained within the 95% credible intervals of the dengue predictive distribution.

4. Discussion and Conclusion

The results described in the previous section demonstrate that it would have been possible to predict the 2010 dengue outbreak in southern coastal Ecuador long before its occurrence (2 years earlier) by using information about the ONI from an ENSO forecasting model. ONI was found to be the key climate explanatory variable for dengue outbreaks in this region by Stewart-Ibarra and Lowe (2013). Moreover, it has a strong inter-annual signal and a weaker seasonal signal, which makes it a good predictor. Also, considering that local temperature and precipitation correlate with ONI, it should be sufficient to use only the climatological values of these variables in combination with the ONI in order to explain the variability in dengue incidence that is related to climate. In this way, even at the very long lead time of 24 months the dengue model used here is skilful at detecting the high chance of a dengue epidemics in March 2010 (Figure 7(f)), which actually

15

happened in the aftermath of the 2009/10 El Niño event. Therefore, long-lead El Niño information could theoretically be used to enhance the prediction of dengue incidence and issue early warning of such epidemics in Ecuador, and thus help to mitigate some of their worst effects.

However, these results must be interpreted with caution. It should be noted that not all of the outbreaks of dengue in El Oro are associated with EN events. In Figure 3(c), for example, it is visible that a couple of outbreaks occurred at a time that was not during El Niño. Other factors not related to climate, such as the introduction of a new serotype or a lapse in vector control could trigger a sizeable epidemic in the region. Additionally, the positive association between warm events and dengue incidence in southern coastal Ecuador has been found for the recent period since 1995, and in the cases of the 1997/98, 2002/03, 2004/05, 2006/07 and 2009/10 EN events in particular (see Figure 2 in Stewart-Ibarra and Lowe (2013)). In fact, for validation purposes, the same forecasts as for the 2009/10 EN and dengue epidemic were repeated for the prominent 1997/98 EN event and the epidemic that followed it (Supplementary figures A1 and A2), and the dengue prediction was highly accurate even at the 24 months lead time. Still, although the effect of EN on precipitation and temperature in Ecuador has been relatively similar for the whole observational record, it is not clear that this positive association between the warm phase of ENSO and dengue would be preserved with time and with the changing climate (Tabachnick 2010). Finally, it should also be recognized that El Niño events are variable. For example, there are eastern Pacific canonical EN and central Pacific "Modoki" EN (Ashok et al. 2007), and they could be associated with possibly different teleconnections (Dewitte et al. 2012), and possible differences in the time lags that characterize them. Moreover, different types of El Niño could be predicted with various rates of success (Barnston et al. 2012), and the smaller events are generally harder to predict. Thus, in an operational forecasting framework it could be the case that a false alarm for EN, for instance, could result also in a false alarm for a dengue epidemic. In this way the credibility

16

of the information provided to the end users could be limited by the predictability of the climate

system, in this case represented by the metric of ONI. It is even more probable that an outbreak

could occur in a non-El Niño year, and one such outbreak actually happened in early 2006, which

was a La Niña year. To prevent such false alarms, the two models briefly presented here should

be recalibrated regularly, the forecast information should be updated frequently (as often as new

information becomes available), and the level of uncertainty should be clearly communicated at

each step.

Although the afore-mentioned issues are important and must be taken into consideration, it is

still valid that a skilful ENSO forecasting scheme could provide highly valuable information and

make possible the development of early warning systems for forthcoming disasters (Connor and

Mantilla 2008). In this study we demonstrated the ability of an ENSO forecasting model to predict

a particular El Niño event and to deliver long-lead forecast information, which was then used in

a partially climate-driven dengue prediction scheme to assess the risk of a dengue epidemic in El

Oro in Ecuador in 2010. The ENSO model successfully predicted the 2009/10 EN at the very

long lead time of 21 months, which provided a lead time of 24 months for the dengue forecast.

Thus, the dengue outbreak that occurred in southern coastal Ecuador in the aftermath of this El

Niño could in theory have been foreseen 2 years in advance. The added value of such a long-

lead prediction is mostly economic, as public funds for health, and in particular for epidemic

prevention and vaccination campaigns could be well planned and optimized. This example of

a potentially successful early warning system and collaborative analysis between the spheres of

climate science and infectious disease serves to demonstrate the benefit of climate services to

health decision makers and professionals, and ultimately to society. Considering that dengue is

a disease that poses substantial strains to the health care systems in tropical countries such as

Ecuador, the results shown here open a possibility for the development of an operational scientific

345 tool to provide local decision makers with targeted health information, and to support and help

346 them in planning their mitigation and adaptation measures.

360 **References**

361 Aceituno, P., 1988: On the functioning of the Southern Oscillation in the South America sector

362   Part I: Sufrace climate. *Monthly Weather Review*, **116**, 505–524.

363 An, S. I., and J. Choi, 2009: Seasonal locking of the ENSO asymmetry and its influence on the

364   seasonal cycle of the tropical eastern Pacific sea surface temperature. *Atmospheric Research*,

365   **94**, 3–9.

366 Ashok, K., S. Behera, R. S.A., H. Weng, and T. Yamagata, 2007: El Niño Modoki and its possible

367 teleconnection. *Journal of Geophysical Research: Oceans*, **112**, –.

368 Ballester, J., S. Bordoni, D. Petrova, and X. Rodo, 2015: On the dynamical mechanism explain-

369 ing the western pacific subsurface temperature buildup leading to ENSO events. *Geophysical*

370 *Research Letter*, **42**, 2961–2967.

371 Barnston, A., M. Tippett, M. L′Heureux, S. Li, and D. DeWitt, 2012: Skill of real-time seasonal

372 ENSO model predictions during 2002-11. Is our capability increasing? *American Meteorologi-*

373 *cal Society*, **93**, 631–651.

374 Bendix, A., and J. Bendix, 2006: Heavy rainfall episodes in Ecuador during El Niño events and

375 associated regional atmospheric circulation and sst patterns. *Advances in Geosciences*, **6**, 43–49.

376 Bendix, J., and W. Lauer, 1992: The rainy seasons in Ecuador and their climatic interpretation.

377 *Erdkunde*, **46**, 118–134.

378 Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather*

379 *Review*, **97**, 163–172.

380 CDC, 2010: Largest dengue outbreak in puerto rico history. CDC Dengue Update 2011; 3. *http:*

381 *//www.cdc.gov/dengue/dengue_upd/resources/DengueUpdateVo3No1.pdf* .

382 Chen, D., M. Cane, A. Kaplan, S. Zebiak, and D. Huang, 2004: Predictability of El Niño over the

383 past 148 years. *Nature*, **428**, 15.

384 Coelho, C., C. Uvo, and T. Ambrizzi, 2002: Exploring the impacts of the tropical Pacific SST

385 on the precipitation patterns over south america during ENSO periods. *Theoretical and Applied*

386 *Climatology*, **71**, 185–197.

Connor, S., and G. Mantilla, 2008: *Seasonal Forecasts, Climatic Change and Human Health.* Springer.

CPC, 2017: Cold and warm episodes by season. *http://www.cpc.ncep.noaa.gov/products/ analysis_monitoring/ensostuff/ensoyears.shtml.*

Dewitte, B., and Coauthors, 2012: Change in El Niño flavours over 1958–2008: Implications for the long-term trend of the upwelling off Peru. *Deep Sea Research Part II: Topical Studies in Oceanography*, **77**, 143–156.

Doornik, J. A., 2013: *Object-Oriented Matrix Programming using Ox 7.0.* Timberlake Consultants Ltd, London, see `http://www.doornik.com`.

Durbin, J., and S. J. Koopman, 2012: *Time Series Analysis by State Space Methods.* 2nd ed., Oxford University Press.

Gilks, W., S. Richardson, and D. Spiegelhalter, 1996: *Markov Chain Monte Carlo in Practice.* Boca Raton, Florida, USA: Chapman and Hall/CRC, 486.

Gonzalez, P., and L. Goddard, 2016: Long-lead ENSO predictability from cmip5 decadal hindcasts. *Climate Dynamics*, **46**, 3127–3147.

Harvey, A., 1989: *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press.

Harvey, A., and S. J. Koopman, 2000: Signal extraction and the formulation of unobserved components models. *The Econometrics Journal*, **3**, 84–107.

Hsieh, Y., and C. Chen, 2009: Turning points, reproduction number, and impact of climatological events for multi-wave dengue outbreaks. *Trop. Med. Int. Health*, **14**, 628–638.

408 Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective analyses of SST and marine

409 meteorological variables for the 20th century using COADS and the Kobe Collection. *Interna-*

410 *tional Journal of Climatology*, **25**, 865–879.

411 Jiang, N., D. Neelin, and M. Ghill, 1995: Quasi-quadrennial and quasi-biennial variability in the

412 equatorial Pacific. *Journal of Climate Dynamics*, **12**, 291–310.

413 Jin, F. F., 1997: An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model.

414 *Journal of Atmospheric Sciences*, **54**, 811–829.

415 Jin, F. F., J. S. Kug, S. An, and I. S. Kang, 2003: A near-annual coupled ocean-atmosphere mode

416 in the equatorial Pacific. *Geophysical Research Letters*, **30**.

417 Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Journal of Basic*

418 *Engineering , Transactions, ASMA, Series D*, **82**, 35–45.

419 Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the*

420 *American Meteorological Society*, **77**, 437–471.

421 Kiladis, G., and H. F. Diaz, 1989: Global climatic anomalies associated with extremes in the

422 Southern Oscillation. *Journal of Climate*, **2**, 1069–1090.

423 Koopman, S. J., A. C. Harvey, J. A. Doornik, and N. Shephard, 2010: *Stamp 8.3: Structural Time*

424 *Series Analyser, Modeller and Predictor*. Timberlake Consultants, London.

425 Koopman, S. J., N. Shephard, and J. A. Doornik, 2008: *Statistical Algorithms for Models in State*

426 *Space Form: SsfPack 3.0*. Timberlake Consultants Press, London.

427 Krishnamurthy, L., G. Vecchi, R. Msadek, A. Wittenberg, T. Delworth, and F. Zeng, 2015: The

428 seasonality of the great plains low-level jet and ENSO relationship. *Journal of Climate*, **28**,

429 4525–4544.

430 Larkin, N., and D. Harrison, 2002: Enso warm (El Niño) and cold (La Niña) event life cycles:

431 Ocean surface anomaly patterns, their symmetries, asymmetries, and implications. *Journal of*

432 *Climate*, **15**, 1118–1140.

433 Lowe, R., T. Bailey, D. Stephenson, R. Graham, C. Coelho, M. S Carvalho, and C. Barcellos,

434 2011: Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning

435 system for dengue in Brazil. *Computational Geoscience*, **37**, 371–381.

436 Lowe, R., T. Bailey, D. Stephenson, T. Jupp, R. Graham, C. Barcellos, and M. Carvalho, 2013:

437 The development of an early warning system for climate-sensitive disease risk with a focus on

438 dengue epidemics in Southeast Brazil. *Stat. Med.*, **32**, 864–883.

439 Ludescher, J., A. Gozolchiani, M. Bogachev, A. Bunde, S. Havlin, and H. Schellnhuber, 2014:

440 Very early warning of next El Niño. *PNAS*, **111(6)**, 2064–2066.

441 Moran-Tejeda, E., and Coauthors, 2016: Climate trends and variability in Ecuador (19662011).

442 *International Journal of Climatology*, –.

443 Petrova, D., S. Koopman, J. Ballester, and X. Rodo, 2016: Improving the long-lead predictability

444 of El Niño using a novel forecasting scheme based on a dynamic components model. *Climate*

445 *Dynamics*.

446 Rasmusson, E., and T. Carpenter, 1982: Variations in tropical sea surface temperature and surface

447 wind fields associated with the Southern Oscillation/El Niño. *Mon. Weather Rev.*, **110**, 354–384.

448 Rasmusson, E., X. Wang, and C. Ropelewski, 1990: The biennial component of ENSO variability.

449 *Journal of Marine Systems*, **1**, 71–96.

Recalde-Coronel, C., A. Barnston, and A. Muñoz, 2014: Predictability of December–April rainfall in coastal and Andean Ecuador. *Journal of Applied Meteorology and Climatology*, **53**, 1471–1493.

Rodó, X., M. Rodriguez-Arias, and J. Ballester, 2006: The role of ENSO in fostering teleconnection patterns between the tropical north Atlantic and the western Mediterranean basin. *CLIVAR Exchanges*, **11**, 26–27.

Ropelewski, C., and M. Halpert, 1987: Global and regional scale precipitation patterns associated with El Niño/Southern Oscillation. *Monthly Weather Review*, **115**, 1606–1626.

Rossel, F., and E. Cadier, 2009: El niño and prediction of anomalous monthly rainfalls in Ecuador. *Hydrological Processes*, **23**, 3253–3260.

Sarachik, E., and M. Cane, 2010: *The El Niño Southern Oscillation Phenomenon*. Cambridge University Press.

Stanaway, J., D. Shepard, E. Undurraga, and et al., 2016: The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *Lancet Infectious Disease*.

Stein, K., A. Timmermann, and N. Shcneider, 2011: Phase synchronization of El Niño - Southern Oscillation with the annual cycle. *Physical Review Letters*, **107**.

Stewart-Ibarra, A., and R. Lowe, 2013: Climate and non-climate drivers of dengue epidemics in southern coastal Ecuador. *Am. J. Trop. Med. Hyg.*, **88**, 971–981.

Stewart Ibarra, A. M., V. Luzadis, M. Borbor-Cordova, and et al., 2014a: A social-ecological analysis of community perceptions of dengue fever and Aedes aegypti in Machala, Ecuador. *BMC Public Health*, 1135–.

Stewart Ibarra, A. M., A. Munoz, S. Ryan, and et al, 2014b: Spatiotemporal clustering, climate periodicity, and social-ecological risk factors for dengue during an outbreak in Machala, Ecuador, in 2010. *BMC Infect Dis*, **14**, 610–.

Stewart Ibarra, A. M., S. Ryan, and E. Beltran, 2013: Dengue vector dynamics (Aedes Aegypti) influenced by climate and social factors in Ecuador: Implications for targeted control. *PLOS ONE*, **8**, 11.

Tabachnick, W. J., 2010: Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. *Journal of Experimental Biology*, **213**, 946–954.

Tziperman, E., S. Zebiak, and M. Cane, 1997: Mechanisms of seasonal - ENSO interaction. *Journal of Atmospheric Sciences*, **54**, 61–71.

Wyrtki, K., 1985: Water displacements in the Pacific and the genesis of El Niño cycles. *Journal of Geophysical Research*, **90**, 7129–7132.

Zebiak, S., 1989: Oceanic heat content variability and El Niño cycles. *Journal of Physical Oceanography*, **19**, 475–486.

[FIG. 1 about here.]

[FIG. 2 about here.]

[FIG. 3 about here.]

[FIG. 4 about here.]

[FIG. 5 about here.]

[FIG. 6 about here.]

24

491 [FIG. 7 about here.]

492 [FIG. 8 about here.]

493 [FIG. 9 about here.]

26

FIG. 1. Mean climatology of (a)-(d) minimum surface temperature (°C), and (e)-(h) precipitation (mm/month) for (a), (e) January-March, (b), (f) April-June, (c), (g) July-September, (d), (h) October-December seasons. The region of El Oro Province, Ecuador is located within the red rectangle.
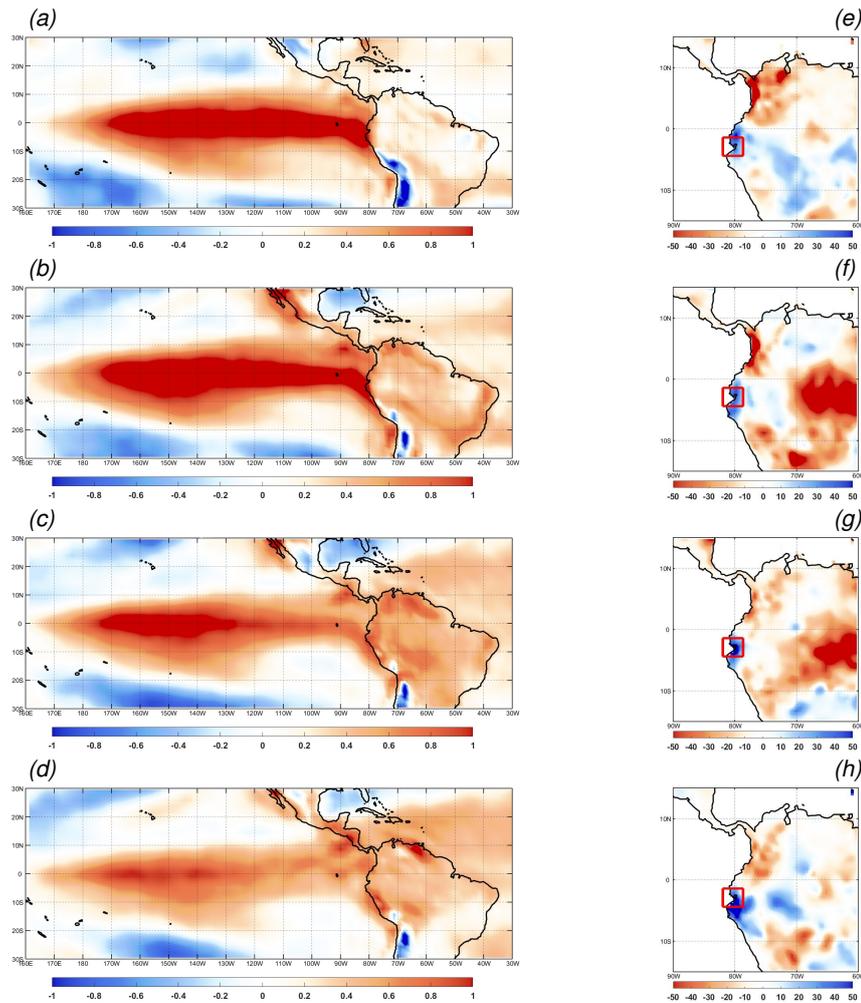
FIG. 2. Anomalies with respect to composites for El Niño of minimum surface temperature (°C) at lag (a) 0, (b) +1, (c) +2, (d) +3 and precipitation (mm/month) at lag (e) 0, (f) +1, (g) +2, (h) +3. The El Niño events used for the calculation of composite anomalies are Dec. 1982, 1991, 1994, 1997, 2002, 2004, 2006, 2009. The region of El Oro Province, Ecuador is located within the red rectangle.
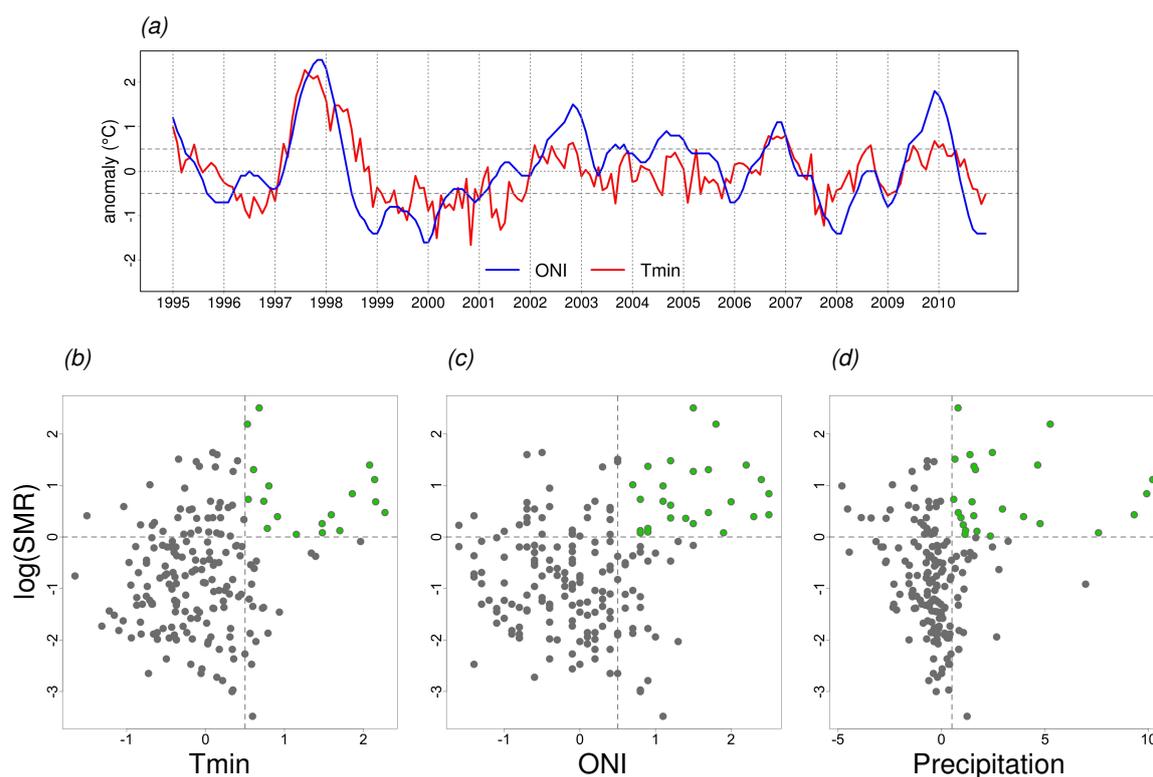
FIG. 3. (a) Time series of the Oceanic Niño Index (blue) and minimum temperature monthly anomalies (red) from the Granja Santa Ines weather station located in Machala, El Oro, from January 1995 - December 2010. Dashed lines indicate the ±0.5 (°C) threshold for El Niño and La Niña events. Scatter-plots of (b) dengue standardised incidence and minimum temperature monthly anomalies in El Oro province, (c) dengue standardised incidence anomalies in El Oro province and the Oceanic Niño Index, (d) dengue standardised incidence and precipitation monthly anomalies in El Oro province. Green dots correspond to Tmin, ONI and precipitation anomaly values > +0.5.
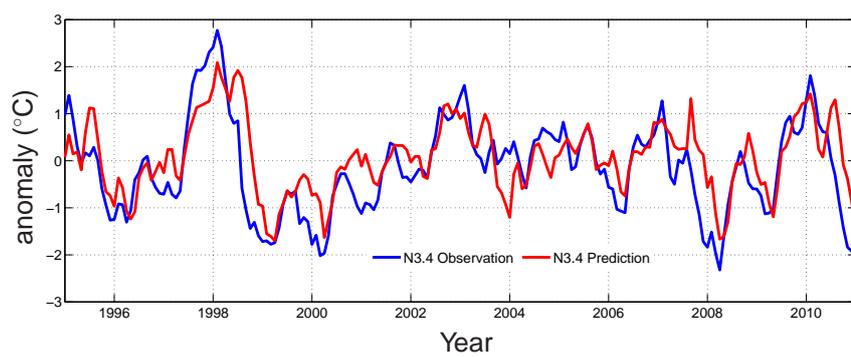
31

FIG. 4. Time series of the Niño3.4 Index (°C), (blue) observation (red) 6-month-lead prediction from January 1995 - December 2010.
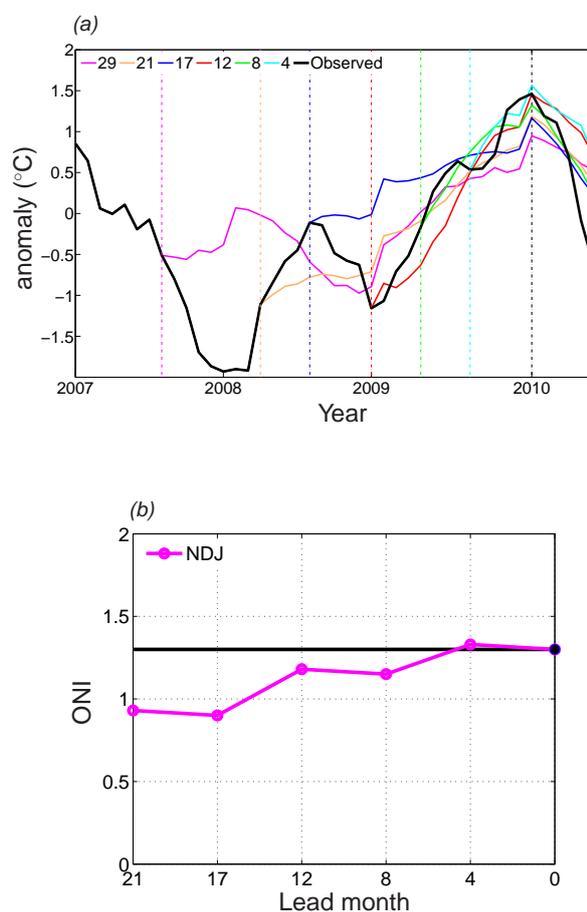
FIG. 5. (a) Forecast of the 2009/10 El Niño event (N3.4 Index) starting 29 (magenta), 21 (beige), 17 (blue), 12 (red), 8 (green) and 4 (cyan) months ahead of the boreal winter peak. Observations are in black. The dashed lines indicate the respective starting months for the forecasts, and the black dashed line indicates the peak of the observed anomaly. (b) NDJ ONI (magenta) calculated from the forecasts started in the respective lead month before the peak of the 2009/10 El Niño. Lead month 0 corresponds to the observed value (black line and filled circle).
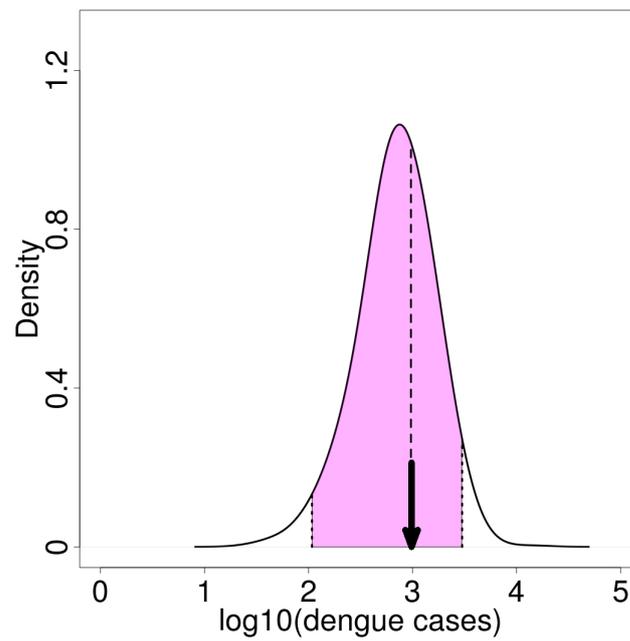
FIG. 6. Posterior predictive distribution of dengue cases (base-10 logarithmic scale) for March 2010, El Oro province, Ecuador. Explanatory variables include local precipitation (1 month lag) and minimum temperature (2 month lag) anomalies, House Index (1 month lag) and ONI (3 month lag observed value for NDJ). The posterior predictive mean (dashed line), 95% credible intervals of the predictive distribution (dotted lines) and observed dengue cases (arrow) are indicated.
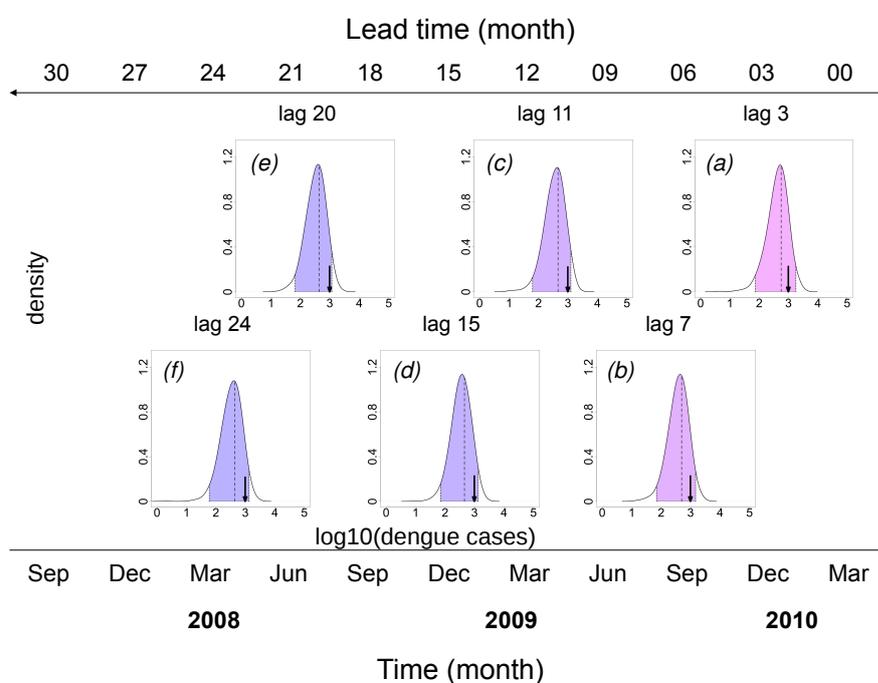
FIG. 7. Posterior predictive distribution of dengue cases (base-10 logarithmic scale) for March 2010, El Oro province, Ecuador, using (a) observed ONI value for NDJ (3), and forecast ONI values for NDJ at lead times of (b) 4 (7), (c) 8 (11), (d) 12 (15), (e) 17 (20), and (f) 21 (24) months with respect to the 2009/10 El Niño (with respect to the peak in dengue). All other explanatory variables are held constant. The posterior predictive mean (dashed line), 95% credible intervals of the predictive distribution (dotted lines) and observed dengue cases (arrow) are indicated.
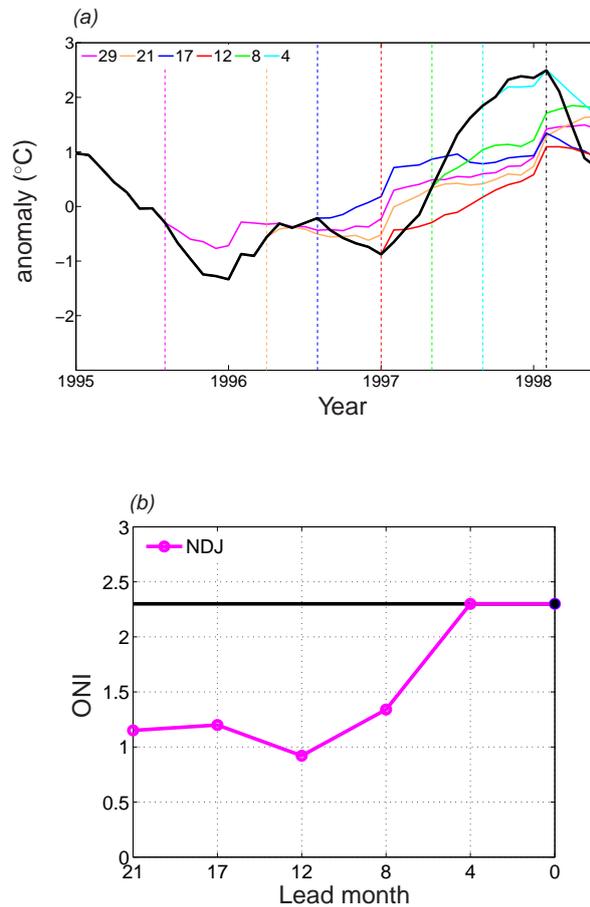
Fig. A1. (a) Forecast of the 1997/98 El Niño event (N3.4 Index) starting 29 (magenta), 21 (beige), 17 (blue), 12 (red), 8 (green) and 4 (cyan) months ahead of the boreal winter peak. Observations are in black. The dashed lines indicate the respective starting months for the forecasts, and the black dashed line indicates the peak of the observed anomaly. (b) NDJ ONI (magenta) calculated from the forecasts started in the respective lead month before the peak of the 1997/98 El Niño. Lead month 0 corresponds to the observed value (black line and filled circle).
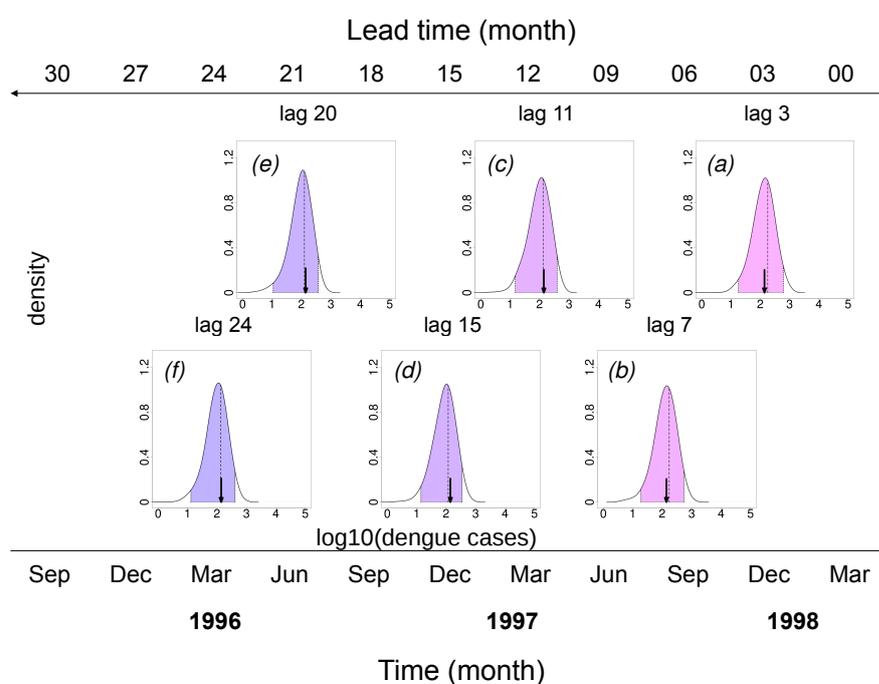
36

Fig. A2. Posterior predictive distribution of dengue cases (base-10 logarithmic scale) for March 1998, El Oro province, Ecuador, using (a) observed ONI value for NDJ (3), and forecast ONI values for NDJ at lead times of (b) 4 (7), (c) 8 (11), (d) 12 (15), (e) 17 (20), and (f) 21 (24) months with respect to the 1997/98 El Niño. All other explanatory variables are held constant. The posterior predictive mean (dashed line), 95% credible intervals of the predictive distribution (dotted lines) and observed dengue cases (arrow) are indicated.

4.2   Climate services for health: predicting the evolution of the 2016 dengue season in Machala, Ecuador

Lowe R (1,2), Stewart-Ibarra A (3), Petrova D (2,4), García-Díez M (5), Borbor-Cordova M J(6), Mejía R (7), Regato M (8), Rodó X (1,4,9)

(1) Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.

(2) Barcelona Institute for Global Health, Barcelona, Catalonia, Spain

(3) SUNY Upstate Medical University, New York, USA

(4) Institut Català de Ciències del Clima, Barcelona, Catalonia, Spain

(5) Predictia Intelligent Data Solutions, Santander, Cantabria, Spain

(6) Escuela Superior Politecnica del Litoral (ESPOL), Guayaquil, Ecuador

(7) National Institute of Meteorology and Hydrology (INAMHI), Guayaquil, Ecuador

(8) National Institute of Public Health Research (INSPI) of the Ministry of Health, Guayaquil, Ecuador

(9) Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

Manuscript Number:

Title: Climate services for health: predicting the evolution of the 2016 dengue season in Machala, Ecuador

Article Type: Article

Corresponding Author: Dr. Rachel Lowe,

Corresponding Author's Institution: London School of Hygiene & Tropical Medicine

First Author: Rachel Lowe

Order of Authors: Rachel Lowe; Anna M Stewart-Ibarra; Desislava Petrova; Markel García-Díez; Mercy J Borbor-Cordova; Raul Mejía; Mary Regato; Xavier Rodó

Abstract: Background: El Niño and its impact on local meteorological conditions potentially influences interannual variability in dengue transmission in southern coastal Ecuador. El Oro province is a key dengue surveillance site, due the high burden of dengue, seasonal transmission, co-circulation of all four dengue serotypes, and the recent introduction of chikungunya and Zika. In this study, we used climate forecasts to predict the evolution of the 2016 dengue season in the city of Machala, following one of the strongest El Niño events on record.

Methods: We incorporated precipitation, temperature and Oceanic Niño Index forecasts in a Bayesian hierarchical mixed model to predict dengue incidence. The model was initiated on 1 January 2016, producing monthly dengue forecasts until October 2016. We accounted for misreporting due to the introduction of chikungunya in 2015, by using active surveillance data to correct reported dengue case data. We then evaluated the forecast retrospectively with available epidemiological information.

Findings: The predictions correctly forecast an early peak in dengue incidence in March 2016, with a 91% chance of exceeding the mean dengue incidence for the previous five years. Accounting for the proportion of chikungunya cases that had been incorrectly recorded as dengue in 2015 improved the prediction of the magnitude of dengue incidence in 2016.

Interpretation: The main advantage of this dengue prediction framework is the use of long-lead seasonal climate and El Niño forecasts, which permits a prediction to be made at the start of the year for the entire dengue season. Combing active surveillance data with routine dengue reports improved not only model fit and performance, but also the accuracy of benchmark estimates based on historic seasonal averages. This study advances the state-of-the-art of climate services for the health, by demonstrating the potential value of incorporating climate information in the public health decision-making process in Ecuador.

Funding: European Union FP7; Royal Society; National Science Foundation.

Climate services for health: predicting the evolution of the 2016 dengue season in Machala, Ecuador

Rachel Lowe[1,2,*] (PhD), Anna M. Stewart-Ibarra[3] (PhD), Desislava Petrova[2] (MSc), Markel García-Díez[4] (PhD), Mercy J. Borbor-Cordova[5] (PhD), Raul Mejía[6] (MSc), Mary Regato[7] (MD), Xavier Rodó[2,8] (PhD)

1. Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, Keppel Street, London WC1E 7HT, UK.
2. CLIMA – Climate and Health Programme, Barcelona Institute for Global Health (ISGLOBAL), Doctor Aiguader, 88, 08003 Barcelona, Spain.
3. Center for Global Health and Translational Science and Department of Medicine, SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY, 13210, USA.
4. Predictia Intelligent Data Solutions, Av. de los Castros, 39005 Santander, Spain.
5. Escuela Superior Politecnica del Litoral (ESPOL), Km 30., Vía Perimetral 5, Guayaquil, Ecuador.
6. National Institute of Meteorology and Hydrology (INAMHI), Pichincha 307 9 de Octubre, Guayaquil, Ecuador.
7. National Institute of Public Health Research (INSPI) of the Ministry of Health, Julián Coronel 905, Guayaquil, Ecuador.
8. Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluís Companys, 23, 08010 Barcelona, Spain.

*Correspondence to: Dr Rachel Lowe, Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK, Tel: +44 (0) 20 7958 8229 Email: rachel.lowe@lshtm.ac.uk

Summary

Background: El Niño and its impact on local meteorological conditions potentially influences interannual variability in dengue transmission in southern coastal Ecuador. El Oro province is a key dengue surveillance site, due the high burden of dengue, seasonal transmission, co-circulation of all four dengue serotypes, and the recent introduction of chikungunya and Zika. In this study, we used climate forecasts to predict the evolution of the 2016 dengue season in the city of Machala, following one of the strongest El Niño events on record.

Methods: We incorporated precipitation, temperature and Oceanic Niño Index forecasts in a Bayesian hierarchical mixed model to predict dengue incidence. The model was initiated on 1 January 2016, producing monthly dengue forecasts until October 2016. We accounted for misreporting due to the introduction of chikungunya in 2015, by using active surveillance data to correct reported dengue case data. We then evaluated the forecast retrospectively with available epidemiological information.

Findings: The predictions correctly forecast an early peak in dengue incidence in March 2016, with a 91% chance of exceeding the mean dengue incidence for the previous five years. Accounting for the proportion of chikungunya cases that had been incorrectly recorded as dengue in 2015 improved the prediction of the magnitude of dengue incidence in 2016.

Interpretation: The main advantage of this dengue prediction framework is the use of long-lead seasonal climate and El Niño forecasts, which permits a prediction to be made at the start of the year for the entire dengue season. Combing active surveillance data with routine dengue reports improved not only model fit and performance, but also the accuracy of benchmark estimates based on historic seasonal averages. This study advances the state-of-the-art of climate services for the health, by demonstrating the potential value of incorporating climate information in the public health decision-making process in Ecuador.

Funding: European Union FP7; Royal Society; National Science Foundation.

Introduction

The burden of dengue fever (DENV: family Flaviviridae, genus flavivirus) has increased globally over the last three decades, from an estimated 8·3 million apparent (symptomatic) cases in 1990, to 58·4 million cases in 2013.[1] The World Health Organization and others have advocated the use of climate information to manage the increasing burden of dengue as part of comprehensive early warning and response systems.[2–4] Predictions of higher than expected dengue incidence (e.g., outbreaks) can optimise the allocation of scarce resources through targeted and focused interventions. Prior studies have shown that climate information, such as seasonal climate forecasts, can be used to improve predictions of dengue outbreaks months in advance.[5,6]

Dengue is sensitive to changes in climate conditions as temperature affects the physiology of the aedes aegypti and aedes albopictus mosquito vectors (e.g., biting and larval development rates)[7,8] and the rate of viral replication in the mosquito.[9,10] Both rainfall and drought can increase the availability of larval mosquito habitats (i.e., containers with standing water), depending on local water storage practices and piped water infrastructure.[11,12]

The El Niño Southern Oscillation (ENSO) is the strongest inter-annual climate cycle on Earth. It occurs in the equatorial Pacific Ocean, and affects weather patterns worldwide through atmospheric teleconnections. Typical examples include excess rainfall in Peru and Ecuador, dry conditions in Indonesia, and a decrease in the number of typhoons in the western Pacific during the warm phase of the cycle, and more or less symmetrical anomalies during the cold phase.[13] To monitor, assess, and predict ENSO, the climate research community has defined the Oceanic Niño Index (ONI), which is calculated as the three-month running-mean sea surface temperature (SST) departures from average in the Niño3·4 region ([120-170°W, 5°S-5°N]). The warm phase, El Niño, occurs when the ONI >= +0·5 for a period of at least 5 consecutive three-month overlapping seasons, and the cold phase, La Niña, when ONI <= -0·5.[14] The recent El Niño in 2014/2016 was one of the strongest on record, similar in magnitude to the prominent 1997/1998 event. The warming in the Niño 3·4 region started in October 2014 and reached a maximum ONI value OF +2·3 in November-December-January (NDJ) 2015/16. The ONI then gradually decreased with a transition to a weak La Niña by autumn of 2016.[14]

Southern coastal Ecuador is an important region to study the effects of ENSO on dengue. El Niño events are associated with heavy rainfall and warmer air temperatures, with anomalies greater than 2°C, as observed during late 2015 and early 2016.[14–16] Prior studies have demonstrated the effects of ENSO and climate on dengue transmission in the region.[12,18,19]

Dengue is hyper-endemic in Ecuador and the principal cause of mosquito-borne febrile illness. There is no official dengue alert system in the country. Each year, the authorities tend to expect the same number of cases as in previous years. Each local health district monitors the behaviour of dengue based on the endemic curve, which is calculated with retrospective dengue case reports from the last five years. The mean number of weekly cases and the upper 95% confidence intervals are calculated and compared with cases reported in the current year. Current surveillance efforts do not formally incorporate climate information, although the public health sector has identified this as a priority area.

Chikungunya virus and Zika virus now co-circulate in the same region. The first recognised autochthonous cases of chikungunya were reported in Ecuador at the end of 2014.[20] The first cases of Zika were confirmed in Ecuador on January 7, 2016, and currently (23 March 2017) 2,804 suspected and 1,058 confirmed cases of Zika have been reported.[21]

In this study, we predicted monthly dengue incidence in the city of Machala, El Oro Province, Ecuador, from January to October of 2016. We incorporated seasonal climate forecasts of precipitation and temperature and a novel ONI forecast in a statistical model framework to make monthly probabilistic predictions of dengue. The forecasts were generated on 1 January 2016, to predict dengue from one to ten months ahead (January – October 2016). We accounted for misreporting due to the introduction of chikungunya virus, by using active surveillance data to correct reported dengue case data. We then evaluated the forecast retrospectively with available epidemiological information.

Methods
Study area
El Oro Province, located in southern coastal Ecuador, is a key dengue surveillance site,[22,23] due the high burden of dengue, seasonal transmission and co-circulation of all four dengue serotypes (DENV 1-4). Further, the region experiences exceptionally high aedes aegypti vector indices, which has implications for the recent emergence of chikungunya in 2015 and Zika in 2016.[12,24] There is significant movement of people and goods through the region, due to proximity to the Ecuador-Peru border, the presence of a major port, and location along the Pan America highway, likely resulting in frequent re-introductions of viruses and the vector.

Data
Passive surveillance data: Monthly clinically suspected cases of dengue from Machala from 2002 to 2016 were provided by the national surveillance system operated by the Ministry of Health. Dengue is a mandatory notifiable disease. Cases were converted to incidence using population data provided by the 2001 and 2010 national censuses (INEC 2001, INEC 2010), and population projections generated by INEC from 2011-2016 (http://www.ecuadorencifras.gob.ec/proyecciones-poblacionales/). Population estimates between 2001 and 2010 were generated by linear extrapolation.

Active surveillance data: The proportion of reported clinically diagnosed dengue cases in 2015, which were later confirmed to be chikungunya infections, were removed from the passive surveillance dengue case dataset (see Table S1). This proportion was determined from the results of a passive and active surveillance study of dengue infections in Machala, which has been described in detail elsewhere.[22] Briefly, patients were referred to the study if they were clinically diagnosed with dengue fever by physicians from sentinel clinics and the central hospital operated by the Ministry of Health in Machala. These individuals were registered as dengue cases in the Ministry of Health passive surveillance system. Serum samples from patients were tested by the study team for acute or recent dengue infections by NS1 rapid test, NS1 ELISA, IgM ELISA, RT-PCR, and for acute chikungunya infections by RT-PCR. Based on these results, we calculated the monthly proportion of clinically diagnosed dengue cases that were dengue negative and chikungunya positive, and used this proportion to adjust the total number of Ministry of Health dengue cases reported in the passive surveillance system from the same period. This allowed us to account for over reporting of dengue cases due to the recent introduction of a new febrile illness with similar clinical presentation as dengue.

Climate data: Local daily weather data (e.g. rainfall, minimum temperature) was obtained from the Granja Santa Ines weather station located in Machala (3°17'26" S, 79°54'5" W, 10 m.a.s.l.) and operated by the National Institute of Meteorology and Hydrology (INAMHI) of Ecuador. The Oceanic Niño Index (ONI) was obtained from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center of NOAA/National Weather Service. The ONI is defined as the 3-month running mean of ERSST.v4 SST anomalies in the Niño 3·4 region, based on centred 30-year base periods updated every five years.

Figure 1 shows the annual cycle of dengue incidence anomalies in Machala (per 100,000 inhabitants), at the monthly time scale from 2002-2015, precipitation and minimum temperature anomalies from the Granja Santa Ines weather station, and the ONI. El Niño events (anomalous warming of SST in Niño 3·4 region) are, in general, associated with positive temperature and precipitation anomalies in Machala, which in turn create ideal environmental conditions for dengue outbreaks (e.g. large dengue outbreak in 2010).

The epidemiological surveillance and climate data, described above, was used to formulate the dengue forecast model, described in the following section.

Dengue forecast model
A statistical mixed model was used to produce probabilistic forecasts of dengue cases per month.[18,25] Dengue cases, y_t, were assumed to follow a negative binomial distribution with mean μ_t and overdispersion parameter κ. At the linear predictor scale, the log of the mean is equal to the log population (included in the model as an offset) and log relative risk r_t for each time t.

$y_t \sim \text{NegBin}(\mu_t, \kappa)$
$\log(\mu_t) = \log(p) + \log(r_t)$
$\log(r_t) = \alpha + f(\beta_{t'(t)}) + \Sigma\gamma_j x_{jt} + \delta_{T'(t)}.$

Using exploratory analysis and model selection criteria, such as the deviance information criterion, the best estimate of the log relative risk comprised a smooth function for the annual cycle $\beta_{t'(t)}$, t'(t)=1,...,12, using a first order random walk model. The variables x_{jt}, represent the selected climate covariates: precipitation (x_{1t}) and minimum temperature (x_{2t}), lagged by one month with respect to dengue, and ONI (x_{3t}), lagged by three months with respect to dengue. Exchangeable non-structured random effects for each year $\delta_{T'(t)}$, T'(t)=1,...,14, were included, to account for interannual changes in dengue risk attributable to unknown factors from 2002–2015, such as changes in vector control practices and the introduction of new serotypes and viruses (such as the introduction of chikungunya in 2015). Given the introduction of another new virus (Zika) in 2016, the random year effect in 2015 was used to approximate the effect of a new virus (and associated misreporting) in 2016.

The model was trained using monthly dengue data from January 2002 – December 2015 and observed climate variables (precipitation, minimum temperature and ONI). The model was then used to produce forecasts for January – October 2016, making use of seasonal climate forecasts of precipitation and minimum temperature, and ONI forecasts from a new ENSO forecasting system (see below for details).

Model parameters were estimated in a Bayesian framework using Integrated Nested Laplace Approximation (INLA, www.r-inla.org), to generate 1000 samples from an approximated posterior of a fitted model.[26] Subsequently, the posterior predictive distribution of dengue cases, $\mathbf{y}_t$, for each month was estimated by drawing random values from a negative binomial distribution with mean corresponding to the elements of $\boldsymbol{\mu}_t$ and scale parameter corresponding to the elements of $\boldsymbol{\kappa}$, estimated from the model. This created a 12 x 1000 matrix, providing a probability distribution for each month. This data was then used to provide probabilistic forecasts of exceeding (a) mean incidence and (b) the upper 95% confidence interval for the mean incidence over the previous five years (2011-2015).

Seasonal climate forecasts
Seasonal forecasts of climate variables, such as precipitation and temperature, take advantage of the parts of the climate system with long-term memory, such as the oceans, to predict climate anomalies one or more months ahead of a given season.[27] To estimate uncertainty, each forecast consists of an ensemble of forecasts, obtained by perturbing the initial conditions. In this study, seasonal forecasts from the Climate Forecast System (CFS) model, developed by the National Center for Environmental Research (NCEP), were used.[28] The data was and accessed via the International Research Institute for Climate and Society (IRI) Data Library (http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.EMC/.CFSv2/index.html). The forecasts (1° zonal resolution) were arranged as a 24-member ensemble, initiated on the 1st of January 2016. The data consisted of monthly averages of precipitation and daily mean and minimum temperatures, for the 10 months following the forecast start date (January-October 2016), taken at the gridpoint nearest to the reference Granja Santa Ines weather station, located in Machala. The forecasts were then bias-corrected by subtracting the mean bias for each forecasts time, to account for the model drift.[29,30] This was done using hindcasts (e.g. retrospective forecasts) for the period 1982-2015, and corresponding observed data from the weather station.

ENSO forecast model
A structural time series model, which uses subsurface ocean temperature, wind stress, and sea surface temperature as predictor variables, was used to forecast the Niño3·4 index and calculate predicted ONI values for 2016.[31] Monthly values of the Niño3·4 were forecast at lead times up to 13 months and three-month averages were calculated. The ENSO prediction model is run with different predictor variables at different lead times. All forecasts of the 2016 ONI values were calculated using the observed Niño3·4 index data for October, November, and December 2015. For example, the November-December-January (NDJ 2015-2016) ONI was obtained by averaging the one-month-ahead forecast values of the Niño3·4 index for these months. Similarly, the December-January-February (DJF 2015-2016) ONI was obtained by averaging the two-month-ahead forecasts of Niño3·4 for these months. Thus, the last forecast used in this study for NDJ 2016-2017 ONI was a 13-month lead forecast.

Role of the funding sources
The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results
A set of analyses were performed and visual aids were co-designed with climate and public health specialists in Ecuador, to communicate the predicted climate and dengue situation for Machala, at the beginning of 2016. This consisted of (a) the climate forecasts for 2016, along with hindcasts (retrospective forecasts) to indicate how well climate models had forecast climate variations over the past 30 years; (b) the dengue prediction for the 2016 season, along with a comparison of observed and predicted dengue over the past 14 years; (c) a detailed forecast of dengue for each month in 2016, from January to October, showing the 2016 model prediction, and the mean and 95% upper confidence interval dengue incidence thresholds, based on observed dengue incidence over the previous five years.

The bias-corrected mean monthly climate forecasts of precipitation and minimum temperature (and 95% confidence intervals, based on the 24-member ensemble) for Machala, Ecuador, from January to October 1986-2016, are shown in Figure 2. Observed values from the Granja Santa Ines weather station are also shown. The forecasts were produced on 1 January each year to predict climate conditions from one to ten months ahead. Observed and forecast SST anomalies in the Niño 3·4 region, from 1986-2016, produced using a structural time series model with a 6 month lead time are also included.[31] In 2016, anomalous warming of Pacific SSTs was predicted, along with above average minimum temperature and below average precipitation in Machala. Results show that the ONI and climate forecasts models were generally successful in predicting past events. For example, the El Niño events in 1998 and 2010 were detected by the ENSO forecasting model. The CFS model correctly predicted above average minimum temperature after these peaks in SSTs.

Observed, posterior predicted mean and 95% prediction (credible) intervals for dengue incidence rates (per 100,000 population) in Machala, for the period 2002-2016 are shown in Figure 3. The model predicted, with some success, the interannual variability in dengue incidence. For example, the model accurately predicted the epidemic that occurred in 2010 and low dengue incidence rates in 2013. However, the model did underestimate incidence rates in 2003 and 2015, although observed incidence fell within the 95% prediction interval. Figure 4a shows the out-of-sample posterior predicted mean and 95% prediction (credible) interval for log dengue incidence rates (per 100,000 population) for 2016, January – October. The five-year mean dengue incidence (blue curve) and upper 95% confidence interval (red curve), calculated for the period 2011-2015, are also shown. This illustrates the typical thresholds used by the Ministry of Health in Ecuador to assess the severity of a dengue season. In Table 1, the probability of exceeding the mean and upper 95% confidence interval, calculated using incidence over the preceding five years (2011-2015) is provided. The model predicted an early peak in dengue incidence in March 2016 (compared to the previous five years), with a 91% chance of exceeding the mean dengue incidence and an 84% chance of exceeding the upper 95% confidence interval threshold (calculated for the previous five years). From June to October the posterior mean prediction was less than the five-year mean incidence, with probabilities of exceeding the mean set to below 30%.

The observed dengue incidence, obtained after the forecast had been made, was included in Figure 4b. Although the posterior predicted mean slightly exceeded the observed dengue incidence for each month, the model correctly predicted that the peak incidence would occur earlier than expected, in March 2016. The model also correctly predicted with confidence that dengue incidence would be greater than the five-year mean incidence between January and April and less than the mean incidence from June onwards.

To illustrate the benefit of incorporating active surveillance data, to correct for dengue misreporting, the model predictions were reproduced using the uncorrected dengue data, i.e. before removing the confirmed chikungunya cases from the dataset in 2015 (see Fig S1). By using the original reported data, both the predicted dengue incidence and five-year benchmark thresholds were inflated.

Discussion

Using a probabilistic dengue prediction model, driven by climate forecasts, dengue incidence rates in Machala were correctly predicted to be greater than the mean incidence over the previous five years (2011-2015) at the start of the season (between January and April 2016). The model successfully predicted the peak to occur earlier than expected, in March, with a 91% chance of exceeding the mean dengue incidence and an 84% of exceeding the upper 95% prediction interval. From June 2016, the model also correctly predicted dengue incidence rates to be less than the mean incidence observed during the previous five years.

In Ecuador, the Ministry of Health informally monitors dengue incidence based on historical passive surveillance data over the previous five years. When case reports exceed the upper 95% confidence interval, the local public health authorities are aware that there is potential for an epidemic. By incorporating forecast climate information, the model provided a more accurate dengue outlook for the upcoming dengue season, than relying on the benchmark risk thresholds of the mean and upper 95% confidence interval over the previous five years. Based on the five-year average alone, public health officials would have expected the peak to occur later in the season.

The main advantage of this new dengue prediction framework is the use of long-lead seasonal climate and ONI forecasts, which permits a prediction to be made at the start of the year for the entire dengue season. This provides advanced warning of the timing and magnitude of peak dengue incidence, which could greatly aid the management of scarce resources throughout the year.

Dengue transmission in Ecuador is seasonal, with most cases occurring during the hot and rainy season, and sporadic transmission during the rest of the year. Over the last five years, the peak in dengue has shifted from the first trimester to the second trimester. This study shows the predicted evolution of the epidemic curve in 2016 was only possible due to the incorporation of forecast climate information in the model. However, this may not be the case every year. Other factors intrinsic to the local population dynamics are likely to play a more dominant role for certain years. These factors are not explicitly accounted for in the model. However, we do include yearly random effects to crudely quantify variability resulting from unmeasured factors, such as variations in mosquito control measures from one year to the next. This allows us to better quantify the impact of climate variation on dengue inter-annual variability and make more realistic predictions of future risk, based on climate information.

The efficacy of a climate-based dengue early warning system depends on the availability of accurate climate information and skilful climate forecasts. Climate forecasts are found to be more accurate during El Niño and La Niña episodes and in ENSO affected regions, such as southern coastal Ecuador.[32] When these events occur, there is a clear opportunity to incorporate climate information into decision-making processes for climate-sensitive sectors. However, these forecasts can be less reliable during ENSO neutral years. The skill of climate model simulations and predictions still represents a major research area for improving the usefulness of health early warning systems to public health decision-makers, particularly in those regions and time-scales for which climate forecast skill is low or non-existent. Further work is in progress to explore different sources of predictability of local meteorological conditions in coastal Ecuador, to improve the skill of seasonal climate forecasts in this region.[33]

Despite these limitations, this work advances the state-of-the-art of climate services for the health sector in Ecuador, by transitioning from proof of concept to application. The successful implementation of climate services for health depends on availability of relevant, high-quality climate data, as well as the institutional and human capacity to transform the data into reliable and tailored climate products and services.[34] In our case, this relied on close collaboration between public health specialists, climate scientists, and mathematical modellers to find a compromise between the quality and resolution of the climate and epidemiological datasets.

As well as taking advantage of the lead-times provided by climate information, the model also considered active surveillance data in the city to correct the dengue dataset, given the introduction of chikungunya virus in the region in 2015. This was a unique opportunity, as active surveillance data is not readily available in

many dengue-endemic regions. By removing the estimated number chikungunya cases from the dataset, which had been erroneously recorded as dengue cases, both the model prediction and the benchmark mean estimates were more realistic (compare Fig. 4B and Fig. S1). The active surveillance study showed that dengue case data in 2015 was in fact made up of dengue, chikungunya and other febrile diseases. The serological survey also tested for Zika virus. However, no cases of Zika were detected in 2015, consistent with Ministry of Health reports. We decided to remove chikungunya cases from the dataset, rather than using confirmed dengue cases only. This is because chikungunya was first introduced to the region in 2015. In previous years, other febrile diseases were likely to have been misreported as dengue cases. Therefore, to be consistent with misreporting practices in years prior to 2015, we did not correct for diseases other than chikungunya. In 2016, Zika virus also began to circulate in Machala but only nine cases of Zika were reported in the city. We assumed that improved knowledge of the clinical differences between dengue and chikungunya among the medical community would have reduced misreporting in 2016. There are ongoing analyses of active surveillance data from 2016 to understand the prevalence and co-infections of Zika, chikungunya, dengue, and other febrile illnesses.

Ultimately, future predictions of dengue outbreaks in areas co-endemic for dengue, chikungunya, and Zika, require laboratory confirmation of cases for accurate differential diagnosis. This study highlights the need to combine climate information and active surveillance data to strengthen early warning systems for arboviruses in Ecuador and other El Niño-sensitive areas, experiencing co-circulation of arboviral diseases.

Panel: Research in context

Evidence before this study
We searched PubMed on March 7 2017, using the terms "climate", "dengue", "model", "early warning system". Several studies have used climate data to formulate dengue models and produce dengue predictions. However, predictions are usually made in retrospective mode, using observed climate data, which would have been available only after the event being predicted had occurred.

Added value of this study
To our knowledge, this work constitutes the first demonstration of the use of long-lead seasonal climate and El Niño forecasts in a dengue early warning model for Ecuador. This study adds value to the body of literature on dengue modelling by 1) using real-time climate forecasts to make long-lead dengue predictions and 2) using active surveillance data to correct for misreporting.

Implications of all the available evidence
The results of this study contribute to an on-going collaboration between the National Institute of Meteorology and Hydrology and the Ministry of Health in Ecuador, to conduct studies on climate and dengue. Prior studies have focused on providing the evidence base of the impact of climate on dengue transmission and improvements in seasonal forecasts in the region. This study takes this collaboration one step further, by co-developing a dengue early warning system using forecast climate information, which could potentially be operationalised as a climate service for the public health sector.

Keywords
dengue; model; climate service; El-Niño Southern Oscillation (ENSO); seasonal climate forecast; probabilistic; early warning system.

Contributors
RL was responsible for the study design, model development, data analysis and wrote the manuscript; AMSI collated the data, contributed to the study design and helped write the manuscript; DP provided the ENSO forecasts; MGD provided the bias-corrected climate forecasts. All authors contributed to the study design, discussed the results, reviewed and approved the final manuscript.

Declaration of interests
The authors declare no conflicts of interest.

References

1   Stanaway JD, Shepard DS, Undurraga EA, et al. The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *Lancet Infect Dis* 2016; **16:** 712–23.

2   Kuhn K, Campbell-Lendrum D, Haines A, Cox J. Using climate to predict infectious disease epidemics. Geneva: World Health Organ 2005; http://www.who.int/globalchange/publications/infectdiseases/en/index.html (accessed 29 Mar 2017).

3   Climate Services for Health - Case Studies. World Meteorol Organ 2016; https://public.wmo.int/en/resources/library/climate-services-health-case-studies (accessed 29 Mar 2017).

4   Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A. El Niño and health. *Lancet* 2003; **362:** 1481–89.

5   Lowe R, Barcellos C, Coelho CA, et al. Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts. *Lancet Infect Dis* 2014; **14:** 619–26.

6   Lowe R, Coelho CA, Barcellos C, et al. Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil. *Elife* 2016; **5:** e11285.

7   Yasuno M, Tonn RJ. A study of biting habits of Aedes aegypti in Bangkok, Thailand. *Bull World Health Organ* 1970; **43:** 319–25.

8   Couret J, Benedict MQ. A meta-analysis of the factors influencing development rate variation in Aedes aegypti (Diptera: Culicidae). *BMC Ecol* 2014; **14:** 3.

9   Watts DM, Burke DS, Harrison BA, Whitmire RE, Nisalak A. Effect of Temperature on the Vector Efficiency of Aedes aegypti for Dengue 2 Virus. *Am J Trop Med Hyg* 1986; **36:** 143–52.

10   Carrington LB, Armijos MV, Lambrechts L, Scott TW. Fluctuations at a Low Mean Temperature Accelerate Dengue Virus Transmission by Aedes aegypti. *PLoS Negl Trop Dis* 2013; **7:** e2190.

11   Pontes RJ, Freeman J, Oliveira-Lima JW, Hodgson JC, Spielman A. Vector densities that potentiate dengue outbreaks in a Brazilian city. *Am J Trop Med Hyg* 2000; **62:** 378–83.

12   Stewart Ibarra AM, Ryan SJ, Beltrán E, Mejía R, Silva M, Muñoz Á. Dengue Vector Dynamics (Aedes aegypti) Influenced by Climate and Social Factors in Ecuador: Implications for Targeted Control. *PLoS ONE* 2013; **8:** e78263.

13   Sarachik ES, Cane MA. The El Nino-southern oscillation phenomenon. Cambridge University Press, Cambridge, 2010.

14   Climate Prediction Center. Cold and Warm Episodes by Season (1951-present). http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml (assecced 29 Mar 2017).

15   Recalde-Coronel GC, Barnston AG, Muñoz ÁG. Predictability of December-April Rainfall in Coastal and Andean Ecuador. *J Appl Meteorol Climatol* 2014; **53:** 1471–93.

16   Muñoz ÁG, López P, Velásquez R, et al. An Environmental Watch System for the Andean Countries: El Observatorio Andino. *Bull Am Meteorol Soc* 2010; **91:** 1645–52.

17   Rossel F, Cadier E. El Niño and prediction of anomalous monthly rainfalls in Ecuador. *Hydrol Process* 2009; **23:** 3253–60.

18   Stewart Ibarra AM, Lowe R. Climate and non-climate drivers of dengue epidemics in southern coastal Ecuador. *Am J Trop Med Hyg* 2013; **88:** 971–81.

19   Stewart Ibarra AM, Munoz AG, Ryan SJ, et al. Spatiotemporal clustering, climate

periodicity, and social-ecological risk factors for dengue during an outbreak in Machala, Ecuador, in 2010. *BMC Infect Dis* 2014; **14:** 610.

20   PAHO. Number of Reported Cases of Chikungunya Fever in the Americas, by Country or Territory. http://www.paho.org/hq/index.php?option=com_topics&view=readall&cid=5927&Itemid=40931&lang=en (accessed 29 Mar 2017).

21   Mitchell C. PAHO WHO | Zika Cumulative Cases. Pan Am. Health Organ. World Health Organ. 2016; http://www.paho.org/hq/index.php?option=com_content&view=article&id=12390&Itemid=42090&lang=en (accessed 29 Mar 2017).

22   Stewart-Ibarra AM, Kenneson A, King CA, et al. The high burden of dengue and chikungunya in southern coastal Ecuador: Epidemiology, clinical presentation, and phylogenetics from a prospective study in Machala in 2014 and 2015. bioRxiv 2017; 102004. http://biorxiv.org/content/early/2017/01/20/102004.

23   Borbor-Cordova M, Beltran Ayala E, Cardenas W, et al. Case study 5.C Vector-virus microclimate surveillance system for dengue control in Machala, Ecuador. In: Climate Services for Health: Improving public health decision-making in a new climate. Geneva, Switzerland: World Meteorological Association and World Health Organization, 2016. http://public.wmo.int/en/resources/library/climate-services-health-case-studies (accessed 29 Mar 2017).

24   Quintero J, Brochero H, Manrique-Saide P, et al. Ecological, biological and social dimensions of dengue vector breeding in five urban settings of Latin America: a multi-country study. *BMC Infect Dis* 2014; **14:** 38.

25   Lowe R, Bailey TC, Stephenson DB, et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Stat Med* 2013; **32:** 864–83.

26   Lowe R, Cazelles B, Paul R, Rodó X. Quantifying the added value of climate information in a spatio-temporal dengue model. *Stoch Environ Res Risk Assess* 2015; **30:** 2067–78.

27   Doblas-Reyes FJ, García-Serrano J, Lienert F, Biescas AP, Rodrigues LR. Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdiscip Rev Clim Change* 2013; **4:** 245–68.

28   Saha S, Moorthi S, Wu X, et al. The NCEP climate forecast system version 2. *J Clim* 2014; **27:** 2185–2208.

29   Wetterhall F, Pappenberger F, He Y, Freer J, Cloke HL. Conditioning model output statistics of regional climate model precipitation on circulation patterns. *Nonlinear Process Geophys* 2012; **19:** 623–33.

30   Guemas V, García-Serrano J, Mariotti A, Doblas-Reyes F, Caron L-P. Prospects for decadal climate prediction in the Mediterranean region. *Q J R Meteorol Soc* 2015; **141:** 580–97.

31   Petrova D, Koopman SJ, Ballester J, Rodó X. Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model. *Clim Dyn* 2017; **48:** 1249.

32   Ballester J, Lowe R, Diggle PJ, Rodó X. Seasonal forecasting and health impact models: challenges and opportunities. *Ann N Y Acad Sci* 2016; **1382:** 8–20.

33   de Guenni LB, García M, Muñoz ÁG, et al. Predicting monthly precipitation along coastal Ecuador: ENSO and transfer function models. *Theor Appl Climatol* 2016; 1–15. doi:10.1007/s00704-016-1828-4.

34   Watts N, Adger WN, Ayeb-Karlsson S, et al. The Lancet Countdown: tracking progress on health and climate change. *Lancet* 2016: **389:** 1151–64.

Table 1: Monthly probabilistic dengue risk forecasts for Machala, Ecuador, January – October 2016. Mean and upper 95% confidence interval (CI) for the mean cases and incidence (cases per 100,000 population) for the last five-year period 2011-2015. Probability of dengue incidence rates exceeding both the five year mean and upper 95% CI are shown.

Month	Mean (2011-2015)		Probability of exceeding mean	Upper 95% CI (2011-2015)		Probability of exceeding upper 95% CI
	Cases	Incidence		Cases	Incidence	
Jan	18	8	89%	27	11	84%
Feb	33	13	85%	58	23	74%
Mar	37	15	91%	56	22	84%
Apr	59	23	79%	68	26	75%
May	133	51	57%	186	71	43%
Jun	248	95	24%	430	162	9%
Jul	154	59	17%	251	97	5%
Aug	79	31	22%	140	54	6%
Sep	52	21	19%	69	27	11%
Oct	33	13	28%	53	22	11%

Figure 1: Annual cycle of (a) dengue incidence anomalies in Machala, Ecuador (per 100,000 inhabitants) (b) precipitation anomalies (mm/day) and (c) minimum temperature anomalies (deg C), from the Granja Santa Ines weather station, located in Machala and (d) Oceanic Niño Index (ONI, sea surface temperatures (SST) anomalies in the Niño 3·4 region) at the monthly time scale from 2002-2015.

Figure 2: Bias-corrected monthly forecasts from the Climate Forecast System version 2 (CFSv2) model for Machala, Ecuador, from January to October 1986-2016 for (a) precipitation (blue curve) and (b) minimum temperature (red curve). The shaded areas represent the 95% confidence intervals for the ensemble forecast (24 members). Observations from the Granja Santa Ines weather station, located in Machala, (black curve) for 1986-2015 are included. Forecasts are produced in January each year to predict climate conditions up to ten months in advance. (c) Forecast (purple curve) sea surface temperature (SST) anomalies in the Niño 3·4 region, 1986-2016. Forecasts are produced using a structural time series model with a six-month lead time. Observed values (black curve) for 1986-2015 are included.

Figure 3: Observed (solid black curve), posterior predicted mean (dashed purple curve) and 95% prediction (credible) interval (purple shaded area) for dengue incidence rates (per 100,000 population) in Machala, Ecuador, 2002-2016.

Figure 4: (a) Posterior predicted mean (dashed purple curve) and 95% prediction (credible) interval (purple shaded area) for log dengue incidence rates (per 100,000 population) in Machala, Ecuador, January – October 2016. The five year mean dengue incidence (blue curve) and upper 95% confidence interval (red curve), for the period 2011-2015, is shown. (b) as (a) with observed incidence shown in black.

Figure 1

Figure 2

Figure 3

Figure 4a
Click here to download high resolution image

Figure 4b
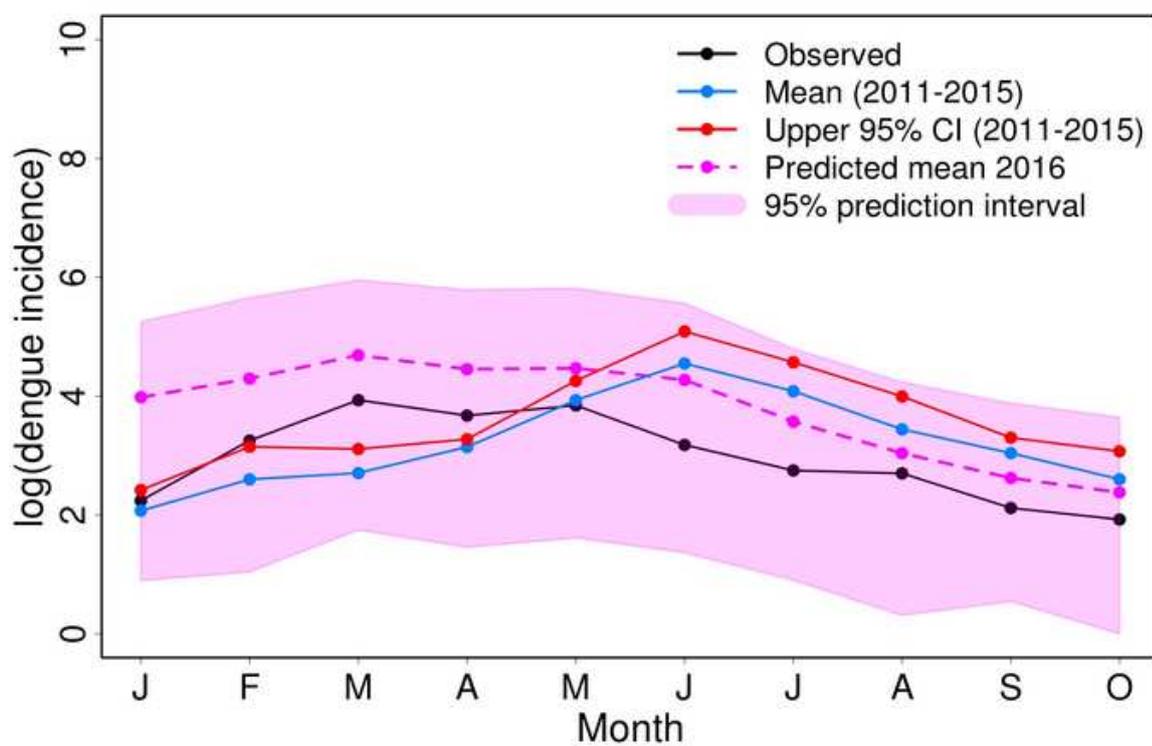
4.3 RossBell Dipole and PSA Mode Dynamics: Combination Forcing by ENSO and the Annual Cycle

Petrova D (1,2), Stuecker M (3,4), Timmermann A (3), Ballester J (1,2), Rodó X (1,2,5)

(1) Institut Català de Ciències del Clima, Barcelona, Catalonia, Spain

(2) Barcelona Institute for Global Health, Barcelona, Catalonia, Spain

(3) School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, USA

(4) Department of Atmospheric Sciences, University of Washington, Seattle, USA

(5) Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

In progress

RossBell Dipole and PSA Mode Dynamics: Combination Forcing by ENSO and the Annual Cycle

D. Petrova, M. Stuecker, A. Timmermann, J. Ballester, X. Rodó.

April 19, 2017

1   Data and Experiments

We use the general circulation model (GCM) experiments described in (Stuecker et al., 2015). In these experiments idealized ENSO-like sea surface temperature (SST) anomaly forcing is prescribed with the aim to explore the nonlinear interaction between ENSO and the annual cycle, and the role of atmosphere-ocean coupling both in the tropics and in the RossBell region of interest in the Southern Ocean. For the experiments an Atmospheric General Circulation Model (AGCM) and and AGCM partially coupled (PARCP) to a thermodynamic slab ocean model (SOM) have been used.

The spatial pattern for the experiments is obtained by regressing the HadISST1 (Rayner et al., 2003) anomalies onto the normalized N3.4 index (Figure 2 in Stuecker et al. (2015)). The forcing region applied for the experiments is bound in the box [170°E-280°E]x[25°S-25°N]. The ENSO pattern is multiplied by a sinusoidal time series in order to obtain the ENSO-related sea surface temperature anomalies (SSTA). In this way, only SSTA in the eastern tropical Pacific (positive or negative) characterize an El Niño (EN) or La Niña (LN) event. The model used in the experiments is the atmospheric components CAM4 of the Community Earth System Model (CESM) version 1.0.3 (Neale et al., 2013) in a T42 horizontal resolution and 26 vertical levels.

1.1   Atmospheric General Circulation Model experiment

The discussed SSTA are applied over an SST and sea ice climatology based on the observations in the period 1982-2001 (Hurrel et al., 2008). A 100-year AGCM run (i.e. 40 idealized ENSO cycles) characterized by a sinusoidal 2.5-year period ENSO SSTA forcing and modern day SST annual cycle has been used. The forcing phase is such that the peak EN phase occurs once in December and once in June during a 5-year cycle (the peak of LN is in March and September). Then the long-term monthly means are removed in order to obtain anomalies. In this way, it is possible to study the sensitivity of the anomalous circulation response related to the ENSO phase and the phase of the annual cycle. The 2.5-year ENSO period is chosen so that the real evolution of an El Niño event, SSTAs peaking in December (forcing frequency 0.4 yr^{-1}) and decaying by the following boreal summer, could be reproduced, and at the same time to be possible to clearly identify the combination tone frequency peaks (0.6 yr^{-1} and 1.4 yr^{-1}). It should be noted, however, that La Niña events do not peak as they do in the observations, which could mean that results about LN may be less realistic. However, some information about the processes during the cold phase should also be revealed, even thoug the phase is different.

1.2   Partially coupled experiment

SST represent a fixed boundary condition in AGCM runs and thus are not able to adjust to heat fluxes. Since observed SST act both as forcing and solution, some previous studies (Lau and Nath,

2006; Kucharski et al., 2007; Lu et al., 2011; Kosaka and Xie, 2013; Lee et al., 2013) have applied PARCP experiments in which SST (or the heat fluxes) are prescribed in parts of the model domain, and in the rest of the domain the atmospheric model is coupled with an underlying thermodynamic and/or dynamic ocean. Here we use the same idealized ENSO pattern as in the AGCM experiment, but SST are prescribed from the dateline to the South American coast within the tropical band [15°S-15°N]. In the rest of the domain a thermodynamical SOM is applied. There is also a buffer zone [170°E-180°E]x[15°N-25°N], where the prescribed and calculated SSTs are gradually merged in order to avoid unrealistic sharp SST gradients (Figure 2 in (Stuecker et al., 2015)).

The SOM applied here to obtain heat fluxes based on the observed climatology is that of (Kiehl et al., 2006). CAM4 is run only with the modern climatology as boundary conditions for 30 years in order to obtain the climatological net heat fluxes F_{net}. Further details about the configuration of this experiment could be found in (Stuecker et al., 2015).

Similarly to the AGCM experiment, the PARCP experiment has been designed to test the effect of atmosphere-ocean coupling on the RossBell dynamics - a 100-year run applying a 2.5-year period sinusoidal ENSO SSTA forcing and modern day SST annual cycle. Again El Niño peaks once in December and once in June during any 5-year cycle, and La Niña peaks in March and September.

2   Results

2.1   Stream function RossBell index and the theoretical C-modes

Composites of the time evolution of the N3.4 SST time series (blue line), the normalized theoretical C-mode (green line), the normalized theoretical C-mode-south (light blue line), the RossBell stream function index for the AGCM (red line) and the RossBell stream function index for the PARCP (black line) for the following four cases (20 member averages each) have been calculated (Figure 3):
a) El Niño events with a peak phase in December (Figure 3a);
b) El Niño events with a peak phase in June (Figure 3b);
c) La Niña events with a peak phase in March (Figure 3c);
d) La Niña events with a peak phase in September (Figure 3d);

The theoretical C-mode-south (N3.4 x annual cycle in the Southern Hemisphere extratropics) captures some of the time evolution of the RB circulation indices for both experiments (AGCM and PARCP) in both El Niño events with winter and with summer peaks (Figure 3a, b). The correlation between the three series is significant (Table 1). Similarly, the theoretical C-mode (N3.4 x annual cycle in the tropics) captures the time evolution of the RB circulation indices for both experiments (AGCM and PARCP) in both La Niña events with autumn and with spring peaks (Figure 3c, d). At the same time the correlations between the N3.4 SST time series and the stream function RB indices are even higher (Table 1). Bi-linear regression has been calculated for the RB circulation index from the PARCP experiment in terms of N3.4 index and the C-mode, as well as the N3.4 index and the C-mode-south index for the ENSO composites. The respective regression coefficients are 0.25 for N3.4, -0.07 for C-mode, and 0.06 for C-mode-south. The same coefficients, but for composites only of EN events are 0.16 for N3.4, -0.42 for C-mode, and 0.44 for C-mode-south; and only of LN events 0.1 for N3.4, -0.3 for C-mode and 0.24 for C-mode-south. Thus, we might conclude that the ENSO forcing is important factor for the dynamics of the RB circulation pattern, but its combination with the seasonal cycles in the south Pacific extratropics (where RB occurs) plays a more significant role during El Niño events.

For El Niño events with winter peaks (Figure 3a) the C-mode-south time series predicts an anomalous RB circulation in the developing EN phase (boreal summer and fall - July - November(0)), as well as during the decaying phase in spring and summer of the following year (April - September (1)). In

2

the following summer when negative La Niña conditions set in, however, the RB stream function index is predicted by the time series of the other C-mode (Figure 3c), indicating that the ENSO forcing and its combination with seasonality in the tropics is more influential during La Niña as compared to El Niño. The same results hold for the EN with summer peaks (Figure 3b) and LN with autumn peaks (Figure 3d). Thus, the RB dynamics is to a large extent imfluenced by ENSO regardless of the exact timing of the cycle.

2.2 SST RossBell index and the theoretical C-modes

Composites of the time evolution of the N3.4 SST time series (blue line), the normalized theoretical C-mode (green line), the normalized theoretical C-mode-south (light blue line), the RossBell SST index for the AGCM (red line) and the RossBell SST index for the PARCP (black line) for the same four cases (20 member averages each) have been calculated (Figure 4).

In the case of the SST RB indices for both EN with winter and with summer peaks, the RB SST evolution is explained by the C-mode time series (Figure 4a, b). On the other hand, in the case of LN the C-mode-south time series predicts the RB indices (Figure 4c, d), suggesting that the ENSO forcing and the local seasonal cycle dominate the dynamics during the cold ENSO phase. An interesting feature stands out in Figure 4a (EN with winter peak), there is a small RB peak in the months August-November(0). This peak is captured by the C-mode-south index, showing that there is some SST response also due to EN dynamics and local Southern Hemisphere seasonality. Moreover, this peak is entirely missing in the case of the EN with summer peak (Figure 4b), indicating that the SST response in the SH RB region indeed is affected by the EN-locking to the annual cycle. The correlation coefficients for the respective time series are given in Table 2. The correlations between the RB SST indices and the N3.4 index are low (Table 2). At the same time the correlations between the RB stream function indices and the RB SST indices are significant (-0.58 in the AGCM experiment and -0.50 in the PARCP experiment), suggesting that the RB SST response is not directly affected by the ENSO forcing, but is rather modulated by the local atmospheric response to ENSO in the southern extra-tropics, and by seasonality. Again, in order to confirm this conclusion, we performed a bi-linear regression for the RB SST index from the PARCP experiment in terms of the N3.4 index and the C-mode index, the N3.4 and the C-mode-south index, and the N3.4 and the RB circulation index. The respective coefficients for the composites are: 0.06 for N3.4, 0.006 for C-mode, and -0.01 for C-mode-south; as well as 0.2 for N3.4, and -0.5 for RB stream function index; for composites only of EN events: -0.1 for N3.4, 0.32 for C-mode, -0.33 for C-mode-south, as well as -0.01 for N3.4 and -0.32 for RB stream function index; for composites only of LN events: 0.1 for N3.4, -0.3 for C-mode, 0.24 for C-mode-south, as well as 0.1 for N3.4 and -0.05 for the RB stream function index.

2.3 Stream function PSA1 index and the theoretical C-modes

Composites of the time evolution of the N3.4 SST time series (blue line), the normalized theoretical C-mode-south (light blue line), the PSA1 stream function index for the AGCM (yellow line) and the PSA1 stream function index for the PARCP (light green line) for the same four cases (20 member averages each) have been calculated (Figure 5). As can be seen in Figure 5 the theoretical C-mode-south generally predicts the PSA1 index in both the AGCM and the PARCP experiment for both EN an LN events and in all 4 cases (i.e. regardless of the time when ENSO peaks). The time series in the two experiments have a high correlation coefficient of 0.64, which rises to 0.88 for the composites (Table 3), indicating that coupling plays some role in the PSA1 atmospheric index, but it is not of fundamental importance. Overall, there is a very high correlation between the N3.4 index and the PSA1 index, and a smaller correlation with the theoretical C-mode-south, thus indicating that the combination between ENSO and the annual cycle in the tropics has some impact on the atmospheric response in this part of the southern extra-tropics, but not as significant when compared to the effect

of ENSO alone. There is also a high correlation between the stream function PSA 1 index and the RossBell stream function index, especially for the composites in the coupled experiment (0.62). The respective bi-linear regression coefficients are -0.02 for N3.4 index, 0.06 for C-mode index, and -0.06 for C-mode-south index.

2.4   Temperature PSA1 index and the theoretical C-modes

Composites of the time evolution of the N3.4 SST time series (blue line), the normalized theoretical C-mode (green line), the normalized theoretical C-mode-south (light blue line), the PSA1 SST index for the AGCM (velvet line) and the PSA1 SST index for the PARCP (magenta line) for the same four cases (20 member averages each) have been calculated (Figure 6). The SST PSA1 index is only slightly correlated with the PSA1 stream function index - a correlation of 0.21 (Table 4). This indicates that the oceanic response in this part of the southern extra-tropics is not simply driven by the local atmospheric response to ENSO. On the other hand, there is a very high lagged correlation (lag -7) of 0.88 between the SST PSA1 index and the N3.4 index in the PCMKR experiment (see Table 4 and Figure 6). Based on this, we can conclude that the PSA1 SST response is rather directly affected by the equatorial ENSO forcing, in fact it leads ENSO by 6-7 months, and combination mode dynamics seems to play little effect (Table 4, Figure 6). The bi-linear regression coefficients are 0.03 for N3.4, -0.03 for C-mode, and 0.02 for C-mode-south.

Table 1: Cross-correlation coefficients R between the RB stream function indices and the C-mode and N3.4 indices for the 2.5 yr Sine ENSO experiments (n = 1200 months). The R values for the composites are given in brackets.

Indices	R
N3.4 and C-mode	0.00 (0.00)
N3.4 and C-mode-south	0.00 (0.00)
N3.4 and RossBell ψ AGCM	0.45 (0.62)
N3.4 and RossBell ψ PARCP	0.45 (0.60)
C-mode and RossBell ψ AGCM	0.00 (-0.06)
C-mode and RossBell ψ AGCM (LN EVENTS)	(0.30)
C-mode and RossBell ψ PARCP	-0.10 (-0.09)
C-mode and RossBell ψ PARCP (LN EVENTS)	(0.49)
C-mode-south and RossBell ψ AGCM	0.00 (0.04)
C-mode-south and RossBell ψ AGCM (EN EVENTS)	(0.33)
C-mode-south and RossBell ψ PARCP	0.10 (0.08)
C-mode-south and RossBell ψ PARCP (EN EVENTS)	(0.50)
RossBell ψ AGCM and RossBell ψ PARCP	0.42 (0.90)

Table 2: Cross-correlation coefficients R between the RB SST indices and the C-mode and N3.4 indices for the 2.5 yr Sine ENSO experiments (n = 1200 months). The R values for the composites are given in brackets.

Indices	R
N3.4 and C-mode	0.00 (0.00)
N3.4 and C-mode-south	0.00 (0.00)
N3.4 and RossBell SST AGCM	0.02 (0.22)
N3.4 and RossBell SST PARCP	0.00 (0.22)
C-mode and RossBell SST AGCM	0.00 (0.00)
C-mode and RossBell SST AGCM (EN EVENTS)	(0.48)
C-mode and RossBell SST PARCP	0.00 (0.01)
C-mode and RossBell SST PARCP (EN EVENTS)	(0.76)
C-mode-south and RossBell SST AGCM	0.00 (0.00)
C-mode-south and RossBell SST AGCM (LN EVENTS)	(0.57, 0.64 at Lag(1))
C-mode-south and RossBell SST PARCP	$0.00 \, (-0.03)$
C-mode-south and RossBell SST PARCP (LN EVENTS)	(0.62, 0.76 at Lag(1))
RossBell SST AGCM and RossBell ψ AGCM	$-0.39 \, (-0.58)$
RossBell SST PARCP and RossBell ψ PARCP	$-0.15 \, (-0.40, \, -0.50 \text{ at Lag(-1)})$
RossBell SST AGCM and RossBell SST PARCP	0.34 (0.79)

Table 3: Cross-correlation coefficients R between the PSA1 (PC3) stream function indices and the C-mode and N3.4 indices for the 2.5 yr Sine ENSO experiments (n = 1200 months). The R values for the composites are given in brackets.

Indices	R
N3.4 and C-mode	0.00 (0.00)
N3.4 and PSA1 (PC1) ψ AGCM	0.74 (0.86)
N3.4 and PSA1 (PC1) ψ PARCP	0.68 (0.80)
C-mode and PSA1 (PC1) ψ AGCM	-0.21 (-0.25, -0.26 at Lag(+1))
C-mode and PSA1 (PC1) ψ AGCM (EN EVENTS)	(-0.27)
C-mode and PSA1 (PC1) ψ AGCM (LN EVENTS)	(-0.41)
C-mode and PSA1 (PC1) ψ PARCP	-0.34 (-0.40)
C-mode and PSA1 (PC1) ψ PARCP (EN EVENTS)	(-0.50, -0.55 at Lag(+1)))
C-mode and PSA1 (PC1) ψ PARCP (LN EVENTS)	(-0.61)
C-mode-south and PSA1 (PC1) ψ AGCM	0.20 (0.25, 0.27 at Lag(+1))
C-mode-south and PSA1 (PC1) ψ AGCM (EN EVENTS)	(0.26), 0.33 at Lag(+1))
C-mode-south and PSA1 (PC1) ψ AGCM (LN EVENTS)	(0.39)
C-mode-south and PSA1 (PC1) ψ PARCP	0.33 (0.39)
C-mode-south and PSA1 (PC1) ψ PARCP (EN EVENTS)	(0.48, 0.56 at Lag(+1))
C-mode-south and PSA1 (PC1) ψ PARCP (LN EVENTS)	(0.58)
PSA1 (PC1) ψ and RossBell ψ AGCM	0.36 (0.57)
PSA1 (PC1) ψ and RossBell ψ PARCP	0.40 (0.57, 0.62 at Lag(+1)))
PSA1 (PC1) ψ AGCM and PSA1 (PC1) ψ PARCP	0.64 (0.88, 0.90 at Lag(-1))

7

Table 4: Cross-correlation coefficients R between the PSA1 SST indices and the C-mode and N3.4 indices for the 2.5 yr Sine ENSO experiments (n = 1200 months). The R values for the composites are given in brackets.

Indices	R
N3.4 and C-mode	0.0 (0.0)
N3.4 and C-mode-south	0.0 (0.0)
N3.4 and PSA1 SST AGCM	0.07 (0.09)
N3.4 and PSA1 SST PARCP	0.02, 0.40 at Lag(+/-7)
	$(0.10, -0.83$ at Lag(-7), 0.67 at Lag(6))
C-mode and PSA1 SST AGCM	-0.05 (-0.08)
C-mode and PSA1 SST AGCM (EN EVENTS)	$(0.45, -0.59$ at Lag(-6))
C-mode and PSA1 SST PARCP	-0.04 (-0.06)
C-mode and PSA1 SST PARCP (EN EVENTS)	(0.29, 0.45 at Lag(2))
C-mode-south and PSA1 SST AGCM	0.04 (0.07)
C-mode-south and PSA1 SST AGCM (LN EVENTS)	(0.66)
C-mode-south and PSA1 SST PARCP	0.03 (0.03)
C-mode-south and PSA1 SST PARCP (LN EVENTS)	(0.40, 0.50 at Lag(1))
PSA1 SST AGCM and PSA1 (PC3) ψ AGCM	-0.12 (-0.07)
PSA1 SST PARCP and PSA1 (PC3) ψ PARCP	-0.06 (-0.09)
PSA1 SST AGCM and PSA1 SST PARCP	0.09 (0.23)
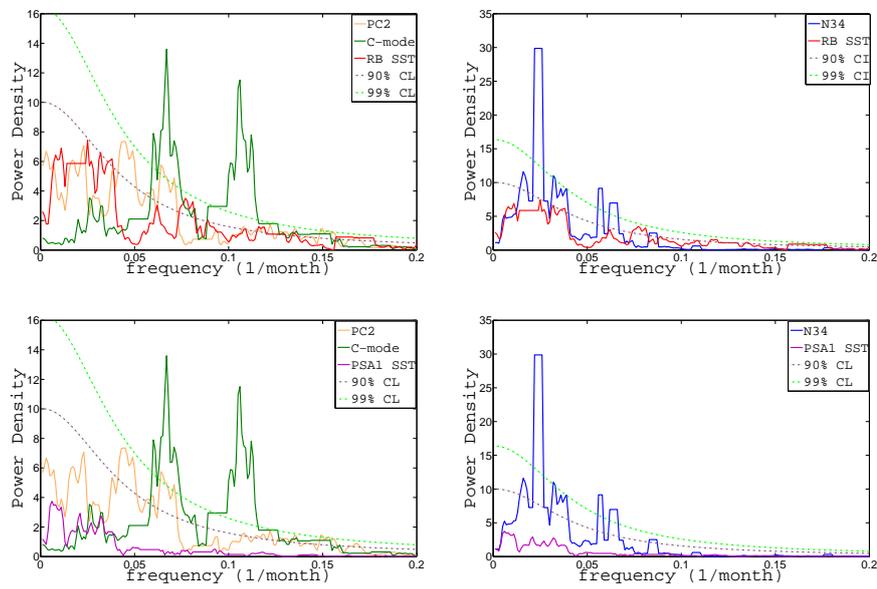
Figure 1: Multi Taper Method (MTM) power spectra for the indicated observed indices. The solid line indicates the power density and dashed lines the respective confidence level (CL) based on a red noise null hypothesis.
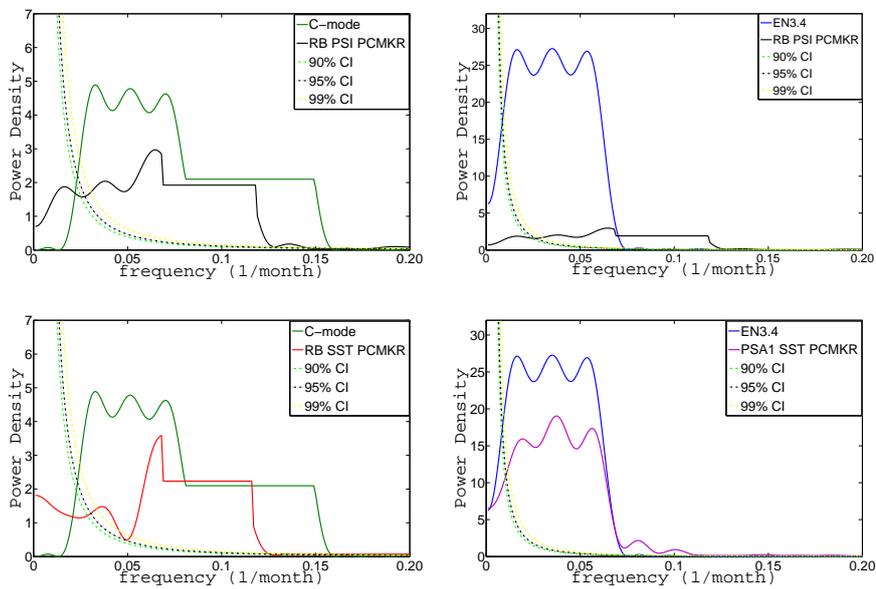
Figure 2: Multi Taper Method (MTM) power spectra for the indicated PARCP experiment indices. The solid line indicates the power density and dashed lines the respective confidence level (CL) based on a red noise null hypothesis.

2.5 Sine ENSO experiment composites: El Niño Phase

a) El Niño with winter peak

b) El Niño with summer peak

Figure 3: Index time evolution composites (20 members each) for the 2.5 yr Sine ENSO experiments. Shown are the N3.4 SST forcing index (blue, [° C]), the normalized theoretical C-mode-south (light blue), the RB ψ index for the AGCM experiment (red, $[10^6 \ m^2 \ s^{-1}]$) and the RB ψ index for the PARCP experiment (black, $[10^6 \ m^2$ $s^{-1}]$). Due to the sinusoidal 2.5 yr ENSO period, the events can be separated and composited into the following groups: a) Composite for the El Niño events with a peak in December; b) Composite for the El Niño events with a peak in June.

2.5 Sine ENSO experiment composites: La Niña Phase

c) La Niña with spring peak *d) La Niña with autumn peak*

Figure 3: Index time evolution composites (20 members each) for the 2.5 yr Sine ENSO experiments. Shown are the N3.4 SST forcing index (blue, [° C]), the normalized theoretical C-mode (green), the RB ψ index for the AGCM experiment (red, [10^6 m^2 s^{-1}]) and the RB ψ index for the PARCP experiment (black, [[10^6 m^2 s^{-1}]). Due to the sinusoidal 2.5 yr ENSO period, the events can be separated and composited into the following groups: a) Composite for the La Niña events with a peak in March; b) Composite for the La Niña events with a peak in September.

2.5 Sine ENSO experiment composites: El Niño Phase

a) El Niño with winter peak   *b) El Niño with summer peak*

Figure 4: Index time evolution composites (20 members each) for the 2.5 yr Sine ENSO experiments. Shown are the N3.4 SST forcing index (blue, [° C]), the normalized theoretical C-mode (green), the RB SST index for the AGCM experiment (red, [° C]) and the RB SST index for the PARCP experiment (black, [° C]). Due to the sinusoidal 2.5 yr ENSO period, the events can be separated and composited into the following groups: a) Composite for the El Niño events with a peak in December; b) Composite for the El Niño events with a peak in June.

2.5 Sine ENSO experiment composites: La Niña Phase

c) La Niña with spring peak   *d) La Niña with autumn peak*
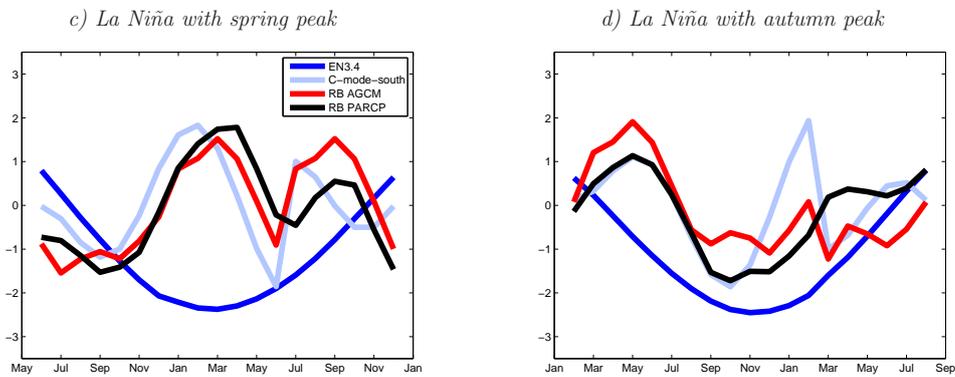
Figure 4: Index time evolution composites (20 members each) for the 2.5 yr Sine ENSO experiments. Shown are the N3.4 SST forcing index (blue, [° C]), the normalized theoretical C-mode-south (light blue), the RB SST index for the AGCM experiment (red, [° C]) and the RB SST index for the PARCP experiment (black, [° C]). Due to the sinusoidal 2.5 yr ENSO period, the events can be separated and composited into the following groups: a) Composite for the La Niña events with a peak in March; b) Composite for the La Niña events with a peak in September.

2.5 Sine ENSO experiment composites: El Niño Phase

a) El Niño with winter peak   *b) El Niño with summer peak*

Figure 5: Index time evolution composites (20 members each) for the 2.5 yr Sine ENSO experiments. Shown are the N3.4 SST forcing index (blue, [° C]), the normalized theoretical C-mode-south (light blue), the PSA1 (PC1) ψ index for the AGCM experiment (yellow, $[10^6 \ m^2 \ s^{-1}]$) and the PSA1 (PC1) ψ index for the PARCP experiment (light green, $[[10^6 \ m^2 \ s^{-1}]$). Due to the sinusoidal 2.5 yr ENSO period, the events can be separated and composited into the following groups: a) Composite for the El Niño events with a peak in December; b) Composite for the El Niño events with a peak in June.

2.5 Sine ENSO experiment composites: La Niña Phase

c) La Niña with spring peak   *d) La Niña with autumn peak*

Figure 5: Index time evolution composites (20 members each) for the 2.5 yr Sine ENSO experiments. Shown are the N3.4 SST forcing index (blue, [° C]), the normalized theoretical C-mode-south (light blue), the PSA1 (PC1) ψ index for the AGCM experiment (yellow, $[10^6 \ m^2 \ s^{-1}]$) and the PSA1 (PC1) ψ index for the PARCP experiment (light green, $[[10^6 \ m^2 \ s^{-1}]$). Due to the sinusoidal 2.5 yr ENSO period, the events can be separated and composited into the following groups: a) Composite for the La Niña events with a peak in March; b) Composite for the La Niña events with a peak in September.

2.5 Sine ENSO experiment composites: El Niño Phase

a) El Niño with winter peak   *b) El Niño with summer peak*

Figure 6: Index time evolution composites (20 members each) for the 2.5 yr Sine ENSO experiments. Shown are the N3.4 SST forcing index (blue, [° C]), the normalized theoretical C-mode-south (light blue), the normalized theoretical C-mode (green), the PSA1 SST index for the AGCM experiment (light velvet, [° C]) and the PSA1 SST index for the PARCP experiment (dark pink, [° C]). Due to the sinusoidal 2.5 yr ENSO period, the events can be separated and composited into the following groups: a) Composite for the El Niño events with a peak in December; b) Composite for the El Niño events with a peak in June.

2.5 Sine ENSO experiment composites: La Niña Phase

c) La Niña with spring peak   *d) La Niña with autumn peak*

Figure 6: Index time evolution composites (20 members each) for the 2.5 yr Sine ENSO experiments. Shown are the N3.4 SST forcing index (blue, [° C]), the normalized theoretical C-mode-south (light blue), the normalized theoretical C-mode (green), the PSA1 SST index for the AGCM experiment (light velvet, [° C]) and the PSA1 SST index for the PARCP experiment (dark pink, [° C]). Due to the sinusoidal 2.5 yr ENSO period, the events can be separated and composited into the following groups: a) Composite for the La Niña events with a peak in March; b) Composite for the La Niña events with a peak in September.

Figure 7: 2.5 yr period sinusoidal ENSO PCMKR experiment regression coefficients. Stream function anomalies regressed onto the respective stream function indices.
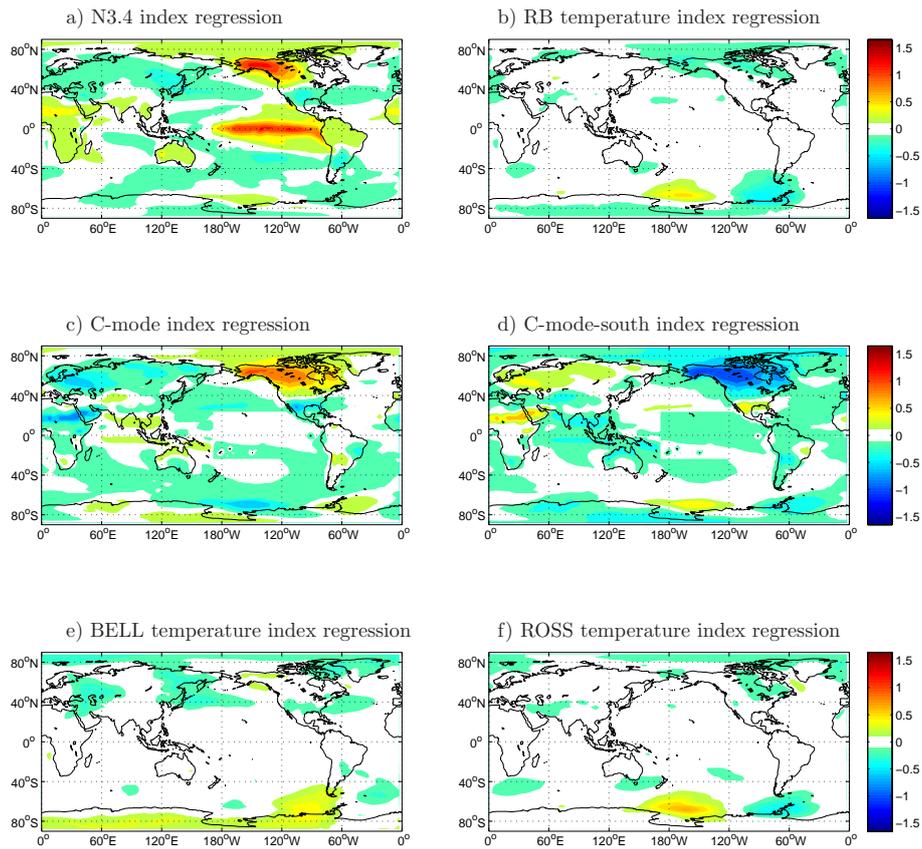
Figure 8: 2.5 yr period sinusoidal ENSO PCMKR experiment regression coefficients. SST anomalies regressed onto the respective SST indices.
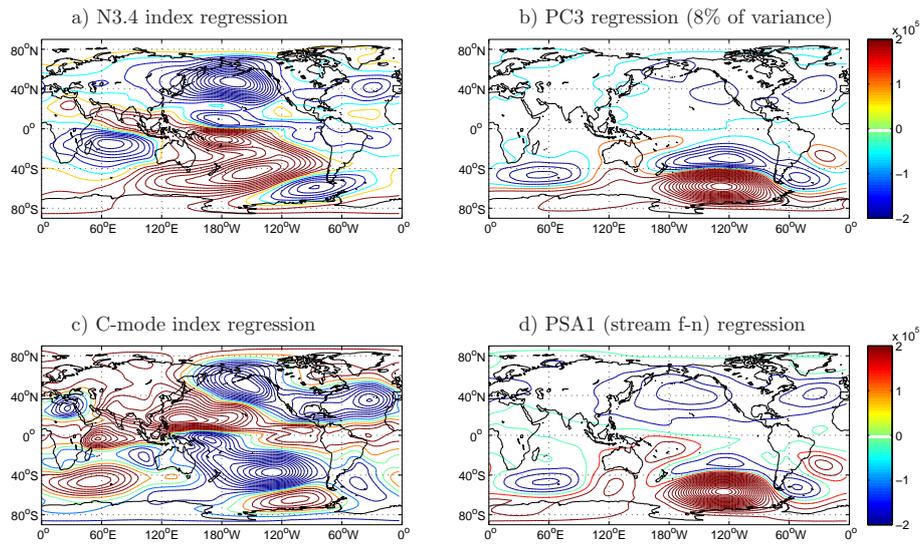
Figure 9: 2.5 yr period sinusoidal ENSO PCMKR experiment regression coefficients. Stream function anomalies regressed onto the respective stream function indices.

References

J.W. Hurrel, J.J. Hack, D. Shea, J.M. Caron, and J. Rosinski. A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *Journal of Climate*, 21:5145–5153, 2008.

J. T. Kiehl, C. A. Shields, J. J. Hack, and W. D. Collins. The climate sensitivity of the community climate system model version 3 (ccsm3). *Journal of Climate*, 19:2584–2596, 2006.

Y. Kosaka and S.P. Xie. Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, 501:403–407, 2013.

F. Kucharski, A. Bracco, J.H. Yoo, and F. Molteni. Low-frequency variability of the Indian Monsoon-ENSO relationship and the tropical Atlantic: the weakening of the 1980s and 1990s. *Journal of Climate*, 20:4255–4266, 2007.

N.C. Lau and M.J. Nath. ENSO modulation of the interannual and intraseasonal variability of the East Asian Monsoon - a model study. *Journal of Climate*, 19:4508–4530, 2006.

S.K. Lee, R. Atlas, D. Enfield, C. Wang, and H. Liu. Is there an optimal enso pattern that enhances large-scale atmospheric processes conducive to tornado outbreaks in the united states? *Journal of Climate*, 26:1626–1642, 2013.

J.M. Lu, B. Cash, and S. Li. Oceanic forcing for the East Asian precipitation in pacemaker AGCM experiments. *Geophysical Research Letters*, 38, 2011.

R.B. Neale, J. Richter, S. Park, P.H. Lauritzen, S.J. Vavrus, P.J. Rasch, and M. Zhang. The mean climate of the Community Atmosphere Model (CAM4) in forced SST and fully coupled experiments. *Journal of Climate*, 26:5150–5168, 2013.

N.A. Rayner, D.E. Parker, E. B. Horton, Folland C.K., L.V. Alexander, D.P. Rowell, E.C. Kent, and A. Kaplan. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108:4407, 2003.

M.F. Stuecker, F.F Jin, A. Timmermann, and Sh. McGregor. Combination mode dynamics of the anomalous Northwest Pacific Anticyclone. *Journal of Climate*, 28:1093–1111, 2015.

Bibliography

E. Sarachik and M. Cane. *The El Niño Southern Oscillation Phenomenon*. Cambridge University Press, 2010.

G. Walker. Correlation in seasonal variations of weather. IX. A further study of world weather. *Memoirs of the Indian Meteorological Department*, 24:275–332, 1924.

NOAA/CPC. The southern oscillaton index. *http://www.cpc.noaa.gov/products/analysis_monitoring/ensocycle/soi.shtml*, 2017.

F. F. Jin and Soon-Il An. Thermocline and zonal advective feedbacks within the equatorial ocean recharge oscillator model for ENSO. *Geophysical Research Letters*, 26:2989–2992, 1999.

F. F. Jin and J. D. Neelin. Modes of interannual tropical ocean–atmosphere interaction - A unified view. Part I: Numerical results. *Journal of the atmospheric sciences*, 50:3477–3503, 1993.

F. F. Jin. An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model. *Journal of Atmospheric Sciences*, 54:811–829, 1997a.

J. Bjerknes. Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review*, 97:163–172, 1969.

J. Bjerknes. A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus*, 18:163–172, 1966.

A. Fedorov and S. G. Philander. A stability analysis of tropical ocean-atmosphere interactions: Bridging measurements and theory for El Niño. *Journal of Climate*, 14:3086–3101, 2001.

A. Fedorov, S. Harper, S. G. Philander, B. Winter, and A. Wittenberg. How predictable is El Niño? *Bull. Am. Meteorol. Soc.*, 84:911–919, 2003.

S. Philander. El Niño and La Niña. *Journal of Atmospheric Sciences*, 77:451–459, 1989.

Hsun-Ying Kao and Jin-Yi Yu. Contrasting eastern-pacific and central-pacific types of ENSO. *Journal of Climate*, 22:615–632, 2009.

E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77:437–471, 1996.

M.J. Suarez and P.S. Schopf. A delayed action oscillator for ENSO. *Journal of the Atmospheric Sciences*, 45:3283–3287, 1988.

A. Fedorov and J. Brown. Equatorial waves. *Encyclopaedia of ocean sciences*, pages 3679–3695, 2009.

A. Troccoli. Seasonal climate forecasting. *Meteorological Applications*, 17:251–268, 2010.

R.H. Weisberg and Ch. Wang. A western Pacific oscillator paradigm for the El Niño-Southern Oscillation. *Geophysical Research Letters*, 24:779–782, 1997.

J. Picaut, F. Masia, and Y. Du Penhoat. An advective-reflective conceptual model for the oscillatory nature of the ENSO. *Science*, 277:663–666, 1997.

F. F. Jin. An equatorial ocean recharge paradigm for ENSO. Part II: a stripped down coupled model. *Journal of Atmospheric Sciences*, 54:830–847, 1997b.

Ch.S. Meinen and M.J. McPhaden. Observations of warm water volume changes in the equatorial pacific and their relationship to el niño and la niña. *Journal of Climate*, 13:3551–3559, 2000.

Ch. Wang. A unified oscillator model for the El Niño-Southern Oscillation. *Journal of Climate*, 14:98–115, 2001.

K. Wyrtki. El Niño - the dynamic response of the equatorial Pacific Ocean to atmospheric forcing. *Journal of Physical Oceanography*, 5:572–584, 1975.

N. Ramesh and R. Murtugudde. All flavours of El Niño have similar early subsurface origins. *Nature Climate Change*, 3:42–46, 2013.

S. Hare and N. Mantua. Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Progress in Oceanography*, 47:103–145, 2000.

J. Ballester, S. Bordoni, D. Petrova, and X. Rodó. On the dynamical mechanism explaining the western pacific subsurface temperature buildup leading to ENSO events. *Geophysical Research Letter*, 42:2961–2967, 2015.

D. Petrova, S.J. Koopman, J. Ballester, and X. Rodó. Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model. *Climate Dynamics*, 2016.

X. Yu and M. McPhaden. Seasonal variability in the equatorial Pacific. *Journal of Physical Oceanography*, 29:925–947, 1999.

J. Ballester, S. Bordoni, D. Petrova, and X. Rodó. Heat advection processes leading to El Niño events as depicted by an ensemble of ocean assmiliation products. *Journal of Geophysical Research: Oceans*, 121:3710–3729, 2016a.

M. McPhaden. Evolution of the 2002/2003 El Niño. *American Meteorological Society*, 85:677–695, 2004.

I. Eisenman, L. Yu, and E. Tziperman. Westerly wind bursts: ENSO's tail rather than the dog? *Journal of Climate*, 18:5224–5238, 2005.

E. Tziperman and L. Yu. Quantifying the dependence of westerly wind bursts on the large-scale tropical Pacific SST. *Journal of Climate*, 20:2760–2768, 2007.

G. Gebbie and E. Tziperman. Predictability of SST-modulated westerly wind bursts. *Journal of Climate*, 22:3894–3909., 2009.

K. Wyrtki. Water displacements in the Pacific and the genesis of El Niño cycles. *Journal of Geophysical Research*, 90:7129–7132, 1985.

F. F. Jin, S. T. Kim, and L. Bejarano. A coupled-stability index for ENSO. *Geophysical Research Letters*, 33, 2006.

J. Ballester, D. Petrova, S. Bordoni, and X. Rodó. Sensitivity of El Niño intensity and timing to preceding subsurface heat magnitude. *Scientific Reports*, 6, 2016b.

M. Ishii, A. Shouji, S. Sugimoto, and T. Matsumoto. Objective analyses of SST and marine meteorological variables for the 20th century using COADS and the Kobe Collection. *International Journal of Climatology*, 25:865–879, 2005.

E. Rasmusson and T. Carpenter. Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Weather Rev.*, 110:354–384, 1982.

G.N. Kiladis and H. F. Diaz. Global climatic anomalies associated with extremes in the Southern Oscillation. *Journal of Climate*, 2:1069–1090, 1989.

L. Goddard, S. Mason, S. Zebiak, Ch. Ropelewski, R. Basher, and M. Cane. Current approaches to seasonal-to-interannual climate predictions. *International Journal of Climatology*, 21:1111–1152, 2001.

A.V. Fedorov. Ocean–atmosphere coupling. *The Oxford Companion to Global Change*, pages 369–374, 2008.

J. Shukla. Predictability of the tropical atmosphere. *NASA Technical Reports*, 1981.

T.P. Barnett, M. Latif, N. Graham, M. Flugel, S. Pazan, and W. White. ENSO and ENSO-related predictability. Part I: prediction of equatorial Pacific sea surface temperature with a hybrid coupled ocean-atmosphere model. *Journal of Climate*, pages 1545–1566, 1993.

E. K. Jin, J. L. Kinter III, B. Wang, C.K. Park, I.S. Kang, B. Kirtman, J.S. Kug, A. Kumar, J.J. Luo, J. Schemm, et al. Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Climate Dynamics*, 31:647–664, 2008.

M. Latif, D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, and E. Schneider. A review of the predictability and prediction of ENSO. *Journal of Geophysical Research*, 103:14375–14393, 1999.

K. Trenberth. The definition of El Niño. *Bulletin of the American Meteorological Society*, 78: 2771–2777, 1997.

D. Petrova, J. Ballester, S.J. Koopman, and X. Rodó. Multi-year statistical prediction of ENSO enhanced by the tropical Pacific observing system. *submitted to PNAS*, 2017.

B.P. Kirtman and S.E. Zebiak. ENSO simulation and prediction with a hybrid coupled model. *Monthly Weather Review*, 125:2620–2641, 1997.

M. Lengaigne, E. Guilyardi, J.Ph. Boulanger, Ch. Menkes, P. Delecluse, P. Inness, J. Cole, and J. Slingo. Triggering of El Nino by westerly wind events in a coupled general circulation model. *Climate Dynamics*, 23:601–620, 2004.

M.A. Cane, S.E. Zebiak, and S.C. Dolan. Experimental forecasts of El Niño. *Nature*, 321: 827–832, 1986.

D. Chen, M. Cane, A. Kaplan, S. Zebiak, and D. Huang. Predictability of El Niño over the past 148 years. *Nature*, 428:15, 2004.

J. Neelin, D. Battisti, A. Hirst, F.-F. Jin, Y. Wakata, T. Yamagata, and S. Zebiak. ENSO theory. *Journal of Geophysical Research: Oceans*, 103:14261–14290, 1998.

P. Gonzalez and L. Goddard. Long-lead ENSO predictability from CMIP5 decadal hindcasts. *Climate Dynamics*, 46:3127–3147, 2016.

M. McPhaden and X. Yu. Equatorial waves and the 1997/98 El Niño. *Geophysical Research Letters*, 26:2961–2964, 1999.

B. P. Kirtman and P. S. Schopf. Decadal variability in ENSO predictability and prediction. *Journal of Climate*, 11:2804–2822, 1998.

F. F. Jin, J. D. Neelin, and M. Ghil. El Niño on the devil's staircase - annual subharmonic steps to chaos. *Science*, 264:70–72, 1994.

E. Tziperman, L. Stone, M. Cane, and H. Jarosh. El Niño chaos: overlapping of resonances between the seasonal cycle and the Pacific ocean-atmosphere oscillator. *Science*, 264:72–74, 1994.

D. Chen, S. Zebiak, A.J. Busalacchi, and M. Cane. An improved procedure for El Niño forecasting: implications for predictability. *Science*, 269:1699–1702, 1995.

A. Barnston, M. Tippett, M. L'Heureux, Sh. Li, and D. DeWitt. Skill of real-time seasonal ENSO model predictions during 2002-11. Is our capability increasing? *American Meteorological Society*, 93:631–651, 2012.

S. E. Zebiak and M. A. Cane. A model El Niño-Southern Oscillation. *Monthly Weather Review*, 115:2262–2278, 1987.

IRI. ENSO forecasting plume. *http://iri.columbia.edu/climate/ENSO/currentinfo/SST_table.html*, 2017.

T.N. Palmer, Č. Branković, and D.S. Richardson. A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Quarterly Journal of the Royal Meteorological Society*, 126:2013–2033, 2000.

M.A. Balmaseda, M.K. Davey, and D. Anderson. Decadal and seasonal dependence of ENSO prediction skill. *Journal of Climate*, 8:2705–2715, 1995.

D. Gu and S.G.H. Philander. Interdecadal climate fluctuations that depend on exchanges between the tropics and the extratropics. *Science*, 275:805–807, 1997.

P.J. Webster. The annual cycle and the predictability of the tropical coupled ocean-atmosphere system. *Meteorology and Atmospheric Physics*, 56:33–55, 1995.

B. N. Goswami and J. Shukla. Predictability of a coupled ocean-atmosphere model. *Journal of Climate*, 4:3–22, 1991.

A.J. Clarke and S. van Gorder. The connection between the boreal spring Southern Oscillation persistence barrier and biennial variability. *Journal of Climate*, 12:610–620, 1999.

Y. Xue, M. Cane, S. Zebiak, and M. Blumenthal. On the prediction of ENSO: a study with a low-order Markov model. *Tellus A*, 46:512–528, 1994.

S. Zebiak. Oceanic heat content variability and El Niño cycles. *Journal of Physical Oceanography*, 19:475–486, 1989.

M. McPhaden. Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophysical Research Letters*, 30, 2003.

S. Zebiak. Tropical atmosphere-ocean interaction and the El Niño/Southern Oscillation phenomenon. *PhD Thesis, Massachusetts Institute of Technology*, 1985.

C. Penland and T. Magorian. Prediction of Niño3 sea surface temperatures using linear inverse modeling. *Journal of Climate*, 6:1067–1076, 1993.

A.J. Clarke and S. van Gorder. Improving El Niño prediction using a space-time integration of indo-pacific winds and equatorial Pacific upper ocean heat content. *Geophysical Research Letters*, 30, 2003.

F.T. Tangang, W.W. Hsieh, and B. Tang. Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Climate Dynamics*, 13:135–147, 1997.

H.M. Van den Dool. Searching for analogues, how long must we wait? *Tellus A*, 46:314–324, 2007.

X. Rodó, M.A. Rodriguez-Arias, and J. Ballester. The role of ENSO in fostering teleconnection patterns between the tropical north Atlantic and the western Mediterranean basin. *CLIVAR Exchanges*, 11:26–27, 2006.

M. McPhaden, A. Busalacchi, R. Cheney, J.-R. Donguy, K. Gage, D. Halpern, M. Ji, P. Julian, G. Meyers, G. Mitchum, et al. The Tropical Ocean-Global Atmosphere observing system: A decade of progress. *Journal of Geophysical Research: Oceans*, 103:14169–14240, 1998.

A. Harvey and S. J. Koopman. Signal extraction and the formulation of unobserved components models. *The Econometrics Journal*, 3:84–107, 2000.

A. Harvey, S. J. Koopman, and J. Penzer. Messy time series. In T.B. Fomby and R. Carter Hill, editors, *Advances in Econometrics*, volume 13. New York: JAI Press, 1998.

J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2 edition, 2012.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering , Transactions, ASMA, Series D*, 82:35–45, 1960.

S. I. An and F. F. Jin. Nonlinearity and asymmetry of ENSO. *Journal of Climate*, 17:2399–2412, 2004.

S. I. An and J. Choi. Seasonal locking of the ENSO asymmetry and its influence on the seasonal cycle of the tropical eastern Pacific sea surface temperature. *Atmospheric Research*, 94:3–9, 2009.

O. Fashé and B. Dewitte. Enhancement of near-annual variability in the equatorial Pacific in 2000-2008. *Advances in Geosciences*, 33:13, 2013.

R. Kleeman, J. McCreary, and B. Klinger. A mechanism for generating ENSO decadal variability. *Geophysical Research Letters*, 26:1743–1746, 1999.

S. Yeh and B. Kirtman. Tropical Pacific decadal variability and ENSO amplitude modulation. *Geophysical Research Letters*, 32, 2005.

J. Dutton. Opportunities and priorities in a new era for weather and climate services. *Bulletin of the American Meteorological Society*, pages 1303–1311, 2002.

J. Ballester, R. Lowe, P. Diggle, and X. Rodó. Seasonal forecasting and health impact models: challenges and opportunities. *Annals of the New York Academy of Sciences*, 1382:8–20, 2016c.

GFCS. Global Framework for Climate Services, the World Meteorological Organization. `http://www.wmo.int/gfcs/`, 2017.

C.F. Ropelewski and M.S. Halpert. Global and regional scale precipitation patterns associated with El Niño/Southern Oscillation. *Monthly Weather Review*, 115:1606–1626, 1987.

T.N. Stockdale, D. Anderson, M. Balmaseda, F. Doblas-Reyes, L. Ferranti, K. Mogensen, T. Palmer, F. Molteni, and F. Vitart. ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Climate Dynamics*, 37:455–471, 2011.

J. Bendix and W. Lauer. The rainy seasons in Ecuador and their climatic interpretation. *Erdkunde*, 46:118–134, 1992.

E. Moran-Tejeda, J. Bazo, J. Lopez-Moreno, E. Aguilar, C. Azorin-Molina, A. Sanchez-Lorenzo, R. Martinez, J. Nieto, R. Mejia, N. Martin-Hernandez, and S. Vicente-Serrano. Climate trends and variability in Ecuador (1966–2011). *International Journal of Climatology*, 2016.

R. Lowe, A.M. Stewart-Ibarra, D. Petrova, M. García-Díez, M. Borbor-Cordova, R. Mejía, M. Regato, and X. Rodó. Climate services for health: predicting the evolution of the 2016 dengue season in Machala, Ecuador. *submitted to the Lanset Planetary Health*, 2017.

A. Levine and M. McPhaden. How the July 2014 easterly wind burst gave the 2015–2016 El Niño a head start. *Geophysical Research Letters*, 43:6503–6510, 2016.

J. Ballester, M. A. Rodriguez-Arias, and X. Rodó. A new extratropical tracer describing the role of the western pacific in the onset of El Nñio: Implications for ENSO understanding and forecasting. *Journal of Climate*, 24:1425–1437, 2011.

Sh. Hu and A. Fedorov. Exceptionally strong easterly wind burst stalling El Niño of 2014. *Proceedings of the National Academy of Sciences*, 113:2005–2010, 2000.

CPC. Cold and warm episodes by season. *http://www.cpc.ncep.noaa.gov/products/ analysis_monitoring/ensostuff/ensoyears.shtml*, 2017.

N. Ramesh, M. Cane, R. Seager, and D. Lee. Predictability and prediction of persistent cool states of the Tropical Pacific Ocean. *Climate Dynamics*, pages 1–17, 2016.

A. Timmermann and F. Jin. A nonlinear mechanism for decadal El Niño amplitude changes. *Geophysical Research Letters*, 29:445–466, 2003.

G. A Meehl, J. M Arblaster, J. Fasullo, A. Hu, and K. Trenberth. Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Climate Change*, 13:360–364, 2011.

K. Kuhn, D. Campbell-Lendrum, A. Haines, and J. Cox. Using climate to predict infectious disease epidemics. Geneva: World Health Organization. *http://www.who.int/ globalchange/publications/infectdisease/en/index.html*, 2005.