UNIVERSITAT DE BARCELONA

Atmospheric characterization and time-series analysis of the impact of environmental factors on disease onset

Alejandro Fontal Navarrete

Atmospheric characterization and time-series analysis of the impact of environmental factors on disease onset

Doctoral Thesis

by

Alejandro Fontal

submitted as pre-requisite to obtain the degree of

DOCTOR OF PHILOSOPHY (PH.D.) IN **BIOTECHNOLOGY**

at

UNIVERSITAT DE BARCELONA

Facultat de Farmàcia i Ciències de l'Alimentació

Institution

INSTITUT DE SALUT GLOBAL DE BARCELONA (ISGLOBAL)

Thesis Director: **Dr. Xavier RODÓ**

Climate & Health program Head at ISGlobal

Thesis Tutor: **Dr. Carlos BALSALOBRE**

Professor of Microbiology at Universitat de Barcelona

Barcelona, September 2024

Acknowledgements

This thesis is the culmination of five years of work, but more than that, it marks five years of my life where I have grown, not just as a professional, a researcher, and a scientist, but more importantly, as a person. While the many pages that follow this section are a testament to part of the work accomplished over these past years, I feel it is necessary to express my gratitude to everyone and everything that has made my journey during this half-decade a much more pleasant experience than it might have been otherwise. It has been a journey filled with ups and downs, with moments of immense personal and professional excitement, but also times of pessimism, stress, and struggle. If it weren't for everyone who has made my days a bit brighter, pushing me and believing in me, I would certainly not have been able to persevere to the end. For that, I want to thank you all.

For starters, I would like to thank my supervisor and thesis director, Xavi, for giving me the opportunity to pursue this PhD and enter the fascinating world of researching the dynamical systems that govern our planet, from atmospheric processes to disease. It's an extremely complex topic but incredibly rewarding, and while at times I have missed studying the more lab-controlled scenarios I was used to until I started this, it's given me the opportunity to have an impact with my research and an understanding of the world that I would not have imagined before. Thanks for your guidance and for your continuous belief on my skills and support on my work, from the office in Barcelona to the cold of the winter in Osaka.

I'm also deeply grateful to everyone involved in the HELICAL ITN. This project didn't just fund my salary, research, and trips; it gave me the chance to work alongside a remarkable group of people, to be part of something bigger than myself. The chance to regularly meet, online and physically, with my fellow ESRs, gave me a sense of community that was incredibly helpful.

I am also incredibly grateful to the women of the AIRLAB: Sílvia, Sofya, Lídia, who adopted me into the group from the very first day, taught me how to work making my hands dirty outside of the bubble of my keyboard and screens, and made me feel like I was completely at home every time I walked through the door of the lab.

To my friends and colleagues who have been with me since day one, Adrià and Maxime: thank you. We've walked this path together, sharing the frustrations and the victories, and there's was

something really comforting in knowing that I was never alone in this. Sharing our pain and our joy made the lows less daunting and the highs much sweeter. Whether it was our long philosophical chats on the terrace, ping-pong games, or our beloved vermuts or kudos on Strava, each moment either motivated me or gave me the escape I needed. Thank you for being part of this journey.

To all my friends and colleagues at PRBB: every time I walked into that building, I knew I would be surrounded by people who cared and who made my days brighter. On the days when I felt disheartened or uncertain about my work, it was the people around me that kept me going. There's too many of you to name you all, and I am writing this at the very verge of running out of time to submit to make sure I don't forget any of you, but I look back in absolute gratefulness for all the moments I have shared in the PRBB and outside of it in a way that will make me really sad the day I leave. Thanks for all the absurd discussions during lunches in the terrace, for all afterwork beers in the football bar or in Pasatapas and all the official and unofficial volleyball matches in the beach. You made Barcelona feel like more than just my hometown, you made me rediscover it with new eyes. I will miss it all when this chapter closes. Thank you, from the bottom of my heart.

Gracias también de todo corazón a mis amigos, todos aquellos que ya formaban parte de mi vida antes de empezar este doctorado, y que hoy en día siguen siendo parte de ella. A mis amigos biotecs, tanto de Barcelona como de Wageningen: Alicia, Agnès, Patty, Miquel, Aran, Juanca… Gracias por estos casi 10 años (y algunos, más) sufriendo juntos, apoyándonos mutuamente y compartiendo risas. Siempre habéis sido ese refugio que me hace sentir en casa sin importar en qué parte del mundo nos encontremos.

Gracias también a ti, Agata, por subir al cielo, bajar al infierno, y volver a la tierra conmigo. Sé que no fue fácil y pese a todo fuiste un soporte vital en un buen tramo de esta maratón, y lo que he aprendido contigo me servirá de por vida.

Gracias a mis amigos de siempre: Cristian, Eric, Balas, Isra, Fran, Pablo. Habéis sido una constante en mi vida desde que tengo memoria, y esa roca sobre la que siempre he podido apoyarme. Gracias por ser mi refugio fuera del mundo académico, por recordarme quién soy y de dónde vengo, pero también por apoyarme mientras busco a dónde quiero ir. A ti también, Joserra, por junto a Cristian hacer de nuestro piso en Sagrada Familia mi segunda familia y un verdadero hogar. Todos esos meses de pandemia, sin balcón y llenos de incertidumbre, fueron realmente mágicos y sea donde sea que estás ahora nunca te olvidaremos.

Finalmente, a mi familia. A mis yayos, por siempre estar ahí, por darme ese lugar donde podía descansar y sentirme seguro. A mi hermana Elena, por ser no solo mi hermana, sino también una amiga cercana y una vecina inmejorable. Y a mis padres, José e Isabel, porque sin vuestro apoyo y sacrificio, jamás habría llegado hasta aquí. Aunque muchas veces no nos entendimos, aunque a veces parecía que hablábamos idiomas distintos, hoy me siento increíblemente afortunado de haberos tenido como padres. Espero poder devolveros, aunque sea en parte, todo el amor y apoyo incondicional que me habéis dado desde siempre.

Preface

This thesis was conducted at the Barcelona Institute for Global Health (ISGlobal), between October 2019 and September 2024, under the supervision of Prof. Dr. Xavier Rodó. The content and structure of this thesis adhere to the doctoral thesis procedures and regulations for the submission of a thesis by compendium of publications, as required for the PhD in Biotechnology program at the Facultat de Farmàcia i Ciències de l'Alimentació of the Universitat de Barcelona.

A significant portion of this thesis was developed under the umbrella of the HELICAL (HEalth data LInkage for ClinicAL benefit) project, a Marie Skłodowska-Curie Actions Innovative Training Network funded by the European Union's Horizon 2020 program. HELICAL focused on leveraging recent advances in data science to link research datasets with longitudinal healthcare records. The PhD candidate participated in this project as an Early-Stage Researcher within Work Package 1, which focused on the environmental impacts on autoimmunity, with vasculitis serving as the primary paradigm.

Due to the COVID-19 pandemic, parts of the original work planned for this thesis had to either be changed, postponed or cancelled. However, these changes also presented new research opportunities, particularly reflected in Paper I, where we investigated the influence of climatic variables, such as temperature and absolute humidity, on the spread of COVID-19 at both local and global scales. The initial thesis plan aimed to explore diseases situated on the continuum between non-communicable and infectious diseases, using them as models to better understand the behavior of Kawasaki Disease: an illness whose etiology remains a topic of considerable debate regarding its classification as infectious or non-communicable. The unexpected emergence of COVID-19 during this period provided a wealth of data, enabling the development of a robust statistical framework to test key hypotheses during a time of global crisis. The publication of the findings attracted significant media attention, with the study being featured on television and printed media worldwide, reflecting the relevance of the research during a time of global concern.

A significant setback due to the pandemic was the inability to perform on-site sampling campaigns as originally scheduled. The thesis initially included the design of sampling campaigns to collect and analyse air samples for biological and chemical characterization in Japan, with the

intention of correlating these data with health outcomes, particularly KD incidence. However, Japan remained closed to foreigners for an extended period, and the sampling campaign could not be performed until January 2023. KD incidence data, which is collected biannually through nationwide surveys and released on odd years, could not be obtained in time for this thesis—the corresponding data for 2023 won't be released until 2025. Despite these challenges, the PhD candidate coordinated and conducted the sampling in Osaka and indirectly from three other sites in Japan, resulting in a unique dataset. This experimental collection of data will prove valuable not only for elucidating the etiological triggers of KD but also for future research on other diseases and for a more thorough understanding of the dynamics of the microbiome of the air.

Abstract

This thesis investigates the complex interplay between atmospheric characterization, environmental factors, and the onset of diseases, specifically focusing on Kawasaki Disease (KD) and COVID-19. Through time-series analysis and multidisciplinary approaches, it explores how air quality, climatic conditions, and airborne microbial diversity influence disease incidence and progression, offering new insights into the role of the environment in shaping public health outcomes.

In the first study, we analyze the association between ultrafine aerosols enriched in metals—originating from intensive farming and urban pollution—and the incidence of KD in Japan. Utilizing daily variability data of fine aerosols and an unprecedented daily KD epidemiological record, we uncover a strong dynamical link. This suggests that exposure to these metal-rich aerosols may trigger immune responses leading to KD, accounting for over 40% of the disease's variability and highlighting the significance of environmental triggers in its etiology.

The second study examines the seasonality of KD through an age-stratified spatiotemporal analysis over a 20-year period in Japan. We identify significant temporal shifts in disease incidence, particularly an abrupt desynchronization between younger and older children after 2016. This shift is most evident in children under two years old, revealing a new autumn peak consistent across most regions. These findings suggest changes in environmental triggers and exposure windows, offering potential explanations for the mysterious patterns of this pediatric vasculitis.

Expanding the scope, the third study explores the climatic signatures in the different COVID-19 pandemic waves across both hemispheres. By employing (and developing) statistical methods designed to detect transitory associations, we demonstrate strong consistent negative effects of temperature and absolute humidity on COVID-19 cases. Our findings classify COVID-19 as a seasonal low-temperature infection, emphasizing the role of climatic factors in its transmission dynamics and underscoring the importance of considering airborne pathways in public health interventions.

The fourth study evaluates the diversity of microorganisms in aerosols above the planetary boundary layer, confirming the long-distance transport—up to 2,000 km—of viable human

pathogens and antimicrobial-resistant bacteria. Through tropospheric aircraft surveys over Japan, we identify a vast array of bacterial and fungal taxa, including potential human pathogens originating from agricultural regions in Northeast Asia. This discovery highlights a novel pathway for disease dispersion and the need to consider atmospheric microbial communities in disease etiology and public health policies.

The final study presents an innovative methodology for the real-time detection of bacteria in bioaerosols using Laser-Induced Fluorescence (LIF) and machine learning. By modifying existing equipment and testing with artificially generated aerosols enriched with different bacterial species, we achieve high accuracy in detecting (96.74%) and classifying (69.24% of class-balanced accuracy between 5 bacterial species) bacterial particles. This approach offers valuable tools for rapid monitoring of airborne microbial communities, enhancing both ecological studies and public health surveillance.

Collectively, these studies provide a comprehensive examination of how atmospheric characterization and environmental factors influence disease onset. By bridging atmospheric science and epidemiology, this thesis contributes new perspectives on the environmental determinants of KD and COVID-19. It lays the groundwork for future research aimed at mitigating the impacts of air pollution, climate variability, and bioaerosol exposure on human health, with significant implications for public health strategies and policies.

Table of Contents

1. Introduction

1.1 Vasculitis

Vasculitis refers to inflammation of blood vessel walls, characterizing a group of rare diseases known as vasculitides. These diseases can affect vessels of any size, from small capillaries to large arteries, and can involve virtually any organ system. This broad potential for organ involvement leads to a wide range of clinical manifestations, from mild symptoms to life-threatening conditions. Depending on the affected organs, patients may present with skin rashes, fever, neurological deficits, renal dysfunction, or respiratory difficulties.

Management and prognosis vary significantly among the different types of vasculitis, highlighting the importance of early and accurate diagnosis. Many forms initially present with nonspecific symptoms such as fatigue, weight loss, and fever, which can delay recognition and treatment. Misdiagnosis or delayed intervention can result in irreversible organ damage or severe complications.

Traditionally, vasculitides are classified based on the predominant size of the vessels affected (Figure 1), which helps guide both diagnostic approaches and treatment strategies. This classification system divides vasculitis into three broad categories:

- **Large-vessel vasculitis**: This category includes diseases such as Takayasu Arteritis and Giant Cell Arteritis (GCA), which primarily affect the larger arteries. These diseases can lead to serious complications, including stroke, aneurysms, and organ ischemia due to the narrowing or obstruction of large blood vessels.
- **Medium-vessel vasculitis**: Diseases in this category, such as Polyarteritis Nodosa and Kawasaki Disease (KD), primarily affect medium-sized arteries. Kawasaki Disease, which will be discussed in greater detail in this thesis, is of particular relevance as it predominantly affects children and is the leading cause of acquired heart disease in pediatric populations in developed countries. Polyarteritis Nodosa, on the other hand, can affect multiple organ systems and is associated with serious vascular complications such as aneurysms and infarction.
- **Small-vessel vasculitis**: This group of vasculitides affects the smaller vessels, including capillaries, arterioles, and venules. It is further subdivided into:

- **ANCA-associated vasculitis (AAV):** This includes Granulomatosis with Polyangiitis (GPA, formerly Wegener's granulomatosis), Eosinophilic Granulomatosis with Polyangiitis (EGPA, formerly Churg-Strauss syndrome), and Microscopic Polyangiitis (MPA). These conditions are characterized by the presence of anti-neutrophil cytoplasmic antibodies (ANCA) and often involve the respiratory and renal systems.
- **Non-ANCA-associated vasculitis**: This subgroup includes conditions such as Cryoglobulinemic Vasculitis, IgA Vasculitis (Henoch-Schönlein purpura), and Anti-C1q Vasculitis. These diseases are not associated with ANCAs but still lead to widespread inflammation and damage in small blood vessels, often affecting the skin, kidneys, and gastrointestinal tract.

- **Variable-vessel vasculitis:** Behçet's Disease is a form of vasculitis that affects blood vessels of all sizes, including both arteries and veins, and it is unique in its involvement of variable vessel types. This multisystem inflammatory disorder primarily presents with recurrent oral and genital ulcers, skin lesions, and uveitis, but it can also affect multiple organ systems, including the cardiovascular, gastrointestinal, and neurological systems. Due to its ability to affect both small and large vessels, Behçet's Disease poses a challenge for diagnosis and treatment, often requiring a multidisciplinary approach.

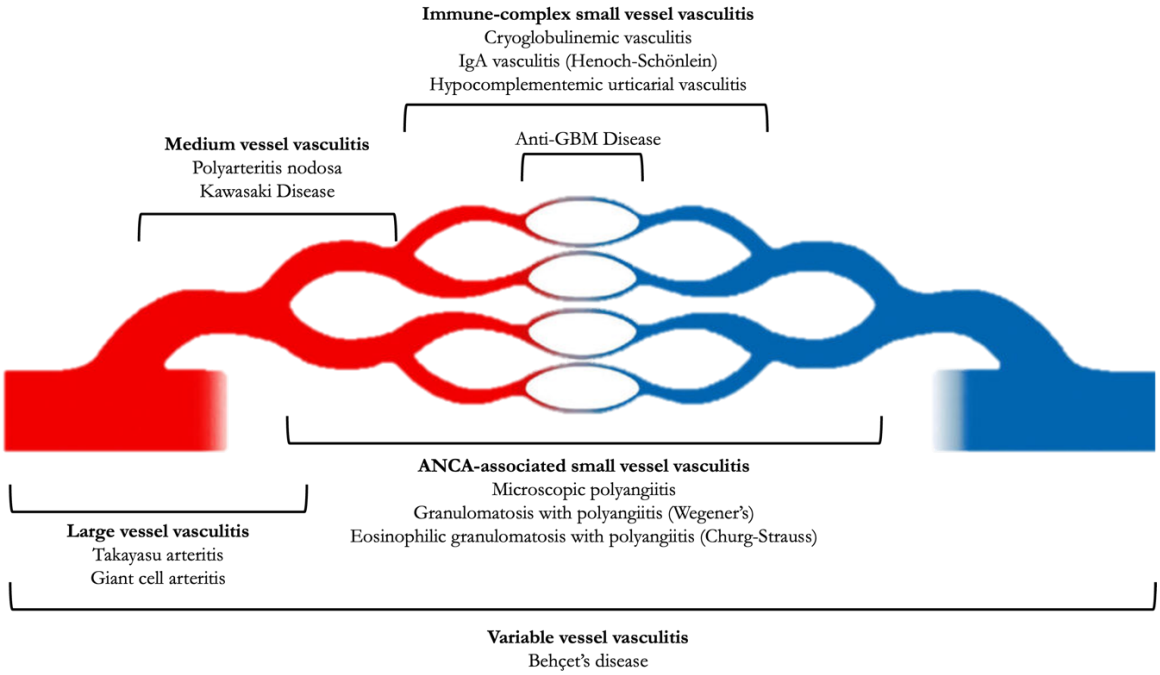


Figure 1. Schematic classification of vasculitides based on the size of the affected vessels. Adapted from the main site of the Vasculitis International association.

Understanding the classification and pathophysiology of these diseases is essential, as treatment options can differ widely. Large- and medium-vessel vasculitis may require high-dose corticosteroids and immunosuppressive therapy to control inflammation and prevent further vascular damage. In contrast, some forms of small-vessel vasculitis, particularly ANCA-associated variants, may involve a combination of immunosuppression and plasma exchange in severe cases.

In the context of this thesis, the focus is on Kawasaki Disease, a medium-vessel vasculitis that has unique epidemiological patterns, presents mainly in young children and for which extensive since the 1970s are available. As we will explore in subsequent sections, KD presents distinct challenges in understanding its etiology, seasonal variations, and the specific risk factors contributing to its onset.

1.2 Kawasaki Disease

1.2.1 General background and history

Kawasaki Disease is an acute, self-limited vasculitis that predominantly affects children under five years of age. It initially presents with high fever, mucocutaneous inflammation, and cervical lymphadenopathy. Without timely administration of intravenous immunoglobulin (IVIG) during the acute phase, approximately 20% of untreated children develop coronary artery aneurysms, leading to serious long-term cardiovascular complications such as myocardial infarction and sudden death. Currently, KD is the leading cause of acquired heart disease in children in developed countries, surpassing rheumatic fever [1,2].

It has now been over 60 years since Dr. Tomisaku Kawasaki first documented cases of what is now recognized as Kawasaki Disease. By 1967, Kawasaki had recorded 50 cases, leading him to publish a seminal report in the *Japanese Journal of Allergy* [3] in Japanese, where he meticulously detailed all the observable clinical aspects of this, at the time, novel illness. He described it as a "Pediatric acute febrile mucocutaneous lymph node syndrome with characteristic desquamation of fingers and toes". The first publication referring to the disease in an international journal did not appear until 1974, when Kawasaki and colleagues published an article in *Pediatrics* describing "A New Infantile Acute Febrile Mucocutaneous Lymph Node Syndrome (MLNS) prevailing in Japan" [4] . This article was not as comprehensive as his initial manuscript in Japanese, but it marked the disease's introduction to the international medical community under what later became known as Kawasaki Disease.

Many of the observations and insights that Kawasaki disclosed in his original Japanese manuscript went unnoticed by the English-speaking medical community until an English translation was made available in 2002 [5], by which time several of his findings had been "re-discovered" by other researchers during the late 20th century [6].

Nevertheless, the report published in 1967 had already garnered significant attention within the Japanese medical community, prompting widespread recognition of the disease. In response, the Japanese Ministry of Health and Welfare funded Dr. Kawasaki's team to conduct the first of many nationwide surveys that have since then been performed and curated by the team at Jichi Medical University since 1970. The results from these surveys [7–11] represent the most

extensive and complete KD database to date, have provided critical insights and fueled decades of research into the disease.

1.2.2 Temporal dynamics

The extensive records for KD in Japan have allowed for a comprehensive observation of the evolution of the disease's incidence over time. These records have enabled practitioners and researchers to gain a deep understanding of the characteristic temporal dynamics of KD, which can be summarized into three key features:

- **Nationwide epidemics**: From the beginning of the records through the end of the 20th century, three significant nationwide epidemics of KD occurred in 1979, 1982, and 1986. These epidemics were remarkable in that they occurred synchronously across Japan [12,13], drawing attention to the potential role of widespread epidemiological factors influencing disease onset.

From the beginning of the 21st century, two new features start to become noticeable:

- **Upward Incidence Trend**: Incidence rates showing a noticeable upward trend - the incidence rate in 2019 was nearly double that of the epidemic peak in 1982, rising from 130 to 370.8 yearly cases per 100,000 children under five years old from 2000 to 2019. The only exception was the last year of the records, 2020, which showed a significant reduction in cases compared to previous years in the context of the COVID-19 global pandemic [11].

- **Yearly seasonality**: A marked seasonal pattern, with a consistent and prominent peak during winter, followed by a less consistent and weaker peak during spring, and a nadir during the fall, has been a hallmark of the epidemiology of the disease since 2000.

Nationwide epidemics

The nationwide epidemics of KD in Japan represent one of the most intriguing phenomena in the epidemiology of the disease. These epidemic waves unfolded in a rather synchronous manner across the whole country, but with distinct features in each case [14,13]. The country-aggregated and prefecture-wise weekly incidence change is depicted for the three epidemics in Figure 2.

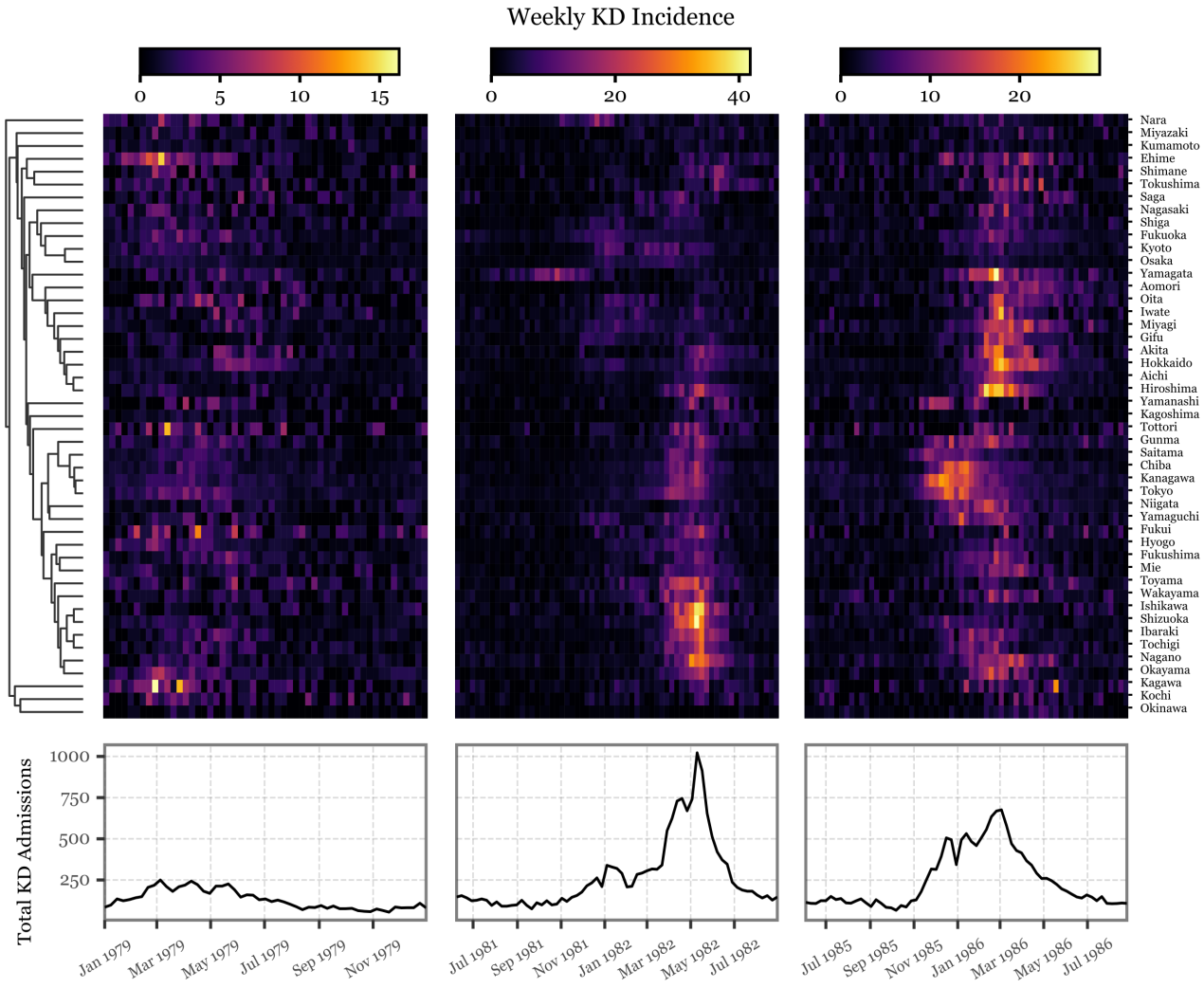


Figure 2. Aggregated and prefecture-wise depiction of the nationwide epidemics in Japan. On top, the three heatmaps represent the prefecture-wise weekly incidence during the 1979, 1982 and 1986 epidemics, respectively. On the bottom, the total number of KD admissions per week aggregated for all of Japan. Ordination of the prefectures by rows was made according to a hierarchical clustering using the inverse of the correlation between the weekly incidences of each prefecture, as the dendrogram in the left side showcases, with the length of the bars representing the distance between each pair of prefectures or to the closest node.

The 1979 epidemic began in March in western Japan, with the eastward progression spreading to all regions within a three-month period. The second epidemic in 1982 exhibited a more complex pattern, with an early peak in January, followed by a significantly larger peak in May. The origin of the 1982 epidemic was multifocal, arising from four distinct areas. By contrast, the 1986 epidemic followed a more linear progression, similar to 1979, but occurred across a longer time frame, peaking in April. These distinctive temporal and spatial patterns raised important questions regarding the nature of the causative agent(s) involved in these epidemics and their mode of spread, which we will explore further in the subsequent section on the etiology of the disease.

Long-term increasing trend

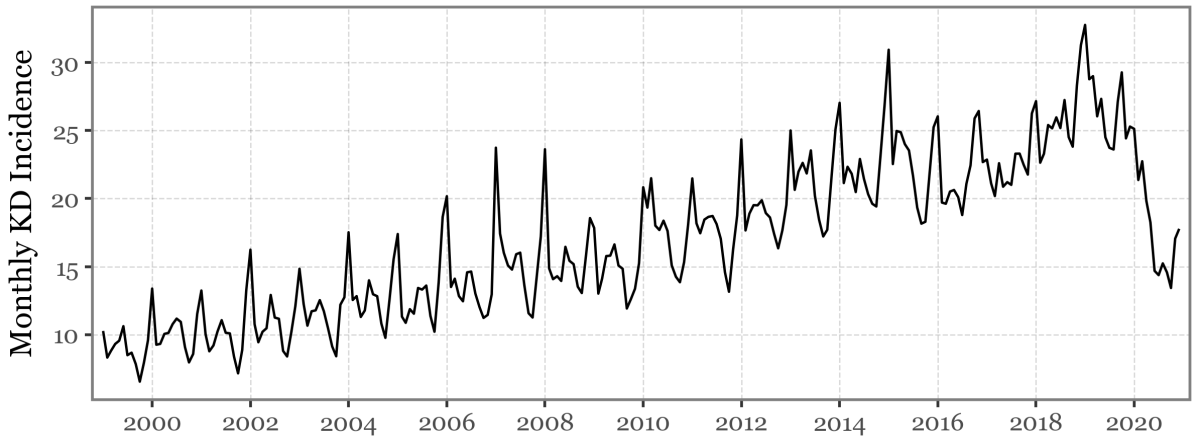


Figure 3. Monthly KD Incidence time series for Japan (1999-2020). KD incidence rates in Japan as the number of cases per 100.000 children under the age of 5 years old for the 1999 to 2020 period.

One of the most alarming aspects of the disease from the public health perspective is the upward trend in KD incidence, which became particularly noticeable at the start of the 21st century. The incidence rate in Japan increased from around 120 cases per 100,000 in the early 2000s to nearly 400 cases per 100,000 by 2019 in children under 5 years old (Figure 3, [11]).

While Japan has the highest incidence by a significant margin, the disease is also highly prevalent in two other East Asian countries with robust nationwide incidence data: Korea and Taiwan. These two countries, along with Japan, had confirmed yearly incidences of over 150 cases per 100,000 children just before the COVID-19 pandemic, with both showing growing trends over recent decades until the significant reduction in cases observed in 2020 and 2021 [11,15,16] due to the pandemic. The contribution of genetics to KD susceptibility is evidenced by the consistently higher incidence rates in East Asian countries but also of children of East Asian ethnicity everywhere in the world, where they have been shown to reliably show higher prevalence of the disease than their peers from other ethnicities [17].
The phenomenon of the increasing trend is not unique to these East Asian countries [18]. Several other Asian countries exhibit a similar situation, with growing incidence rates observed in China, India [19], Malaysia [20] and the Philippines [21].

The incidence in North America [22], Europe [23], and Australia [24] is at a much lower level, ranging from 5 to 22 cases per 100,000 children under 5 years old per year. In these regions, the incidence was increasing until around 2010, when the rates appeared to plateau.

Determining whether the recorded increases in incidence are due to improved case ascertainment (stemming from increased awareness and access to specialist medical services) or reflect a true rise in disease burden remains challenging. It is likely that both factors play a role, especially in countries outside East Asia, where KD recognition among medical professionals has been more recent. This delayed recognition could account for the initial rise in incidence, followed by stabilization around 2010. However, it is unlikely that the strong and sustained increase observed in Japan, Korea, and Taiwan can be solely attributed to enhanced case detection, given the long-standing recognition and surveillance of the disease in these countries. The persistent upward trend in these regions suggests that other factors, potentially environmental or epidemiological, may be driving the continued increase in KD incidence.

Seasonality

The seasonality of KD has long been one of the most distinctive and studied epidemiological features of the disease. Researchers have investigated its patterns not only in Japan [25] but across the globe [20,21,24,26–28], recognizing seasonality as a key clue that may eventually lead to a deeper understanding of KD's etiology. A comprehensive study comparing KD cases globally found that significant seasonal signals were evident in countries across the extra-tropical Northern Hemisphere, including Canada, China, Finland, France, India, Israel, Italy, Japan, Korea, the Netherlands, Russia, Spain, Turkey, the UK, and the USA [29]. The findings also showed a non-random distribution of seasonal peaks, with winter being the most common period of increased KD incidence. On the other hand, the data from Tropical and Southern Hemisphere countries were less conclusive, with the sparsity of data making it difficult to identify significant seasonal patterns.

In Japan, seasonality has been remarkably consistent over much of the 21st century, with a pronounced peak in cases during the winter months, a secondary but weaker peak during the summer, and a distinct nadir during the fall (see Figure 3). This regular seasonal pattern persisted through much of the period, providing strong evidence that the timing of KD outbreaks is closely tied to environmental and possibly biological cycles that recur each year. However, after 2016, this seasonal signal began to shift, with the previously well-defined patterns becoming more complex. The emergence of new seasonal peaks and the waning of the traditional winter peak during the last years of the records raise intriguing questions about the underlying drivers of KD and their potential changes in recent years.

The seasonality of KD, along with the consistency (or lack thereof) in its patterns, has been one of the key variables used to investigate the potential etiological drivers of the disease. In the following section, we will delve deeper into how these seasonal dynamics may help uncover the underlying causes.

1.2.3 Etiology

The etiology of KD remains one of the most enduring mysteries in pediatric medicine. Despite over five decades of research, no single causative agent has been definitively identified. Multiple hypotheses have been proposed, broadly divided into infectious and environmental origins, with a growing body of evidence suggesting the disease may have a multifactorial etiology. The unique spatiotemporal patterns of KD, particularly its seasonality and epidemic-like waves just mentioned in the previous section, have been key in shaping these hypotheses, pointing to an interplay between environmental triggers and host susceptibility.

Recently, in March 2024, Burns published a comprehensive review on the current state of research regarding KD etiology [30], providing an overview of the prevailing theories and emerging hypotheses. This review further emphasizes the ongoing challenge of pinpointing the precise mechanisms responsible for KD, despite substantial progress in understanding the disease.

Infectious hypothesis

One of the long-standing theories postulates that KD is triggered by an infectious agent [31]. The abrupt onset of high fever in previously healthy children, coupled with symptoms such as conjunctival injection, oral and pharyngeal erythema, cervical lymphadenopathy, and rash, mirrors the clinical findings of various pediatric infections typically acquired through respiratory routes. Moreover, the presence of self-limited aseptic meningitis, hepatitis, and arthritis in a subset of KD patients further supports the idea of a systemic infection. The non-response of KD to antibiotic therapies points away from a bacterial pathogen, while the self-limited nature and generally non-recurring fever suggest an acute infection, rather than an autoimmune process.

Epidemiologically, KD primarily affects very young children, particularly those aged under 5 years, with peak incidence observed between 9 and 11 months of age. This age corresponds to the typical period when children experience common childhood infections due to the waning of passive maternal antibodies. The rarity of KD before 3 months of age supports the idea that maternal antibodies offer some protection against the disease. Furthermore, the seasonal nature

of KD, with higher incidence in the winter-spring period, parallels many common respiratory infections, which reinforces the infectious theory.

Numerous viral and bacterial pathogens have been implicated as potential triggers of KD, with suggestions ranging from retroviruses to streptococcal infections. Efforts to link KD incidence with common viral infections, such as adenoviruses, coronaviruses, and enteroviruses, have been made, but no single pathogen has been consistently identified across different regions or outbreaks [32–35]. A notable meta-analysis published in *Hospital Pediatrics* reviewed multiple studies showing viral co-infections in children with KD [36]. The prevalence of viral co-infections in these cases ranged from 7.5% to 42%, though the methodology and populations studied varied. This variability has led to a lack of consensus on whether viral infections are a direct trigger or merely coincidental in some KD cases.

Clinical studies examining both the innate and adaptive immune responses in KD patients reveal significant variability in circulating antibodies and immune activation [37]. This variability makes it increasingly unlikely that a single infectious agent is solely responsible for triggering the cascade that leads to the onset of KD.

As a counterpoint to the infectious hypothesis, Rodó and colleagues [38,39] developed a SEIR (Susceptible-Exposed-Infected-Recovered) compartmental model to assess the likelihood of an infectious agent being responsible for the short interval between exposure and fever onset, as well as the synchronization of KD outbreaks across cities. Their findings indicate that even the most rapidly replicating infectious agents could not propagate at a speed necessary to account for these patterns, with no known pathogen capable of producing same-day infections over distances between cities through secondary infections alone.

Environmental trigger hypothesis

Parallel to the ongoing but still inconclusive search for an infectious agent, the role of environmental factors in KD etiology has garnered increasing attention. Links between environmental triggers and vasculitides have long been suspected, with multiple studies exploring the connection for conditions like ANCA-associated vasculitis (AAV), rheumatoid arthritis (RA), and systemic vasculitis [40–44].

The environmental trigger case for KD is especially supported by the continuous rise in yearly cases for the past two decades as discussed in the previous section, with nearly a 5-fold change in incidence since the mid-1990s [45]. Apart from the aforementioned acute and synchronic waves across Japan, no infectious disease models match the long-term steady increases in KD incidence documented by the Japanese records either. The coherent increases in school-age children across most of the Japanese archipelago are, however, consistent with a wind-born exposure that might simultaneously affect a broad geographic region.

Rodó and colleagues, through a series of studies, have further suggested that an airborne agent, possibly transported by winds from northeastern China, could be a key factor triggering KD. This theory is supported by the consistent links they have found between tropospheric wind patterns arriving to Japan from NE China and increases in the incidence of the disease not in Japan, but also in Hawaii and San Diego [46,47,38,39], showing a potential atmospheric connection through the North Pacific. Later, a study by Jorquera and colleagues identified a similar mechanism, linking tropospheric winds originating from the Atacama Desert and reaching Santiago, Chile, with a rise in KD cases in the urban area. This pointed to wind-borne desert dust as a potential environmental trigger for the disease as well. [48].

Collectively, this evidence reinforces the idea of the arrival and dispersion of airborne agents via tropospheric winds may trigger KD across broad areas in the Japanese archipelago. These agents, capable of eliciting an autoinflammatory response, may account for the observed waves of KD, as well as the seasonality and increasing trends in Japan.

In Papers I and II, presented as chapters of this thesis, we explore the role of air masses and their contents as potential etiologic triggers of KD, contributing to a deeper understanding of the environmental factors involved.

1.3 Climate and Infectious Diseases (COVID-19)

1.3.1 General background

Climate and meteorological factors play a significant role in the epidemiology of infectious diseases. Environmental conditions such as temperature, humidity, and precipitation can influence the survival, reproduction, and transmission dynamics of pathogens, vectors, and hosts [49]. For vector-borne diseases like malaria and dengue fever, climate directly affects vector populations and pathogen development within vectors [50]. Similarly, for respiratory infections such as influenza and coronaviruses, climatic conditions can impact virus stability, transmission efficiency, and human susceptibility [51].

Understanding the interplay between climate and infectious diseases is crucial for predicting outbreaks, implementing timely interventions, and developing public health strategies. Seasonal patterns observed in many infectious diseases suggest that climatic factors can modulate transmission rates, leading to periodic increases or decreases in incidence [52]. However, disentangling the effects of climate from other factors—such as human behaviour, population immunity, and public health measures—remains a complex challenge.

1.3.2 Historical Context

Throughout the 20th century, several influenza pandemics highlighted the potential influence of climate on the spread of infectious diseases. The Spanish flu of 1918, Asian flu of 1957, and Hong Kong flu of 1968 each caused widespread morbidity and mortality, with patterns that suggested environmental factors played a role in their transmission [53–55]. Influenza viruses are known to exhibit seasonality, with higher transmission rates during colder months in temperate regions [56]. Studies have linked low temperature and low absolute humidity to increased influenza virus survival and transmission [57].

The 2009 H1N1 influenza pandemic provided further insights: early spread of the virus was shaped by climatic conditions, where low relative humidity and high population density favoured virus persistence [58–60]. These observations highlighted the importance of considering climatic factors in designing control and prevention measures for influenza pandemics.

1.3.3 The COVID-19 Pandemic and Climatic Drivers

In late 2019, a novel coronavirus, SARS-CoV-2, emerged in Wuhan, China, causing coronavirus disease 2019 (COVID-19). Declared a pandemic by the World Health Organization (WHO) in March 2020 [61], COVID-19 rapidly spread globally. As of September 2024, over 776 million confirmed cases and 7 million deaths have been reported worldwide [62], with significant impacts on public health, economies, and societies.

Unlike previous coronaviruses that caused relatively mild illnesses, SARS-CoV-2 exhibited efficient human-to-human transmission, including asymptomatic and pre-symptomatic spread [63]. Initial observations suggested that COVID-19 could exhibit seasonality similar to influenza. The early propagation of COVID-19 emerged in a latitudinal band between 30°N and 50°N, characterized by low humidity levels and temperatures between 5°C and 11°C [64]. Studies from China reported negative associations between absolute humidity and daily COVID-19 cases and deaths [65].

The potential role of climate in COVID-19 transmission was (and still is) a subject of intense research and debate. Climatic factors such as temperature, humidity, and UV radiation can influence virus stability in the environment, transmission efficiency, and host susceptibility [66]. Laboratory studies have shown that SARS-CoV-2 survives longer under cooler and less humid conditions [67]. However, the extent to which climate factors modulate COVID-19 transmission in real-world settings remained uncertain.

Determining the role of climate drivers in the transmission dynamics of emerging pathogens like SARS-CoV-2 is burdened with difficulties. Limited data availability, nonlinear responses, and multiple confounding factors complicate analyses, especially during the early stages of a pandemic. Factors such as variable intervention measures (e.g., lockdowns, mask mandates), changes in human behaviour, and the widespread susceptibility of populations can overshadow climatic influences [68,69].

Mathematical models addressing the potential for seasonality in COVID-19 dynamics suggested that, given the large number of susceptible individuals, climate forcing might not establish clear seasonal patterns initially, but as population immunity builds through infection and vaccination, climatic factors could become more influential in modulating transmission [70,71].

1.3.4 Environmental effects on aerosol transmission and seasonal dynamics

At a mechanistic level, the role of aerosols in SARS-CoV-2 transmission has significant implications for understanding climate effects. Aerosol transmission involves the spread of virus-laden particles small enough to remain suspended in the air over time and distance [72]. Environmental conditions like temperature and humidity can affect aerosol stability and virus viability [73].

Observations from the 2003 SARS outbreak raised the possibility of long-range aerosol transmission under specific environmental conditions [74]. Influenza studies have explored the connection between aerosol transmission, environmental persistence, and seasonality, suggesting that climatic factors can influence airborne transmission dynamics [75]. Similarities between influenza and SARS-CoV-2 in terms of environmental sensitivities highlight the need to consider climate in understanding COVID-19 transmission [76].

1.3.6 Context, research gap and significance

The exploration of climate and environmental drivers in infectious disease epidemiology extends beyond COVID-19. As discussed earlier, Kawasaki Disease (KD) exhibits seasonality and geographical patterns that suggest environmental triggers. The potential influence of atmospheric conditions, airborne particles, and microbial agents on KD incidence parallels investigations into climate effects on respiratory viruses and the seasonal dynamics observed in the disease.

Given the conflicting evidence up to that point and the critical need for clarity, our study, presented as PAPER III in this thesis, sought to (a) examine whether COVID-19 exhibited consistent seasonal patterns influenced by climatic factors such as temperature and humidity, (b) the found patterns where coherent across waves and regions and (c) develop methodological alternatives to be able to isolate the effects of climate on transmission dynamics given the nonlinearity of the responses and the potential confounders.

By addressing these objectives, we aimed to fill a critical gap in understanding the role of climate in COVID-19 transmission. This knowledge was (and still is) essential for forecasting future pandemic waves, informing public health strategies, and preparing for potential seasonal resurgences.

1.4 The Aerobiome

1.4.1 General background

The atmosphere is not merely a medium for weather patterns and gas exchanges; it is also a dynamic ecosystem crowded with life. The term aerobiome—the microbiome of the air—refers to the community of biological particles, including microorganisms and organic matter, that inhabit and circulate through the atmosphere [77]. This diverse environment comprises bacteria, fungi, viruses, pollen, spores, and fragments of plants and animals, all of which are transported through air currents and play vital roles in ecological and atmospheric processes [78].

The aerobiome is a complex, dynamic system influenced by both natural and anthropogenic sources. Microorganisms in the air are constantly dispersed from environments such as oceans, forests, soil, and even urban areas, through processes like wind erosion, sea spray, or human activities [79]. Once airborne, these particles can travel vast distances, crossing continents and oceans before settling in new environments. This long-range transport contributes to microbial connectivity between ecosystems, influencing biodiversity and ecological balance [80,81].

1.4.2 Impact on Human Health

The interaction between the aerobiome and human health is an area of growing interest: airborne microorganisms can affect human health both directly and indirectly. Direct impacts include respiratory infections, allergies, and exposure to pathogens, while indirect effects involve changes to environmental conditions that influence disease vectors and the spread of infectious diseases [82].

For instance, bioaerosols containing pathogenic microorganisms can lead to outbreaks of diseases such as influenza, tuberculosis, and Legionnaires' disease [83–86]. Allergens like pollen and fungal spores are well-known triggers of allergic reactions and asthma [87,88]. Moreover, the aerobiome can play a role in the dispersion of antimicrobial resistance genes, posing a significant threat to global health security [89].

In the context of KD, the potential influence of the aerobiome is particularly intriguing. As mentioned in the previous sections, the epidemiological patterns and geographical distribution of KD suggest that airborne agents could act as triggers for the disease [38,39,46]. Understanding

the aerobiome's composition and transport mechanisms is therefore essential for uncovering possible environmental factors involved in KD etiology.

1.4.3 Long-distance transport and pathogen dispersal

Microorganisms can be uplifted into the atmosphere and transported over long distances by atmospheric currents. This long-range transport occurs both within the planetary boundary layer (PBL) and in the free troposphere above it [90]. The PBL is the lowest part of the troposphere directly influenced by the Earth's surface, where vertical mixing is common [91]. Above the PBL, reduced friction and air density allow particles to travel thousands of kilometres before deposition [92].

While the presence of microorganisms in surface air is well-documented, with rich biodiversity and high concentrations of bacteria and fungi [93], less is known about microbial life in the free troposphere. The harsh conditions at high altitudes, such as low temperatures, reduced moisture and nutrients, and increased ultraviolet (UV) radiation, pose significant challenges to microbial survival. However, some extremophiles, like *Deinococcus radiodurans*, are known to withstand these conditions [94].

Studies have shown that viable microorganisms can be transported long distances attached to soil dust or organic aggregates [95]. For example, dust storms originating from the Sahara Desert have been linked to the transatlantic transport of bacteria and fungi to the Caribbean, affecting coral reef ecosystems and plant health [96,97]. Similarly, wind patterns have been associated with the dispersion of pathogens contributing to regional outbreaks of diseases such as meningococcal meningitis [98].

1.4.4 Research gap and significance of the study

Despite the understanding that some microorganisms can reach high altitudes, knowledge about microbial species richness and diversity in the free troposphere remains limited. Studies with quantitative data on the propagation of microorganisms in the free troposphere are scarce [99], with key questions remain unanswered: Can microorganisms remain viable during long-distance atmospheric transport in the harsh conditions of the troposphere? What is the extent of microbial species richness and diversity at high altitudes? How different is the surface-level aerobiome after deposition when compared to the samples collected over the PBL? Do these

airborne microorganisms pose a risk to human health upon deposition, potentially triggering diseases or spreading antimicrobial resistance?

In the section presented as PAPER IV of this thesis, we sought to answer some of these questions with a study involving the collection of air samples from the free troposphere using tropospheric aircraft surveys over Japan. By employing advanced DNA extraction, metagenomic sequencing techniques, and microbial culturing methods, we aimed to identify and characterize the microorganisms present at these altitudes. The findings of this study should allow us to fill critical gaps in knowledge about the aerobiome's composition and its implications for human health.

1.5 LIF Methods for microbial recognition

1.5.1 Challenges in characterizing the Aerobiome

Despite the recognized importance of airborne microorganisms in ecological processes and human health, the study of the aerobiome faces significant technical challenges. Historically, methods for characterizing airborne microbial communities have relied on culture-based approaches. However, these methods are limited because a substantial proportion of environmental microorganisms are non-culturable under laboratory conditions, leading to an incomplete understanding of the true diversity and composition of airborne microbial populations [100].

The advent of metagenomic next-generation sequencing (mNGS) revolutionized microbial detection in both clinical and environmental samples, offering culture-independent analysis with higher throughput, faster turnaround times, and greater taxonomic resolution [101]. mNGS enables the identification of a wide array of microorganisms by sequencing their genetic material directly from environmental samples. However, the ultralow biomass nature of air samples poses significant challenges for mNGS applications. Air samples often yield extremely low amounts of DNA, which can hinder both qualitative and quantitative analyses of microbial diversity. To accumulate sufficient material for analysis, researchers often need to sample over extended periods, compromising temporal resolution and the ability to detect rapid changes in microbial populations [102–104].

1.5.2 Fluorescence-Based Techniques for Bioaerosol Detection

To overcome these challenges, fluorescence-based techniques have been developed for the detection, monitoring, and identification of bioaerosols. Laser-Induced Fluorescence (LIF) is a particularly promising approach that enables real-time, in situ analysis of airborne biological particles without the need for sample collection and laboratory processing [105]. Initially driven by military applications for the detection of biological warfare agents [106], LIF technologies have expanded into various civil and scientific fields where timely bioaerosol surveillance is essential, such as environmental monitoring, public health, and allergen detection [107].

Several instruments using LIF for bioaerosol detection have been commercialized in recent years. Notable examples include the Wideband Integrated Bioaerosol Spectrometer (WIBS) series from Droplet Measurement Technologies (USA) [108], the KH-3000 from Yamatronics (Japan) [109], the BAA500 from Helmut Hund GmbH (Germany) [110], and the Rapid-E series from Plair SA (Switzerland) [111,112]. These instruments are primarily focused on the detection and classification of larger bioaerosol particles, such as pollen and fungal spores, due to their relatively strong fluorescence signals and larger sizes [113]. Maya-Manzano et al. conducted a comprehensive comparison and validation of these instruments for detecting and classifying pollen, using Hirst-type traps as a ground truth reference [114].

1.5.3 Limitations in the characterization of Bioaerosols

While the automation of pollen and spore particle counting and identification is valuable, extending these capabilities to the microbial fraction of bioaerosols is challenging. The relatively small size and weak fluorescence signals of microbial particles make them more difficult to detect and classify using LIF techniques, especially in uncontrolled, real-world environments. Additionally, microbial bioaerosols are often present in heterogeneous mixtures with non-biological particles, such as dust and pollutants, which can interfere with detection and analysis [115].

Studies have demonstrated that UV-LIF can discriminate between different types of bioaerosol particles when combined with advanced data analysis methods, including machine learning classifiers. For example, research has shown the potential to detect and discriminate even the smallerst viruses (picornaviruses) using LIF under laboratory conditions [116]. However, translating these capabilities into field-deployable devices suitable for real-time surveillance remains a significant challenge due to the need for high sensitivity, specificity, and robustness against environmental variability.

1.5.4 Addressing the gap: novel methodology

Recognizing these challenges, the study and prototype we present in PAPER V aims to develop and validate a novel methodology for the automated detection and classification of bacterial particles in bioaerosols using Laser-Induced Fluorescence. We use the Rapid-E real time airborne particle analyzer, developed by Plair SA (Switzerland), and modify existing equipment to enhance its capability to detect smaller microbial particles.

Our approach involves optimizing the instrument's detection settings, changing the internal equipment, developing specialized data processing algorithms, and employing machine learning techniques to improve classification accuracy. By conducting controlled laboratory experiments with known bacterial species, we aim to establish a proof of concept for real-time bacterial detection in bioaerosols. This methodology seeks to be suitable for deployment in both indoor and outdoor environments, providing valuable tools for environmental monitoring, public health surveillance, and early detection of airborne pathogens.

This study complements the broader themes of this thesis by extending our investigation into atmospheric factors influencing public health. While previous sections have focused on the role of environmental variables and the aerobiome in disease etiology, this work contributes by providing tools to monitor and characterize bioaerosols in real time. By bridging experimental methodology with atmospheric science, we aim to enhance our ability to detect potential airborne triggers of diseases and contribute to preventive public health strategies.

2. Objectives and significance

The overarching objective of this thesis is to explore and elucidate the complex relationships between atmospheric environmental factors and the onset of diseases, with a focus on KD but also using COVID-19 as an example of the infectious disease paradigm. To achieve this, the thesis is structured around the following specific objectives:

Investigate the association between the chemical fraction of air masses and the incidence of KD (Paper I)

Objective: To analyse the coherence between daily variability of fine aerosols and the incidence of KD.

Aim: To identify potential environmental triggers contributing to KD onset and understand the role of specific air pollution in the disease's epidemiology.

Examine the spatiotemporal dynamics of KD for clues on its etiology (Paper II)

Objective: To conduct an age-stratified spatiotemporal analysis of KD incidence in Japan, focusing on temporal shifts on it seasonal signal.

Aim: To deepen insights into the epidemiological patterns of KD, particularly the shifts in seasonality across different age groups, and explore potential changes in environmental triggers.

Assessment of the influence of climate on the transmission dynamics of COVID-19 (Paper III)

Objective*:* To investigate the climatic signatures in the different COVID-19 pandemic waves across both hemispheres, focusing on the effects of temperature and absolute humidity on disease transmission.

Aim*:* To determine the role of climatic factors in COVID-19 transmission dynamics, classify the disease's seasonality, and inform public health strategies for controlling respiratory viruses.

Evaluation of the diversity and transport of microorganisms in the troposphere

Objective: To characterize the microbial richness and air chemistry in aerosols above the planetary boundary layer, assessing whether viable human pathogens and antimicrobial-resistant bacteria can be found at such altitudes and survive long-distance transport.

Aim: To understand the potential role of atmospheric microbial communities in disease dispersion and highlight implications for human health and disease etiology.

Development of novel methodology for real-time characterization of bioaerosols

Objective: To benchmark a prototype based on LIF and ML models for the real-time identification and classification of bacteria in bioaerosols.
Aim: To advance capabilities for rapid and precise monitoring of airborne microbial communities, providing innovative tools for ecological studies and public health surveillance.

By fulfilling these objectives, this thesis aims to:

Advance understanding of environmental determinants of disease onset
Provide new insights into how air pollution, climatic conditions, and atmospheric microbial diversity influence the incidence and progression of KD and COVID-19, enhancing our understanding of their environmental determinants.

Inform public health strategies and policies
Offer evidence-based recommendations for disease prevention and control, contributing to public health policies aimed at mitigating the impacts of air pollution, climate variability, and bioaerosol exposure on human health.

Innovate methodologies for atmospheric sciences and epidemiology
Develop and validate novel approaches for atmospheric characterization and bioaerosol detection, enhancing the integration of atmospheric science and epidemiology, and fostering interdisciplinary collaboration.

Lay the groundwork for future research
Establish a foundation for ongoing research that bridges atmospheric science and epidemiology, encouraging further exploration into the environmental factors affecting human health and supporting the development of early-warning systems for disease prevention based on the usage of not so common climatic and atmospheric variables.

3. Publications

The main piece of work presented in this thesis comprises five manuscripts, included as the main chapters. Three of these (Paper I, Paper III, and Paper IV) have been peer-reviewed and published in Q1 journals in their respective fields. The other two (Paper II and Paper V) are currently manuscripts in preparation and will be submitted for publication in the coming months.

Before each publication, a page is provided with the following details: the publication title, authors, author affiliations, full journal reference, DOI, and the impact factor of the journal. The full version of each published paper, as distributed by the journal, is attached as a chapter. Supplementary Information is not included for the published papers but is available online through the respective DOI. For unpublished manuscripts, Supplementary Information is attached directly after the main text.

PAPER I

Sub-weekly signatures relate ultrafine aerosols enriched in metals from intensive farming and urban pollution to Kawasaki disease

Xavier Rodó[6,7,1,2,3], Albert Navarro-Gallinad[6,4], Tomoko Kojima[5], Josep-Anton Morguí[2], Silvia Borràs[2] and **Alejandro Fontal**[2]

Author affiliations

1 ICREA, Barcelona, Catalonia, Spain

2 Climate and Health (CLIMA) Program, IS Global, Barcelona, Catalonia, Spain

3 Osaka University, Graduate School of Human Sciences, Osaka, Japan

4 ADAPT Centre for Digital Content, Trinity College Dublin, Dublin, Ireland

5 Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto, Japan

Author notes

[6] These authors contributed equally to this study.

[7] Author to whom any correspondence should be addressed.

Published peer-reviewed article:

Environmental Research Letters, Volume 18, Number 7

DOI: 10.1088/1748-9326/acd798

IF: 5.8

LETTER • OPEN ACCESS

Sub-weekly signatures relate ultrafine aerosols enriched in metals from intensive farming and urban pollution to Kawasaki disease

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Sub-weekly signatures relate ultrafine aerosols enriched in metals from intensive farming and urban pollution to Kawasaki disease

Xavier Rodó[1,2,3,6,*], **Albert Navarro-Gallinad**[4,6], **Tomoko Kojima**[5], **Josep-Anton Morguí**[2], **Silvia Borràs**[2] and **Alejandro Fontal**[2]

[1] ICREA, Barcelona, Catalonia, Spain
[2] Climate and Health (CLIMA) Program, IS Global, Barcelona, Catalonia, Spain
[3] Osaka University, Graduate School of Human Sciences, Osaka, Japan
[4] ADAPT Centre for Digital Content, Trinity College Dublin, Dublin, Ireland
[5] Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto, Japan
[6] These authors contributed equally to this study.
[*] Author to whom any correspondence should be addressed.

E-mail: xavier.rodo@isglobal.org

Keywords: aerosols, human health, Kawasaki disease, ultrafine particles, agricultural by products, wind intrusions, transient signatures

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
8 February 2023

REVISED
8 May 2023

ACCEPTED FOR PUBLICATION
22 May 2023

PUBLISHED
19 June 2023

Abstract

Air pollution (urban, industrial or rural) has been linked to a myriad of human ailments despite clear mechanistic associations that are often not thoroughly established. Daily variability of fine aerosols in a surveillance campaign in south Japan shows a striking coevolution between their trace elements (metal and metalloid, MM) content and Kawasaki disease (KD) admissions, suggesting a strong dynamical link. These aerosol MM could instigate an immune response that, along with genetic susceptibility, would lead to KD development. This association may account for over 40% of the total variability in the disease, being dominated by a clear sub-weekly cycle (SWC_1). Thanks to both an unprecedented daily KD epidemiological record going back to 1970, light detection and ranging (LIDAR) atmospheric backscattering profiles for the interval 2010–2016 and HYSPLIT simulations with numerous sensitivity analyses, we can trace this SWC_1 variability to occur concomitantly from sub-seasonal to interannual timescales in both KD and aerosols. This SWC_1 appears to connect or disconnect Japan to air intrusions from above the planetary boundary layer (PBL), having their source in industrial and agricultural areas in NE Asia and points to a stronger case for an agricultural source for the exposure as opposed to urban pollution. The KD maxima always occur in full synchrony with the arrival of very small (<1 μm; PM_1) particles showing that ultrafine aerosols appear as a necessary cofactor in the occurrence of KD and sets the field to associate other similar human diseases. Our study shows how signal-detection approaches can be useful to uncover hidden associations between the environment and human health, otherwise unnoticed and help set new early-warning systems for disease prevention.

1. Introduction

The causes conducive to Kawasaki disease (KD) pediatric syndrome are yet unknown, despite the intensive research on different fronts (genetic, immunological, environmental, epidemiological, etc Marrani *et al* 2018, Nagata 2019). Potential drivers include environmental, biological or chemical triggers (e.g. bacteria, fungi, viruses, toxins, dust, pollution, Leung *et al* 1993, Barton *et al* 2002, Matsubara *et al* 2006, Lee *et al* 2011). Unfortunately, no conclusive result has yet unequivocally solved this long-standing puzzle. On the etiology, the leading paradigm is that an unidentified agent enters through the upper respiratory tract and causes a dramatic immunologic response in certain genetically predisposed children (Rowley *et al* 2008, Onouchi 2009, Rodó *et al* 2014). Coronary artery aneurysms develop in 20%–25% of cases, with this outcome being at times clinically silent and often misdiagnosed with other more benign rash

and fever syndromes, but that may lead, in rare occasions, years later, to sudden death or myocardial infarction (Burns and Glodé 2004). The presumed link in KD epidemiology to an environmental forcing factor is supported by a clear seasonal pattern, first described for Japan (Burns *et al* 2005) and then worldwide (Burns *et al* 2013). Increases in KD cases in the Northern Hemisphere winter in locations on both sides of the North Pacific Ocean were seen to occur associated with the seasonal enhancement of low- and high-tropospheric wind currents from the Asian continent (Rodó *et al* 2011). The ulterior identification of a fungal pathogen predominant in air masses over the planetary boundary layer (PBL) at times of KD maxima, led to the hypothesis that in Japan, KD could be promoted by a pathogen or environmental trigger carried by winds blowing from agricultural areas in the Asian continent (Rodó *et al* 2011, 2014, Frazer 2012, Maki *et al* 2019). This seminal study identified an area of intensive cropland production in NE China, known as the 'breadbasket of China' as the center of the source region of winds arriving in Japan a few hours to two days later.

At a time when PM attributed to automobiles and city origins have decreased in general, because of efforts to reduce air pollution, we are nonetheless still seeing large numbers of KD cases in Japan. If the source originated from urban pollution, we would expect the number of cases to lower, which is not the case. Instead, because of the growing resistance to pesticides among agricultural pests, pesticides may have become more potent and even more toxic, strengthening the hypothesis for an agricultural origin. In China (as in other regions), since the 1960s, the adoption of high-yield crop varieties and wide use of artificial fertilizers and pesticides has increased production markedly (Green and Grow 2020). These agricultural practices accumulate high levels of trace elements (in particular MM) in soils, affecting key biological processes with spoiling consequences on the quality of both soil and plant health. While in adequate levels metals, such as Fe, Mn, Mo, Cu, Zn, and Ni are necessary for plant growth, high concentrations of these MM can result in toxicity for humans (Peralta-Videa *et al* 2009). Phosphate fertilizers, for instance, are hazardous to the exposed rural communities, which may be a causative factor of chronic kidney disease (Jayasumana *et al* 2015). These phosphate rocks used in manufacturing fertilizers may contain micronutrients necessary for plant growth (such as magnesium (Mg), manganese (Mn), potassium (K), sodium (Na) and cobalt (Co)), but also environmental pollutants such as cadmium (Cd), copper (Cu), chromium (Cr), nickel (Ni), lead (Pb) and zinc (Zn, Tufail and Khalid 2008). Those MM are one of the major environmental problems around the world. MM constitutes an ill-defined group of inorganic elements, including transition metals, metalloids, lanthanides, and actinides. The most common MMs found at soil sites are Pb, Cr, arsenic (As), Zn, Cd, Cu and Ni. It is well-known that most metals may persist in soils for a long time after their introduction, resisting microbial or chemical mobilization. As for their inhalation from aerosols, data on the chemical composition of air masses and source transport processes are usually not available and, therefore, relationships to human health are extremely hard to substantiate.

2. Methods and datasets

When associating meteorological and air quality factors to health outcomes, the transient nature of such relations should be considered (Bragazza 2008, Andersen *et al* 2009, Lindegren and Eero 2013). We used statistical techniques aimed at characterizing these couplings, often involving thresholds that have proven effective. For instance, the singular spectrum analysis (SSA, Dettinger *et al* 1995) and the scale-dependent correlation (SDC) analysis (Rodriguez-Arias and Rodó 2004), have successfully been applied to relate climate with cholera (Rodó *et al* 2002), malaria (Rodó *et al* 2021), KD (Rodó *et al* 2014), and more recently, also to flu and COVID-19 (Fontal *et al* 2021).

Air pollution from Asian dust and aerosols are routinely monitored at around 20 locations in East Asia at two-wavelengths (1064 nm and 532 nm) using polarization-sensitive LIDARS. The data from the Asian Dust and Aerosol Lidar Observation Network (AD-Net) are processed and published in near real time (Nishizawa *et al* 2016). Our study uses this rich database of vertical profiles that contains high resolution (<10 m) optical measurements of the atmospheric column. Measurements can trace the changing structure and physics of the air column and of particles. As no routine detailed data exists for MM in air quality stations other than a few metals, we used a unique daily survey of aerosol's chemical composition collected in Kumamoto (Japan) for 37 consecutive days in spring 2011 covering nearly 60 major and trace elements (mostly MM) determined from air filter samples (see Moreno *et al* 2013 for a full description of the methodology, figure 1).

SSA was then used to separate the variability components of KD and the total amount of MM (figures 1(B) and 2(D); see Methods).

As in previous studies, the origin of these aerosols was traced to the NE Asia region with the use of Lagrangian particle model simulations. To this end, we used a similar approach to seek for the origins and trajectories of the air masses in this particular event. With the HYSPLIT particle model Version 5.2.0; Stein *et al* (2015) (downloaded from www.arl.noaa.gov/hysplit/), we compared all air trajectories and emerging sources between peak KD days and lowest KD days (see Methods and supplementary information).

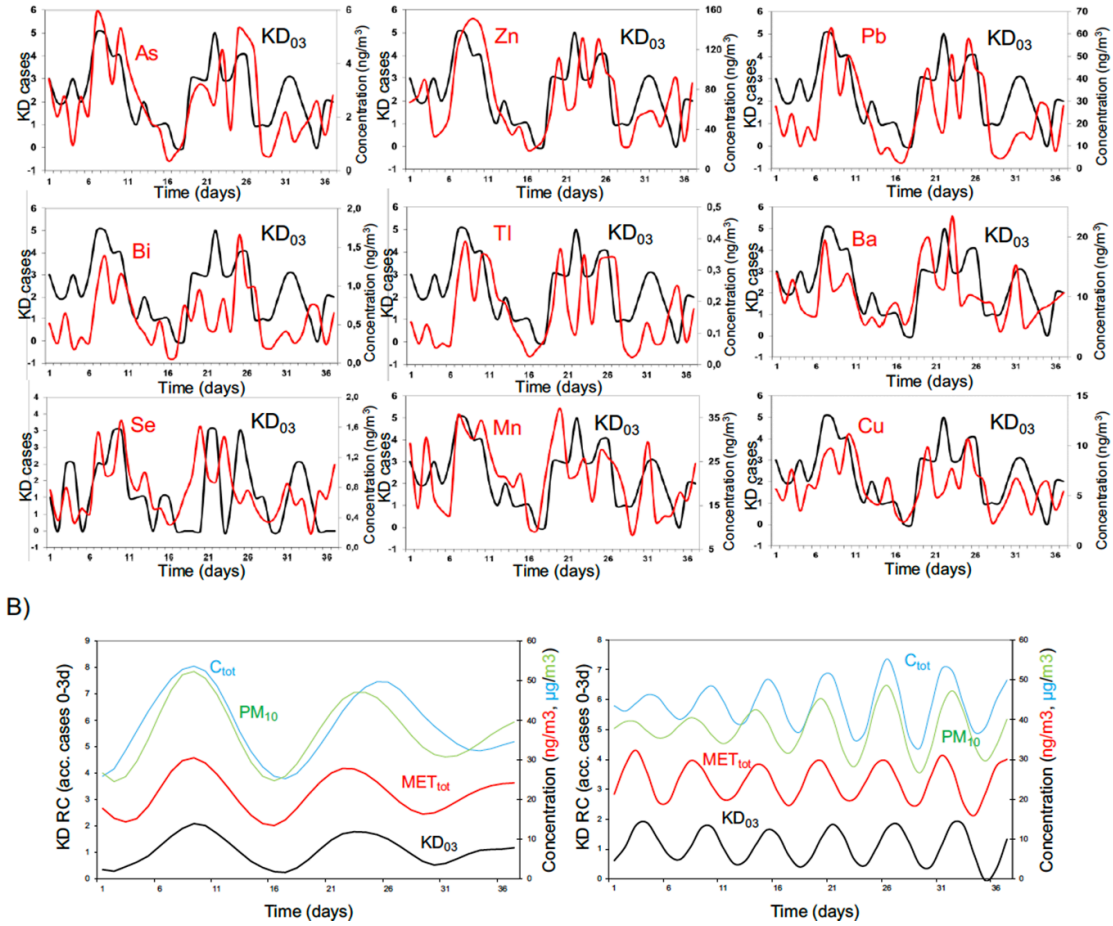


Figure 1. Aerosol's metal concentrations and synchrony with the KD incidence in Kumamoto (Japan, see methods). (A) Coevolution between the most correlated individual MM (red) contained in aerosols and the number of accumulated KD03 cases (integrated from 0 to 3 d (black) in a sequence of 37 d) in Spring 2011(8). (B) SSA reconstructed variability at the bi-weekly timescale (left, see methods and 19) in the case of total MM (red), total carbon (blue), PM_{10} (green) and KD (black). Right plot displays the same as the left one but for the weekly reconstructed components. Note consistencies and differences among phases and convoluting amplitudes of the reconstructed signatures in the different variables. Embedding dimensions for the SSA analysis were always varied to test for signal stability ($p < 0.05$). Concentration is in ng m^{-3} for MET_{tot} and in μg m^{-3} for C_t and PM_{10} but were shown in the same graph for comparison. As KD03 has accumulated cases from day 1 to day 3, no intra-weekly variability is present.

3. Results

Figure 1(A) displays a clear temporal covariation between MM measurements in the 37 days interval and the evolution of KD-admitted children in the group of Kumamoto prefecture hospitals (supplementary figure 1). The chemical results for the main elements analyzed are in full agreement with measurements routinely obtained in that period at 33 air quality monitoring stations in urban, suburban and rural areas operated by the Kumamoto Prefectural Government (see also www.pref.kumamoto.jp). The most abundant MMs were Zn, Pb, barium (Ba) and Cu, despite a very similar evolution which also manifests for many other trace elements in aerosols pointing to the same behavior. We used a cluster analysis to group the MM along ten different clusters (supplementary figures 3 and 2(A), (B)), with cluster 1 selecting 38 MM (figures 1(A) and (B) red curves). When compared to the aggregated KD evolution in the same interval (e.g. with KD_{03} being KD cases added from day 0 to day 3 as the response time is established to vary between hours and 3 d, see Methods and Rodó *et al* 2014), a remarkable degree of covariation shows up (figure 2(B)). A linear dose-response relationship arises in this interval between the new KD cases and total MM' concentration (e.g. with bootstrapped low-high μ difference being significant, $p = 0.0001$, figure 2(C)). A Poisson linear regression model based on total metal concentrations accounts for over 40% of the variability in KD (figure 2(D) showing a yellow line with 95% confidence intervals; $p = 4.35^{*}10^{-4}$; see Methods). As an example, figure 1(A) displays the highest individual coevolution of KD_{03} with eight of these MM species (e.g. As, Zn, Pb, Bi, Tl, Ba, Mn, Cu) and Se. The

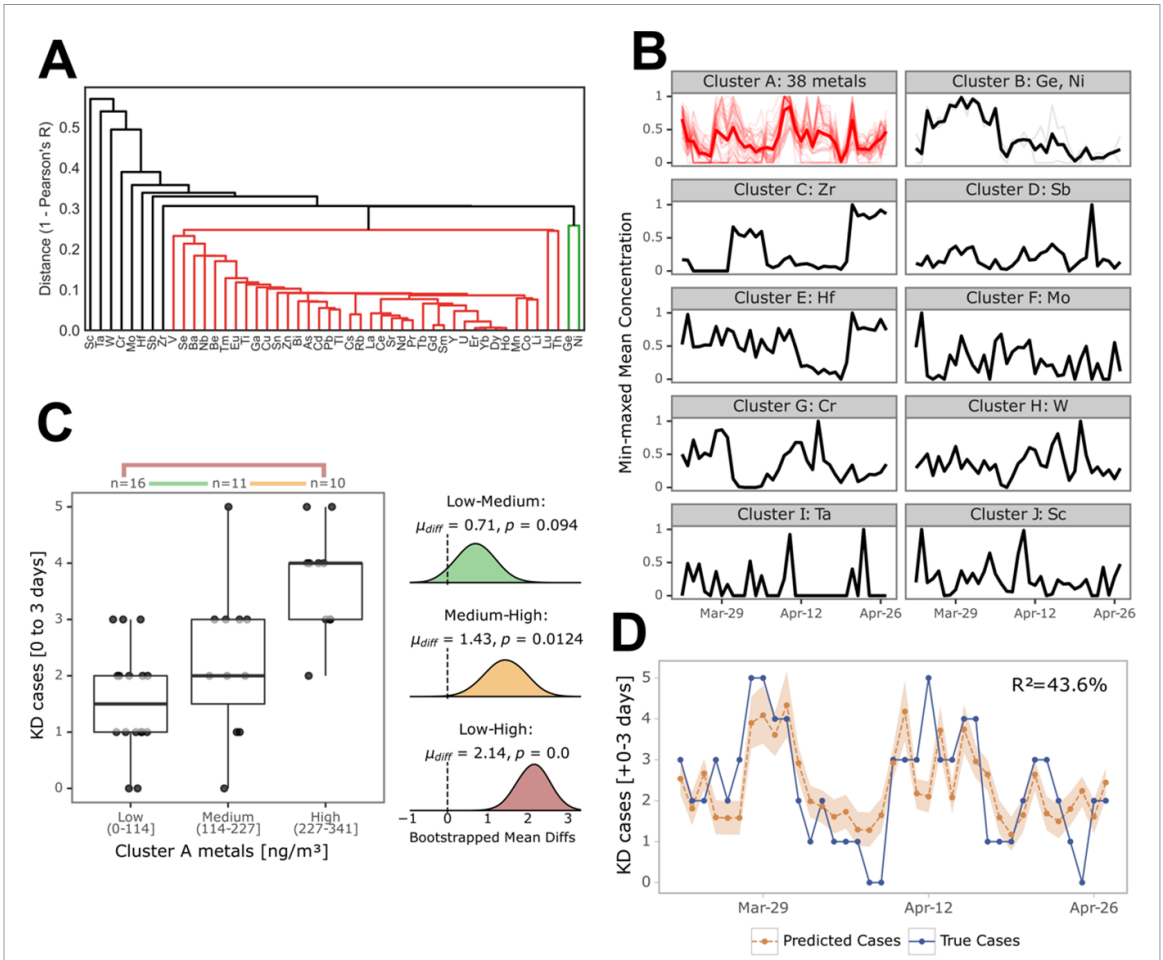


Figure 2. Cluster analysis of the metal content in aerosols and its synchrony with KD incidence in Kumamoto (A) Dendrogram generated when using Pearson's correlation as distance metric and the nearest point algorithm to compute distances across the newly formed clusters. Colors denote the different clusters assigned when setting a distance threshold of 0.3. (B) Variation during the 37 sampled dates of the (min-max normalized) concentration of MM of each cluster. For clusters with more than a single member, the thick line represents the median value of all cluster members at each time-point, with the individual metal contributions shown in the shaded thin lines. (C) Boxplots for the distribution of daily KD cases in the Kumamoto prefecture as a function of the daily concentration of Cluster A MM are categorized into three groups: low (0–114.00 ng m^{-3}, $n = 16$), medium (114.01–227.00 ng m^{-3}, $n = 11$) and high (227.01–341.00 ng m^{-3}, $n = 10$). On the other hand, the distribution of the difference of means for the pairwise comparison of groups performing a bootstrap test, showing significant differences in KD cases at an alpha =0.05 for days with high concentrations compared to those, in both medium, and low concentrations. (D) Reported (true) KD cases in the Kumamoto prefecture (blue) and predicted cases with a Poisson model using the concentration of Cluster A MM as predictor. The shaded area represents a bootstrapped 95% CI.

percentage of variance accounted for by individual MM in this 37 days atmospheric event reaches around 50% on average, with values nearing 60% in the case of Zn (figure 1(A)). The statistical model also denotes how an increment of around 70 ng m^{-3} in the concentration of total MM is associated with the appearance of one new KD case.

The SSA yielded a clear separation of this variability into four pairs of significant components, with two pairs showing a 15 d periodicity that varied fully in line with the oscillations reconstructed for KD_{03} (figure 1(B)). Two other paired components recovered a marked weekly cycle (WC), being the 15 d cycle super-harmonic of the latter. Strikingly, coarse particulate matter (denoted by PM_{10}) and total carbon content (C_{tot}) also display the same two periods above, albeit they both peak with slight delay in the latter part of the survey (e.g. 1 d after) and manifest a clearly different convoluting amplitude. Both C_{tot} and $PM_{10,}$ in the coarser aerosol fraction, are therefore not fully concurrent with neither MM (heareafter MET_{tot}) nor KD_{03}, whereas instead, figure 1(B) shows how the same evolving amplitudes of the reconstructed components (RCs) of MET_{tot} and KD_{03} (with at times a one-day lead for the former).

To obtain regional source estimates, we ran from Kumamoto city, 4 d back trajectories four times every day (at 00:00, 6:00, 12:00 and 18:00 h locally), for each event with either 5 cases (maxima) or days with 0 cases (minima) (figure 3(A)). A display of the trajectories inferred for all 37 d is included in supplementary figure 2. For comparison with former results (generated up to 2010), we also ran similar back trajectories for all extreme 5% percentile days in the interval 2010–2016 (figure 3(B)). Figures 3(A) and (B) show very similar patterns, with maxima predominantly in

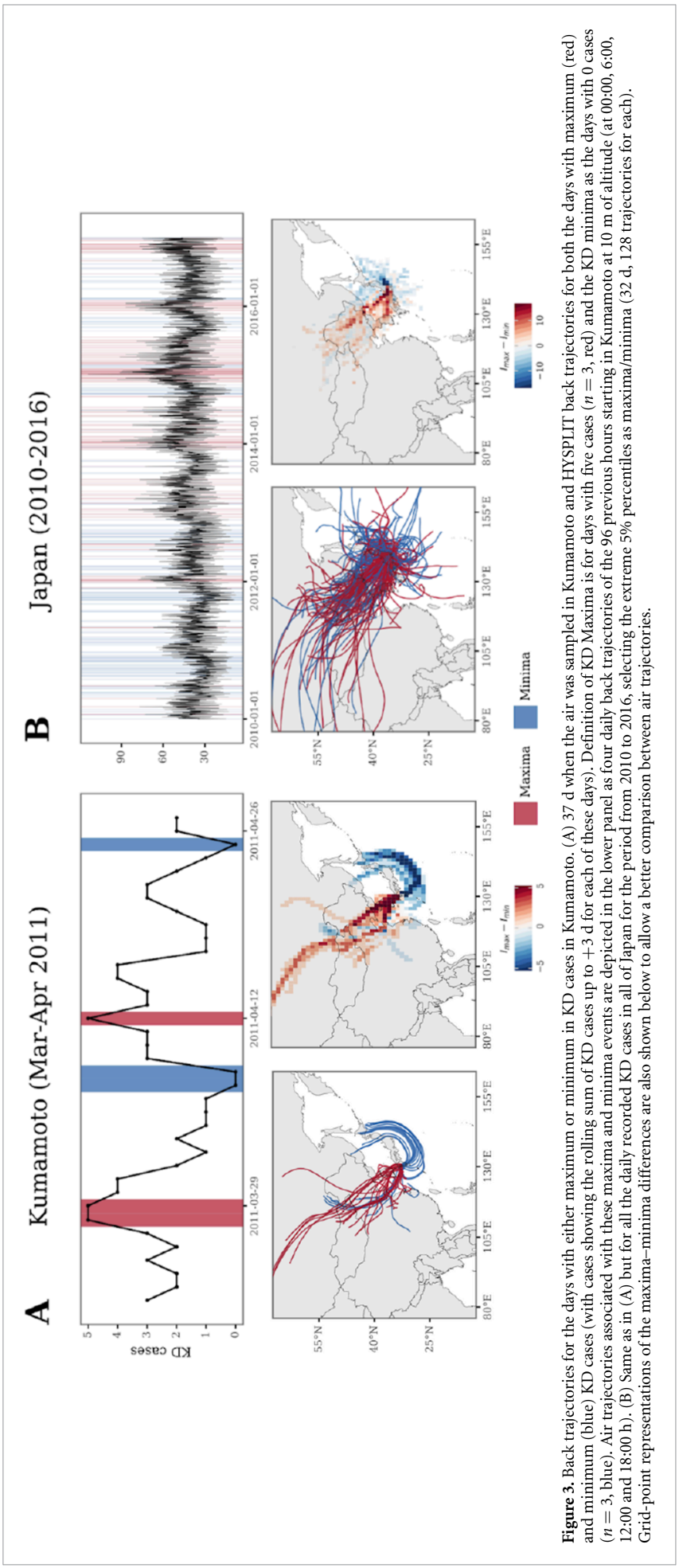


Figure 3. Back trajectories for the days with either maximum or minimum in KD cases in Kumamoto. (A) 37 d when the air was sampled in Kumamoto and HYSPLIT back trajectories for both the days with maximum (red) and minimum (blue) KD cases (with cases showing the rolling sum of KD cases up to +3 d for each of these days). Definition of KD Maxima is for days with five cases ($n = 3$, red) and the KD minima as the days with 0 cases ($n = 3$, blue). Air trajectories associated with these maxima and minima events are depicted in the lower panel as four daily back trajectories of the 96 previous hours starting in Kumamoto at 10 m of altitude (at 00:00, 6:00, 12:00 and 18:00 h). (B) Same as in (A) but for all the daily recorded KD cases in all of Japan for the period from 2010 to 2016, selecting the extreme 5% percentiles as maxima/minima (32 d, 128 trajectories for each). Grid-point representations of the maxima–minima differences are also shown below to allow a better comparison between air trajectories.

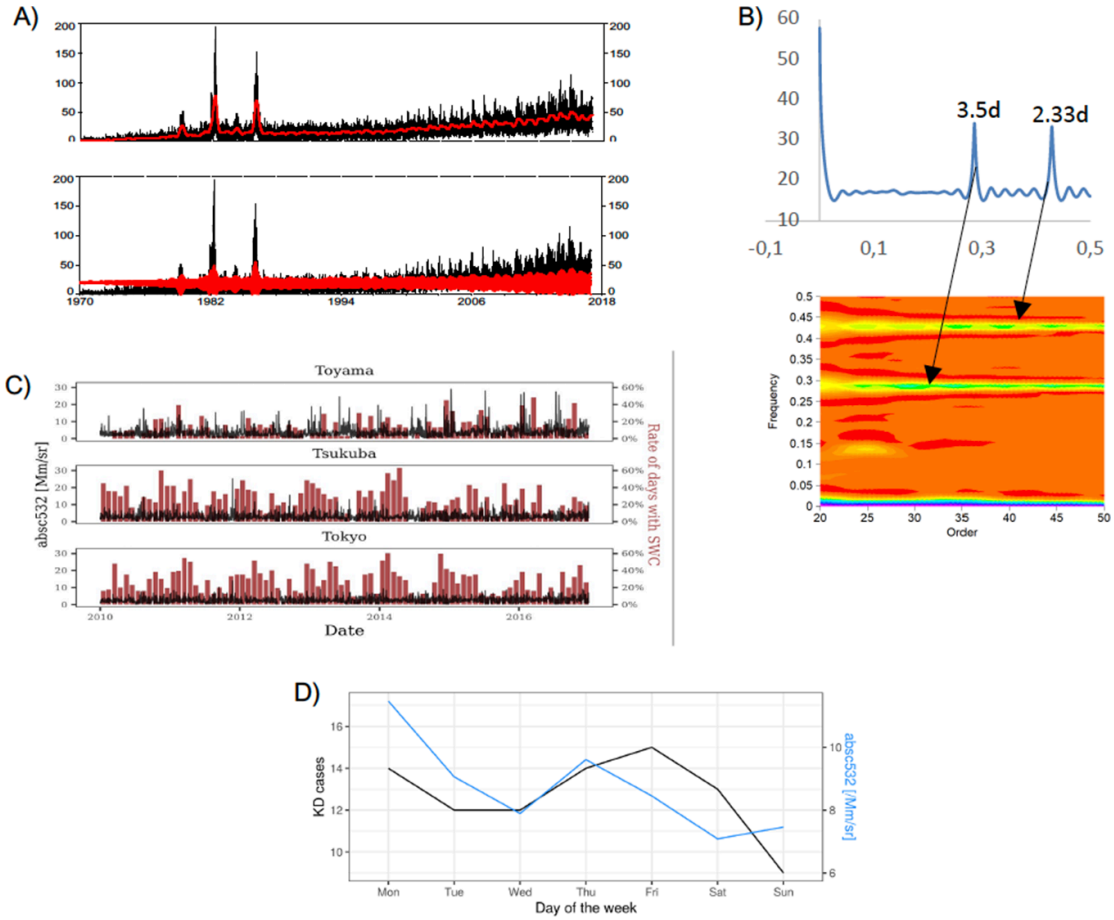


Figure 4. The long-term evolution of Kawasaki disease (KD) incidence for the entire of Japan (1970–2016) and contribution of the constituents of the weekly cycle. (A) Top: aggregated KD incidence for all of Japan (black) and reconstructed variability of the seasonality (red, corresponding to RC_{12} of the SSA decomposition; see Methods for spectral decomposition and reconstruction (19). Bottom: same as before but for the weekly cycle variability. (B). Top panel corresponds to the AR spectrum of the KD time series with FB (forward–backward) singular value decomposition algorithms. Bottom panel displays the constituent sub-weekly frequencies obtained (namely 3.5 d SWC_1 and 2.33 d SWC_2, respectively), and an exploration of different orders in the decomposition for stability of results. (C) Temporal evolution of the absc532 coefficient and the SWC in the three stations. Bars display the monthly percentage of days with 'SWC presence' (with an $r > .4$), while the black line displays the daily readings of the absorbance coefficient at 532 nm for each station. (D) Co-evolution of the weekly cycle in both the SWC of aerosols and KD in Tokyo for the interval 2010–2016.

the NE China region as previously identified in Rodó *et al* (2014), but with the Kumamoto event showing the source location also slightly southern with respect to the 2010 results. A detailed sensitivity analysis on this back tracking of the source has also been conducted as well as an estimation of the associated errors and uncertainties (see Methods and supplementary figures 10–12).

The rich record of epidemiological incidence of KD in Japan going back to the 1970s enables a very detailed and unprecedented statistical approximation to the long-term evolution and structure of the WC shown above for the Kumamoto event (Rodó *et al* 2011, Moreno *et al* 2013). As cases were there grouped from day 0 to day 3, inspection of intra-weekly variation is not possible in this short record. Instead, this motivated the analysis of the epidemiological records of daily KD admissions for the entirety of Japan (figure 4) and for nine regions grouping prefectures (*not shown*). Applying an SSA decomposition ($M = 100$; see Methods and Ghil *et al* 2002) to study temporal variability, a prominent SWC along with marked seasonality shows up as the dominant variability components (figure 4(A)). An autoregressive (AR) spectrum with a backward-forward singular value decomposition algorithm was computed for KD (figure 4(B)) to further determine the sources of spectral variability and assess the nature of these weekly oscillations. First two RC recovered seasonality (RC_{12}, 70% of the variability, $p < 10^{-5}$), and the WC showed up strongly in RC_{34} with the latter accounting altogether for a constant 20% of the remaining variance (figure 4(A)). This SWC is composed of two clearly distinguishable sub-weekly periods (figure 4(B)) and not period 7 d, with the fundamental one being 3.5 d ($SWC_1 > 10.6\%$; $p < 10^{-5}$), and its harmonic at 2.33d ($SWC_2 = 9.2\%$; $p < 10^{-5}$, see Methods and supplementary figure 4).

We conducted a two-way SDC (TW-SDC) analysis (Rodó 2001, Rodó and Rodriguez-Arias 2006)

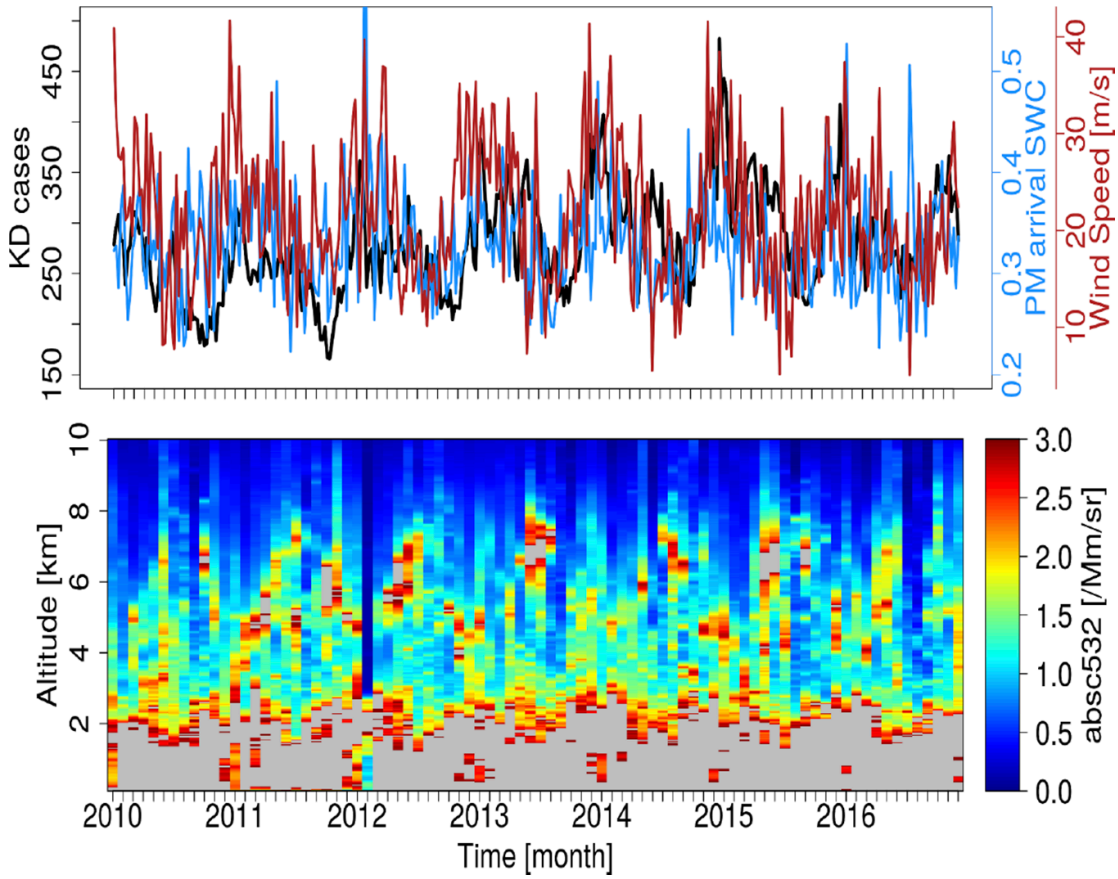


Figure 5. Association of PM arrival (absc532) to KD cases in Japan at a weekly scale. (A) Weekly association of KD in Japan (black) with PM arrival with the SWC_1 dynamics in Tokyo (blue) and wind speed (red) for the 2010–2016 period. The maxima synchronization between the PM arrival and wind speed emphasizes the external source for the aerosols. (B) Attenuated backscatter coefficient at 532 nm (absc532) vertical profile displaying the aerosol intrusions during the winter months when KD, PM arrival and wind speed have a peak. The absc532 scale has been adjusted to highlight the vertical transport from high altitudes (circa 6 km) towards the surface, which tends to be between 0–3 [Mm^{-1} s^{-1}]. The absc532 values below 2 km are of higher magnitude (∼4–8 [Mm^{-1} s^{-1}]), greyed in this figure.

to further confirm the base frequencies present in KD daily records. This technique is specifically aimed at uncovering transitory associations between joint variability structures in time series (see Methods and Fontal *et al* 2021). We used the daily series of KD in the recent interval for which LIDAR atmospheric profiles are available (from 2010 in supplementary figure 4, and January 2010—December 2016 in figures 4 and 5) and built a synthetic time series of 3.5 d (SWC_1), 2.33 d (SWC_2) and 7 d (SWC_3) periods to seek for similar signatures in KD (supplementary figure 3 and see Methods). As an example, supplementary figure 4 displays for the year 2010, the TW-SDC results for SWC_1 (3.5 d period; panels a, d), SWC_2 (2.33 d period; b, e) and SWC_3 (7 d period; c, f) at different window sizes (S) to capture the periodicities whenever they appear (namely $S = 6$ d in *a, b, c* and $S = 22$ d in *d, e, f*). The 3.5 d cycle (SWC_1) stands out strongly, both in terms of its high correlation magnitude as well as for being present throughout the interval 2010–2016 (e.g. compare *a vs c, d vs f* in supplementary figure 4). In contrast, a very faint SWC_3 (7 d) that is linked to the 3.5 d cycle, correlates mostly to periods of very low KD incidence (supplementary figures 4(c) and (f)). A sensitivity analysis with an intermediate S of 10 -to maximize a potential 7 d cycle- applied to the entire 2010–2016 interval, similarly states the dominant role of SWC_1 in KD (*not shown*). Summing up, two maxima and two minima clearly appear in a week (figure 4(D)), with this SWC_1 also showing up strongly in all the three large epidemics (e.g. see its maximum in supplementary figure 5 for the largest KD epidemic in 1982 and a marked contribution in 1979 and 1986). SWC_1 exists throughout the entire KD record since the surveillance was initiated in the early 1970s in Japan (and until the end of 2016). The SWC_1 cycle amplifies towards the present concomitantly with the appearance of a more marked seasonality in the epidemiological time series (supplementary figure 5).

We further checked whether bias-reporting could have generated those persistent periodicities. One-way SDC analysis of daily KD per prefecture were grouped annually for the entirety of Japan, resulting in strong positive correlations close to the diagonal that we summed up (supplementary figure 5, red line;

see Methods). The results show a clear amplification of the SWC_1 towards the present, with a rising trend and also an intriguing 4 year oscillation. The total yearly aggregated KD incidence for Japan is here shown for comparison (supplementary figure 5, black line), showing a steeper upward trend towards the present. These long-term patterns are incompatible with a mere weekend bias in reporting, which would have only had a dominant presence at those very short time scales of intra-weekly variation. Supplementary figure 6 (supplementary figure 7) displays the different distribution of KD occurrences (significances) throughout the week, when reported data is considered as either days of admission (DoA) or days of onset (DoO). DoO is, as a consensus, estimated by the medical community to be 5 d before the day of admission. However, our study shows a strong median value centered around 2–3 d. Furthermore, the results reveal some slight discrepancies between the WC when considered from DoA (supplementary figure 6(A)) and DoO (supplementary figure 6(B)). When split in years, the patterns found are quite consistent within the two categories, displaying the aforementioned Sunday and Wednesday–Thursday minima and the Monday–Friday maxima in DoA (compare supplementary figures 6(A) and 4(D)). When DoO is analyzed, a similar SWC emerges, despite having different magnitudes and a much weaker Sunday minimum. DoO is, however, a much more volatile value due to its subjective nature, albeit consistentcy in the appearance of the SWC_1 in both cases provides further strength to the real nature of this type of variability in KD.

As a comparison, similar representations -both of the WCs and the fundamental frequencies- were derived for other concurrent diseases in Tokyo (namely heat strokes, HST; supplementary figures 8(A), (C) and 9(A)) and influenza A; supplementary figures 8(B), (D) and 9(B)). Additionally, a number of environmental covariates were also tested for comparison (namely, temperature, absolute humidity, AQ, PM_{10}, NO_x and ozone; supplementary figure 9(C). None of the alternate diseases nor environmental covariates displayed a similar SWC signature to KD (supplementary figure 9).

The rate of daily presence of the SWC_1 in the LIDAR aerosols datasets was similarly extracted with the aid of SDC and is displayed in figure 4(C) for Toyama, Tsukuba and Tokyo (red bars), aligned with the base value measured from the LIDAR (absc532 as black series, serving as an analogue of the fine aerosol fraction; see Methods). These three locations are unique in that they all have routine LIDAR instruments being operated from 2010–2016 (www-lidar.nies.go.jp/AD-Net/index.html), as well as long-term KD epidemiological records. The SWC_1 presence in aerosols is increasing as we move closer to the air intrusion vertical coordinates denoted by synoptic-scale meteorology. The SWC_1 is roughly constant throughout the record, but displays large oscillations in its relative contribution, ranging from near 0 to around 60% and with an increasing value as we approach Tokyo that is consistent with atmospheric deposition. Clear seasonal oscillations also appear, but with maxima centered in the winter and early spring (figure 4(C)).

The two WCs are overlaid for Tokyo in figure 4(D) (blue line for columnar LIDAR up to 6 km and black for KD new cases) and a clear co-variation and slight anticipation consistently manifests. Expansion of this same analysis to cover all the available LIDAR data period (2010–2016) also displays high agreement when compared to KD new cases grouped for all Japan (supplementary figure 10(A)). The LIDAR data is again an average of the atmospheric layers from near-surface up to 6 km high and a striking and consistent 1 month lead time in both maxima and minima appears between this fraction of aerosols and KD new cases. The amount of these aerosols in particulate matter (PM) clearly grows during events associated with KD maxima (supplementary figure 10(A)) as a striking monthly positive co-evolution exists between the SWC_1-PM_1 cycle (blue) and KD incidence (black). The presence of the 3.5 d cycle shortly anticipates KD (supplementary figure 10(A)). Synchronous annual cycles show up in both variables, with coincident maxima in the winter and minima in the summer, as expected. This large seasonal coherence among wind and SWC_1-PM_1 further appears to reinforce the central role of aerosols in this disease. Again, source-tracking errors and sensitivities were further assessed, both by increasing and reassorting the number and type of HYSPLIT simulations conducted (see Methods and supplementary figures 11–13). Further confirmation of a higher altitude of air circulation from NE Asia is obtained for high-KD days in comparison to low-KD days (supplementary figure 13).

This surrogate variable representing the presence and movement of the SWC_1 in particles in the Tokyo air column (absc532 standing for 532 nm attenuated backscattering coefficient) is inferred from LIDAR optical measurements, and shown averaged at both the surface and above the PBL for Tokyo, Toyama and Tsukuba (supplementary figure 10(B)). A clear co-variation again shows up among the three locations when changes are studied at the seasonal scale. Month-to-month average synchronic evolution in the transport of this SWC_1 fraction of aerosols (here followed by comparing changes in the SWC_1 at different height layers in the troposphere), can be seen plotted together with monthly KD average (brown bars). This way, surface (green bars) and above the PBL (oPBL, blue bars), appear to coherently covary, showcasing the top-to-bottom movement of the SWC_1 in fine aerosols and the associated KD changes (supplementary figure 10(B)). The former might indicate an all-year round similarity, a consistent source

region and the same mechanisms associated with KD at all times, with only the maritime intrusions cleansing air during times of KD minima. In fact, when inspecting in detail such a multi-scale consistency between the arrival and entry of the SWC_1 and KD occurrence in each of the three locations, an abnormally leptokurtic distribution centered at zero lag emerges (supplementary figure 14). While it is expected that this distribution of events takes place centered at zero lag, it is significantly different from a similarly computed random model with the same amount of scale comparisons (blue line in supplementary figure 14). This indicates a fast response in KD following strong air intrusions, but we should stress that DoA accounts for a potential delay of 0–5 d since exposure. The attenuated backscattering ratio between 1064 nm and 532 nm (absc1064/absc532) has indeed been described as an appropriate analogue for tracing the sizes of aerosol particles (Raut and Chazette 2009, Grainger 2022), yielding values well below 1 (supplementary figure 15(A)); and, therefore, clearly denoting entrainment and dominance of particles of size below 1 μm (supplementary figure 15). This is again fully consistent among the three locations studied (supplementary figure 15(A)) and the movement of air can be tracked among different layers of the atmosphere and down to locations (supplementary figures 15(B) and 16, 17).

To expand on the above results, we applied a SDC correlation analysis in supplementary figure 16 between the same two variables for the interval 2010–2016. A strong link emerges between them that enhances towards the present, denoting a leading role for airborne particles (top series) on KD (left series; see Methods). This lead time is clearly evident as strong positive associations (red dots) denoting correlation values above +0.6 and up to +0.8 ($p < 0.001$) (see black box in lower panel) are present at and below the main diagonal, with the distance to the main diagonal representing the response time. A strong but non-linear relationship clearly emerges (e.g. see red striped area wandering below the main diagonal and intensifying around 2012). This same atmospheric aerosol transport over Tokyo was studied for an independent assessment also in the air column up to 6 km height and the direction of the movement traced (figure 5 and Bourgeois *et al* 2018). In this way, downward movement of weekly particles in the air column for 2010–2016, that arrive at Tokyo (e.g. difference calculated as before between the SCW above and below the PBL denoting the aerosol's arrival to the surface of the SWC_1; blue line) is shown in figure 5 (top panel) (see Methods). Wind speed (red line) and KD incidence (black line) are also depicted for inspection. A clear co-variation is evident between the SWC_1 and KD maxima co-occurring at times of maximum wind speed, and not conversely, with air column stability (supplementary figures 15–17). These maxima therefore correspond to air intrusions and rapid downward motion over Tokyo, as indicated by the bottom panel in figure 5 showing the monthly averaged LIDAR atmospheric profile for 2010–2016 (see also supplementary figure 15). High (low) seasonal concentration of aerosols exists trapped in the surface due to strong (weak) downward wind advection typically occurring in the winter (summer) since 2010 (figure 5). This weekly evolution in the three aforementioned variables further reinforces their strong co-variation (figure 5).

4. Discussion

Previous studies on the causes of KD have essentially addressed variability at seasonal and longer time scales while discarding any detailed analyses on the day-to-day disease and aerosol changes. The discovery of a consistent sub-weekly (SWC_1) signature in all the KD epidemiological records, at both the prefecture level and for the entirety of Japan, which is present continuously since the early 1970s, sheds new light on the necessary concurrent factors leading to the development of KD events. This variability has not been previously considered, or was simply interpreted as a weekend bias in reporting, as only high-frequency Rossby wave (RW SHG) events approaching 8 d have sporadically been described that generate via cascade from low-frequency waves. They occur mainly in the subtropics and therefore are extremely rare and identified in the middle atmosphere (He and Forbes 2022). Now our results clearly show that the fundamental cycle of 3.5 d does not match a mere weekend effect (figure 5, supplementary figure 10). Such co-variability cannot simply arise by chance and sustains a strong factual demonstration.

The implications of this finding are considerable as they point either to the chemical composition and/or the physical size of particles as factors important to KD epidemiology. Both features have been hypothesized -as shown above- as leading to the exacerbation of human immune responses that are also typical in KD.

The existence of a WC in human diseases -e.g. in clinical data- has long been attributed to the weekday–weekend dynamics in hospital admissions (e.g. worse prognosis for patients admitted during weekends in poor health care centers with reduced staffing levels or less experienced staff; Zare *et al* 2007, Cavaliere *et al* 2008, Uematsu *et al* 2016), or to lower number of admissions for mild diseases. Also, to biases in reporting (e.g. admissions being assigned to the following Monday or working day; Barnett *et al* 2002). Lower admission rates also occur over the weekend for asthma and mild diseases (Rosselló-Urgell *et al* 2004) and in prevalence surveys of nosocomial infections, due to the reduced effects of air pollution on hospital admissions and mortality (Fung *et al* 2003). The association of mortality with

day and time of admission has been tackled by many studies, with considerable heterogeneity in the results.

It is well-known that no natural WCs exist, being the weekly period one with a unique anthropogenic signature (Cerveny and Balling 1998, Gong *et al* 2007, Xia *et al* 2008), other than the atmospheric long waves, but having mostly a subtropical origin. Human activity in industrialized countries, instead, largely follows a 7 d cycle, where fossil fuel combustion is expected to be reduced during weekends (Cleveland *et al* 1974, Martin *et al* 2003, Kanda 2007, Sanchez-Lorenzo *et al* 2012). This weekend effect is well known from local, ground-based measurements, and may even translate into a small temperature signature, associated with a reduction in transportation (Beirle *et al* 2003). In fact, the influence of pollution on weather patterns is long known (going back to at least Ashworth 1929), and illustrates how weekdays are usually warmer than weekends (Cleveland *et al* 1974, Gordon 1994), and how atmospheric pollution shows substantial weekly variability (Wu *et al* 2022).

Trace elements detected in the 37 d record analyzed are enriched to varying degrees in the different particle size distributions, and varying concentrations of Zn and Pb and to a lesser degree also Mn, Ba, Se and Cu are evident with coherent PM_{10} and C_{tot} loads. This composition denotes a mixture of dust and aerosol of small sizes. A mixture of dust and pollution aerosols from different parts of East Asia (e.g. corresponding to air stagnation over central and the eastern polluted China regions, a Gobi Desert plume from the NW with agricultural debris) creates a complex air quality plume. Considering the distinct air quality phases described in this of 37 d, days with maximum KD cases occur during high continental pollution episodes. These events are mainly characterized by fine sulphatic transboundary aerosol intrusions driven by the established anticyclone, accompanied by higher ambient concentrations of metallic trace elements, such as Pb, As, Zn and Bi. In contrast, the advection of cleansing marine air, recorded by NaCl peaks, removes the Asian mainland influence and results in days with minimum KD cases. These marine aerosols were associated with minimum levels of the most toxic elements, such as As, Pb, Bi and Cd. The coarsening of aging recirculating sulphate aerosols, with low toxic metal loading, seems to be associated with minimum KD occurrence.

These trace elements in airborne PM indicate anthropogenic combustion processes (Dordevic *et al* 2014). Factor analysis in a study of the deposition of MM from the atmosphere in the most polluted city in China (Chengdu) indicated that As, Pb, Zn, Cu, Cd, and antimony (Sb) mainly originated from anthropogenic sources. Also, another type of sulphatic aerosol (rich in Zn and Cu) has been recognized in the finer fraction. Although As, Zn, and Cu could all result from coal combustion, as and the other MM have potentially different sources (Cheng *et al* 2018). Overall, the patterns found are extremely consistent with an agricultural origin, somewhat enriched by local events adding air pollution from adjacent urban areas. Interestingly, when inspecting whether the sampled atmospheric event in Kumamoto corresponds to a high or low KD incidence episode, supplementary figure 1 clearly indicates a low-to-moderate KD incidence (see also supplementary figure 2).

Previous studies addressing a relationship between urban air pollution and KD have shown negative or inconclusive results and controversy yet remains in the scientific literature (Zeft *et al* 2016). However, recently, Buteau *et al* (2020), using a population-based cohort comprising 505 336 children and including 539 with KD, found that both prenatal exposure to ambient and industrial air pollution were associated with the incidence of KD in childhood. Similarly, Jung *et al* (2017) analyzing hospital admissions in Taiwan showed that O_3 was positively associated with KD in the same day of admission, as well as during earlier seasons. Our results harmonize all the former studies in this direction, thanks to the unprecedented analysis of this high-resolution 46 yr daily record of KD admissions. This should therefore stimulate further research on a potential link between KD -or with other similar diseases- and the chemical nature of aerosols as a contributing or concomitant factor.

Clear WCs of anthropogenic origin, such as for NO_2 and PM_x have been described by satellite measurements (e.g. GOME; Cleveland *et al* 1974, Beirle *et al* 2003). Peaks and troughs, however, also vary among regions for cultural and even religious traditions. Some studies have equally shown WCs of tropospheric NO_2 for different regions of the world, with a clear minimum in the US, Europe and Japan and even with 35% lower NO_2 levels on Sundays than on working days in Germany (Wickert 2001). Therefore, variability at this intra-weekly scale can only be linked to an anthropogenic effect on air quality, with two maxima and two minima in both KD and aerosols. The former is clearly evident for the three locations studied (Tokyo, Toyama and Tsukuba). These same patterns are effectively traced to exist for air particles at all levels in the lower troposphere up to 6 km, being the SWC_1 enhanced at times of air intrusions (and concurrent with KD maxima). However, the fact that SWC components altogether account at most for around 20% of the overall KD variability, denotes that the factors associated with SWC_1 are, however, not sufficient to determine alone a KD outcome. The former is in vast agreement with other studies where seasonal-to-interannual changes were studied as the main source of KD variability (Ballester *et al* 2013, Buteau *et al* 2020), but our study sets the numerical basis to substantiate this assertion. For instance, while the three large epidemics might be related to climate extremes in the source region (e.g. floods or droughts), the trend in KD is possibly mostly related

to other processes distinct to the SWC here described (e.g. increases in cropland yield and cover; Rodó *et al* 2016), despite being somewhat also present in supplementary figure 5. MM contamination accumulating in soils is also a major environmental concern that affects large areas worldwide. Agricultural practices have been the main source of trace elements in soil such as Pb, Cr, As, Zn, Cd, Cu and Ni as plants can uptake these toxic MM and nitrogen fertilization also increases Cd concentrations in soil and plants (Wangstrand *et al* 2007). Phosphate fertilizers are also a significant source for contaminated soils with potentially toxic elements, such as Cd, F, Hg and Pb (Guo and Zhou 2006, Moraes 2009) and As. Our results also confirm that the anthropogenic component of suspended PM in the atmosphere usually builds up more slowly and does not fully decline as rapidly as the mineral and organic coarse fraction dust (figure 1(B)).

Exposure to some of the MM found (e.g. Zn, Cu and Pb) have shown a high correlation with pulmonary inflammation by inhalation toxicological studies using animal models (Park *et al* 2018). The long- and short-term exposure to these hazardous air pollutants, individually or collectively, may cause respiratory (eye, nose, throat and sinus) irritation, allergic dermatitis and adverse cardiovascular effects (Mills *et al* 2009). Inhalation of air pollutants thus induces pulmonary oxidative stress (ROS) and inflammation and the presence of soluble transition metals in aerosols enhances the inflammatory responses via increased oxidative stress and the release of ROS (Davis *et al* 2021). Fine ($PM_{2.5}$) and ultrafine particles (those <0.1 μm) induce innate immune response via ROS generation by transition metals and/or polyaromatic hydrocarbons (Roberts *et al* 2003). They modulate both the innate and adaptive immune responses (Hollingsworth *et al* 2007, Mikerov *et al* 2008, Miyata and van Eeden 2011, Zhao *et al* 2014) in ways that can be similar to those observed in KD pathogenesis, although specific research along these topics is needed.

5. Conclusions

Our study is the first to successfully demonstrate how the close scrutiny of LIDAR profiles can be used as effective sentinels for alerting of high-risk KD events. The highly-informative real-time data that the wide AD-Net network in Japan and nearby countries provide, should be promptly used for developing this important air quality service for health.

This study establishes that KD exacerbation is associated with external air intrusions having an anthropogenic signature that might also be associated to aerosol's MM. The analysis of metal contents can also be used as an early-warning system helping in preventing high KD episodes despite this link to MM requires further research and longer-term datasets. Although causality cannot be directly inferred, strength in these results definitely point towards an unequivocal link among them all and to the involvement of tropospheric fine aerosols in the epidemiology of KD. This aerosol facilitation can at times account for around half of the total variability in this disease, despite it being most often seen to act as a secondary -albeit necessary- cofactor in KD epidemiology. While we do not dispute that a portion of the variability in disease epidemiology can be associated with other local factors increasing individual susceptibility, what is clear is that these would, in any case, be very minor when compared to this atmospheric transport from NE Asia. Transiently, different air mixing apportions take place from both highly polluted urban centers and intensive cereal cropland areas. These results point to the chemical (and possibly also the biological) composition of agricultural aerosols as a likely cause for KD and to a dose/response relationship to MM conducive to cell inflammation. However, it also remains to be investigated as other studies have suggested, whether embedded in aerosols and affected by their chemistry, or other factors, such as microbes or organic by-products (e.g. coming from pesticides or fertilizers), in genetically predisposed children, might also be playing a key role in KD pathology and in other similar vasculitis diseases.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/alfontal/kd-metals-swc. Data will be available from 15 December 2023.

Acknowledgments

KD datasets were kindly provided by Professor Y Nakamura at Jichi Medical Hospital by means of Japan nationwide surveys. LIDAR data were kindly provided by Atsushi Shimizu and the AD-Net database at NIES, Tsukuba (www-lidar.nies.go.jp/AD-Net/index.html). We would like to thank Jane Burns, Teresa Moreno, Xavier Querol, Dan Cayan, Joan Ballester, Atsushi Matsuki and Hiroshi Tanimoto for their helpful assistance, technical support and data sharing. A N-G and A F acknowledge financial support from HELICAL as part of the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie Grant Agreement No. 81354. A N-G conducted the research at both the ADAPT Centre for Digital Content Technology (Grant Number 13/RC/2106_P2) at Trinity College Dublin and ISGlobal. We also acknowledge the Daniel Bravo Andreu Private Foundation for their financial support through the research Grant WINDBIOME and the KD parents' association Asenkawa for both their

economic and enthusiastic support. X R acknowledges support from the grant CEX2018 000806S funded by MCIN/AEI/10.13039/501100011033 and support from the Generalitat de Catalunya through the CERCA program.

Author's contributions

X R took the leading role in the conception and design of this study, A N-G and J A M participated in the design, X R, A N-G, S B, J C B, T K, A F participated in the acquisition of data, A N-G, A F and X R analyzed the data and all authors participated in the interpretation of data and discussion of results. X R drafted the article and all authors edited and approved the manuscript.

Code accession

All codes are available on request to the corresponding author and the publicly shareable data and code will be accessible at publication time in the open-source repository https://github.com/alfontal/kd-metals-swc.

ORCID iDs

Xavier Rodó https://orcid.org/0000-0003-4843-6180
Albert Navarro-Gallinad https://orcid.org/0000-0002-2336-753X
Silvia Borràs https://orcid.org/0000-0002-0331-0720
Alejandro Fontal https://orcid.org/0000-0003-1138-2158

References

2020 Green and grow *Nat. Sustain.* **3** 253
Andersen T, Carstensen J, Hernandez-Garcia E and Duarte C M 2009 Ecological thresholds and regime shifts: approaches to identification *Trends Ecol. Evol.* **24** 49–57
Ashworth J R 1929 The influence of smoke and hot gases from factory chimneys on rainfall *Q. J. R. Meteorol. Soc.* **55** 341–50
Ballester J, Burns J C, Cayan D, Nakamura Y, Uehara R and Rodó X 2013 Kawasaki disease and ENSO-driven wind circulation *Geophys. Res. Lett.* **40** 2284–9
Barnett M J, Kaboli P J, Sirio C A and Rosenthal G E 2002 Day of the week of intensive care admission and patient outcomes: a multisite regional evaluation *Med. Care* **40** 530–9
Barton M, Melbourne R, Morais P and Christie C 2002 Kawasaki syndrome associated with group A streptococcal and Epstein-Barr virus co-infections *J. Trop. Pediatr.* **22** 257–60
Beirle S, Platt U, Wenig M and Wagner T 2003 Weekly cycle of NO_2 by GOME measurements: a signature of anthropogenic sources *Atmos. Chem. Phys.* **3** 2225–32
Bourgeois Q, Ekman A M L, Renard J-B, Krejci R, Devasthale A, Bender F A-M, Riipinen I, Berthet G and Tackett J L 2018 How much of the global aerosol optical depth is found in the boundary layer and free troposphere? *Atmos. Chem. Phys.* **18** 7709–20
Bragazza L 2008 A climatic threshold triggers the die-off of peat mosses during an extreme heat wave *Glob. Change Biol.* **14** 2688–95
Burns J C *et al* 2013 Seasonality of Kawasaki disease: a global perspective *PLoS One* **8** e74529
Burns J C, Cayan D R, Tong G, Bainto E V, Turner C L, Shike H, Kawasaki T, Nakamura Y, Yashiro M and Yanagawa H 2005 Seasonality and temporal clustering of Kawasaki syndrome *Epidemiology* **16** 220–5
Burns J C and Glodé M P 2004 Kawasaki syndrome *Lancet* **364** 533–44
Buteau S, Belkaibech S, Bilodeau-Bertrand M, Hatzopoulou M, Smargiassi A and Auger N 2020 Association between Kawasaki disease and prenatal exposure to ambient and industrial air pollution: a population-based cohort study *Environ. Health Perspect.* **128** 107006
Cavaliere F, Conti G, Costa R, Masieri S, Antonelli M and Proietti R 2008 Intensive care after elective surgery: a survey on 30-day postoperative mortality and morbidity *Minerva Anestesiol.* **74** 459–68
Cerveny R S and Balling R C 1998 Weekly cycles of air pollutants, precipitation and tropical cyclones in the coastal NW Atlantic region *Nature* **394** 561–3
Cheng X, Huang Y, Zhang S-P, Ni S-J and Long Z-J 2018 Characteristics, sources, and health risk assessment of trace elements in PM_{10} at an Urban Site in Chengdu, Southwest China *Aerosol Air Qual. Res.* **18** 357–70
Cleveland W S, Graedel T E, Kleiner B and Warner J L 1974 Sunday and workday variations in photochemical air pollutants in New Jersey and New York *Science* **186** 1037–8
Davis W T, Fu J S and Godish T 2021 *Air Quality* (Boca Raton, FL: CRC Press)
Dettinger M D, Ghil M, Strong C M, Weibel W and Yiou P 1995 Software expedites singular-spectrum analysis of noisy time series *EOS Trans. Am. Geophys. Union* **76** 12–21
Đorđević D, Stortini A M, Relić D, Mihajlidi-Zelić A, Huremović J, Barbante C and Gambaro A 2014 Trace elements in size-segregated urban aerosol in relation to the anthropogenic emission sources and the resuspension *Environ. Sci. Pollut. Res.* **21** 10949–59
Fontal A, Bouma M J, San-José A, López L, Pascual M and Rodó X 2021 Climatic signatures in the different COVID-19 pandemic waves across both hemispheres *Nat. Comput. Sci.* **1** 655–65
Frazer J 2012 Infectious disease: blowing in the wind *Nature* **484** 21–23
Fung K Y, Krewski D, Chen Y, Burnett R and Cakmak S 2003 Comparison of time series and case-crossover analyses of air pollution and hospital admission data *Int. J. Epidemiol.* **32** 1064–70
Ghil M *et al* 2002 Advanced spectral methods for climatic time series *Rev. Geophys.* **40** 3-1-3–41
Gong D-Y, Ho C-H, Chen D, Qian Y, Choi Y-S and Kim J 2007 Weekly cycle of aerosol-meteorology interaction over China *J. Geophys. Res.* **112** D22202
Gordon A H 1994 Weekdays warmer than weekends? *Nature* **367** 325–6
Grainger R G 2022 Some useful formulae for aerosol size distributions and optical properties *Lect. Notes (University of Oxford)* 12–13 (available at: http://eodg.atm.ox.ac.uk/user/grainger/research/aerosols.pdf)
Guo G L and Zhou Q X 2006 Evaluation of heavy metal contamination in Phaeozem of northeast China *Environ. Geochem. Health* **28** 331–40
He M and Forbes J M 2022 Rossby wave second harmonic generation observed in the middle atmosphere *Nat. Commun.* **13** 7544
Hollingsworth J W, Maruoka S, Li Z, Potts E N, Brass D M, Garantziotis S, Fong A, Foster W M, Schwartz D A and Bourgeois Q 2007 Ambient ozone primes pulmonary innate immunity in mice *J. Immunol.* **179** 4367–75
Jayasumana C, Fonseka S, Fernando A, Jayalath K, Amarasinghe M, Siribaddana S, Gunatilake S and Paranagama P 2015 Phosphate fertilizer is a main source of arsenic in areas affected with chronic kidney

disease of unknown etiology in Sri Lanka *SpringerPlus* **4** 1–8
Jung C-R, Chen W-T, Lin Y-T and Hwang B-F 2017 Ambient air pollutant exposures and hospitalization for Kawasaki disease in Taiwan: a case-crossover study (2000–2010) *Environ. Health Perspect.* **125** 670–6
Kanda M 2007 Progress in urban meteorology: a review *J. Meteorol. Soc. Jpn.* II **85B** 363–83
Lee M N, Cha J H, Ahn H M, Yoo J H, Kim H S, Sohn S and Hong Y M 2011 Mycoplasma pneumoniae infection in patients with Kawasaki disease *Eur. J. Agron.* **54** 123–7
Leung D Y, Kotzin B L, Meissner H C, Fulton R D, Murray D L and Schlievert P M 1993 Toxic shock syndrome toxin-secreting Staphylococcus aureus in Kawasaki syndrome *Lancet* **342** 1385–8
Lindegren M and Eero M 2013 Threshold-dependent climate effects and high mortality limit recruitment and recovery of the Kattegat cod *Mar. Ecol. Prog. Ser.* **490** 223–32
Maki T *et al* 2019 Vertical distributions of airborne microorganisms over Asian dust source region of Taklimakan and Gobi Desert *Atmos. Environ.* **214** 116848
Marrani E, Burns J C and Cimaz R 2018 How should we classify Kawasaki disease? *Front. Immunol.* **9** 2974
Martin R V *et al* 2003 Global inventory of nitrogen oxide emissions constrained by space-based observations of NO_2 columns *J. Geophys. Res.* **108**
Matsubara K *et al* 2006 Development of serum IgM antibodies against superantigens of Staphylococcus aureus and Streptococcus pyogenes in Kawasaki disease *Clin. Exp.* **143** 427–34
Mikerov A N, Adair-Kirk T L, Kelley D G, Demello D and Senior R M 2008 Ablation of SP-A has a negative impact on the susceptibility of mice to Klebsiella pneumoniae infection after ozone exposure: sex differences *Respir. Res.* **9** 1–16
Mills N L, Donaldson K, Hadoke P W, Boon N A, MacNee W, Cassee F R, Sandström T, Blomberg A and Newby D E 2009 Adverse cardiovascular effects of air pollution *Nat. Clin. Pract. Cardiovascular Med.* **6** 36–44
Miyata R and van Eeden S F 2011 The innate and adaptive immune response induced by alveolar macrophages exposed to ambient particulate matter *Toxicol. Appl. Pharmacol.* **257** 209–26
Moraes M F D 2009 Micronutrientes e metais pesados tóxicos: Do fertilizante ao produto agrícola *Doctoral Thesis* Catalogo USP (Piracicaba: Universidade de São Paulo) (https://doi.org/10.11606/T.64.2009.tde-14102009-100525)
Moreno T *et al* 2013 Daily and hourly chemical impact of springtime transboundary aerosols on Japanese air quality *Atmos. Chem. Phys.* **13** 1411–24
Nagata S 2019 Causes of Kawasaki disease-from past to present *Front. Pediatr.* **7** 18
Nishizawa T, Sugimoto N, Matsui I, Shimizu A, Higurashi A and Jin Y 2016 The Asian dust and aerosol LIDAR observation network (AD-NET): strategy and progress *EPJ Web Conf.* **119** 19001
Onouchi Y 2009 Molecular genetics of Kawasaki disease *Pediatr. Res.* **65** 46–54
Park M *et al* 2018 Differential toxicities of fine particulate matters from various sources *Sci. Rep.* **8** 1–11
Peralta-Videa J R, Lopez M L, Narayan M, Saupe G and Gardea-Torresdey J 2009 The biochemistry of environmental heavy metal uptake by plants: implications for the food chain *Int. J. Biochem. Cell Biol.* **41** 1665–77
Raut J-C and Chazette P 2009 Assessment of vertically-resolved PM 10 from mobile LIDAR observations *Atmos. Chem. Phys.* **9** 8617–38
Roberts E S, Richards J H, Jaskot R and Dreher K L 2003 Oxidative stress mediates air pollution particle-induced acute lung injury and molecular pathology *Inhal. Toxicol.* **15** 1327–46
Rodó X 2001 Reversal of three global atmospheric fields linking changes in SST anomalies in the Pacific, Atlantic and Indian oceans at tropical latitudes and midlatitudes *Clim. Dyn.* **18** 203–17
Rodó X *et al* 2014 Tropospheric winds from northeastern China carry the etiologic agent of Kawasaki disease from its source to Japan *Proc. Natl Acad. Sci. USA* **111** 7952–7
Rodó X, Ballester J, Cayan D, Melish M E, Nakamura Y, Uehara R and Burns J C 2011 Association of Kawasaki disease with tropospheric wind patterns *Sci. Rep.* **1** 152
Rodó X, Ballester J, Curcoll R, Boyard-Micheau J, Borràs S and Morguí J-A 2016 Revisiting the role of environmental and climate factors on the epidemiology of Kawasaki disease *Ann. New York Acad. Sci.* **1382** 84–98
Rodó X, Martinez P P, Siraj A and Pascual M 2021 Malaria trends in Ethiopian highlands track the 2000 'slowdown'in global warming *Nat. Commun.* **12** 1–12
Rodó X, Pascual M, Fuchs G and Faruque A S G 2002 ENSO and cholera: a nonstationary link related to climate change? *Proc. Natl Acad. Sci.* **99** 12901–6
Rodó X and Rodríguez-Arias M-À 2006 A new method to detect transitory signatures and local time/space variability structures in the climate system: the scale-dependent correlation analysis *Clim. Dyn.* **27** 441–58
Rodríguez-Arias M A and Rodó X 2004 A primer on the study of transitory dynamics in ecological series using the scale-dependent correlation analysis *Oecologia* **138** 485–504
Rosselló-Urgell J, Vaqué-Rafart J, Armadans-Gil L L, Vaquero-Puerta J L, Elorza-Ricart J M, Quintas-Fernández J C, Hidalgo-Pardo O and Arévalo-Alonso J M 2004 The importance of the day of the week and duration of data collection in prevalence surveys of nosocomial infections *J. Hosp. Infection* **57** 132–8
Rowley A H, Baker S C, Orenstein J M and Shulman S T 2008 Searching for the cause of Kawasaki disease–cytoplasmic inclusion bodies provide new insight *Nat. Rev. Microbiol.* **6** 394–401
Sanchez-Lorenzo A, Laux P, Hendricks Franssen H-J, Calbó J, Vogl S, Georgoulias A K and Quaas J 2012 Assessing large-scale weekly cycles in meteorological variables: a review *Atmos. Chem. Phys.* **12** 5755–71
Stein A F, Draxler R R, Rolph G D, Stunder B J B, Cohen M D and Ngan F 2015 NOAA's HYSPLIT atmospheric transport and dispersion modeling system *Bull. Am. Meteorol. Soc.* **96** 2059–77
Tufail M and Khalid S 2008 Heavy metal pollution from medical waste incineration at Islamabad and Rawalpindi, Pakistan *Microchem. J.* **1** 77–81
Uematsu H, Kunisawa S, Yamashita K, Fushimi K and Imanaka Y 2016 Impact of weekend admission on in-hospital mortality in severe community-acquired pneumonia patients in Japan *Respirol. Carlton Vic.* **21** 905–10
Wångstrand H, Eriksson J and Öborn I 2007 Cadmium concentration in winter wheat as affected by nitrogen fertilization *Eur. J. Agron.* **26** 209–14
Wickert B 2001 Berechnung anthropogener Emissionen in Deutschland für Ozonsimulationen: modellentwicklung und Sensitivitätsstudien *Calculation of anthropogenic emissions in Germany for ozone simulation: model development and sensitivity studies* (https://doi.org/10.18419/opus-1544)
Wu L, Xie J and Kang K 2022 Changing weekend effects of air pollutants in Beijing under 2020 COVID-19 lockdown controls *npj Urban Sustain.* **2** 23
Xia X, Eck T F, Holben B N, Phillippe G and Chen H 2008 Analysis of the weekly cycle of aerosol optical depth using AERONET and MODIS data *J. Geophys. Res.* **113** D14217
Zare M M, Itani K M F, Schifftner T L, Henderson W G and Khuri S F 2007 Mortality after nonemergent major surgery performed on Friday versus Monday through Wednesday *Ann. Surg.* **246** 866–74
Zeft A S *et al* 2016 Kawasaki disease and exposure to fine particulate air pollution *J. Pediatr.* **177** 179–183.e1
Zhao H, Li W, Gao Y, Li J and Wang H 2014 Exposure to particular matter increases susceptibility to respiratory Staphylococcus aureus infection in rats via reducing pulmonary natural killer cells *Toxicology* **325** 180–8

PAPER II

Shift in seasonal incidence of Kawasaki Disease in Japan: an age-stratified spatiotemporal analysis

Alejandro Fontal[1, 2], and Xavier Rodó[1,3*]

[1]Climate and Health Group, Climate, Air pollution, Nature and Urban Health, Barcelona Institute for Global Health, Barcelona 08003, Spain

[2]Department of Microbiology, Genetics and Statistics, Faculty of Biology, University of Barcelona, Barcelona 08028, Spain

[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain

Article in preparation

Shift in seasonal incidence of Kawasaki Disease in Japan: an age-stratified spatiotemporal analysis

Alejandro Fontal[1, 2], and Xavier Rodó[1,3*]

[1]Climate and Health (CLIMA) Program, ISGlobal, Barcelona, Spain

[2]University of Barcelona (UB), Barcelona, Spain

[3]ICREA, Barcelona, Spain

Abstract

Kawasaki Disease (KD) is a systemic vasculitis primarily affecting young children and is the leading cause of pediatric acquired heart disease in developed countries. Despite extensive research over the past five decades, the etiology of Kawasaki Disease remains elusive, with the causes of its incidence patterns still unexplained.

In this cross-sectional time-series study, we explore the intricacies of KD's seasonality using data from Japan—the country with the highest worldwide incidence of KD—spanning over 20 years (2000 to 2019). Our focus is on examining how these seasonal patterns manifest across different age groups and regions. By extracting and isolating the temporal signals from age-stratified time-series data, we identify not only a marked distinction between the seasonal signals of different age groups but also an abrupt desynchronization between younger and older children after 2016. This shift is particularly evident in the under-2 age group, where we observe a strong change in seasonality, with a new autumn peak that is consistent across most regions in the country.

These findings suggest a shift in the interplay between environmental factors triggering the disease and the specific exposure windows for children under 2, pointing towards a potential nationwide change that could help explain the triggers behind this mysterious vasculitis.

1. Introduction

Kawasaki Disease (KD) is an acute self-limited vasculitis which is the leading cause of acquired heart disease in developed countries (1). Over 50 years after Dr. Tomisaku Kawasaki reported the first cases in the 1960's, there is still no clear answer to which the causative agent of the disease is nor what leads to the seasonal and epidemic-like dynamics or the increasing trend in incidence in Japan. The prevailing theory suggests that these patterns result from a complex interaction among children's biological responses, circulating pathogens and the transport of potential environmental triggers, paired with seasonal lifestyle changes modulating their exposures, underlining the need for interdisciplinary efforts to solve the long-standing mystery.

The extensive epidemiological data collected in Japan through nationwide surveys, meticulously curated by the team at Jichi Medical University since 1970, has been instrumental in advancing our understanding of the disease. It is through examination of these records that the idiosyncratic temporal dynamics of KD in Japan have been able to be studied. A comprehensive understanding of these dynamics—encompassing seasonality, incidence trends, and epidemic-like outbreaks—ought to provide critical insights into the factors that trigger the onset of KD.

The epidemiological records from the last decades of the 20th century stood out for the three significant nationwide epidemics of KD in 1979, 1982, and 1986; which occurred synchronously across

Japan (2,3). However, since the beginning of the 21st century, the dynamics have changed, with two main remarkable features:

1. Incidence rates showing a noticeable upward trend - the incidence rate in 2019 was nearly double that of the peak in 1982, rising from 196.1 to 370.8 cases per 100,000 children under five years old. The only exception was the last year of the records, 2020, which showed a significant reduction in cases compared to previous years (4).
2. A marked seasonal pattern, with a consistent and prominent peak during winter, followed by a less consistent and weaker peak during spring, and a nadir during the fall.

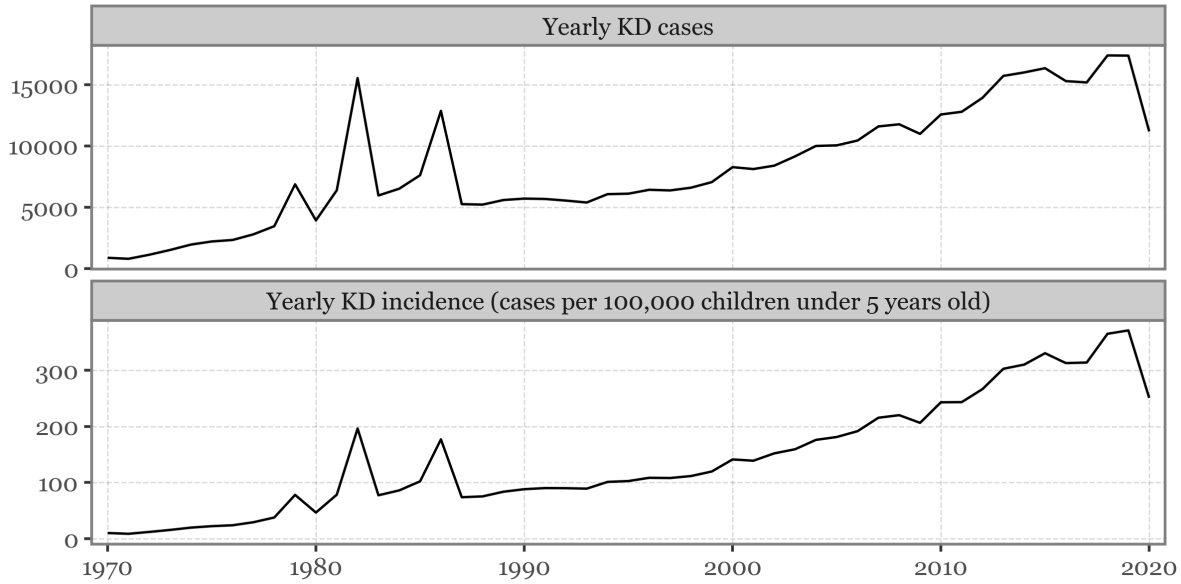


Figure 1. KD incidence trends. The top panel shows the total number of KD admissions per year from the initiation of nationwide surveys in 1970 through 2020. The bottom panel presents the incidence rate, adjusted for the population size of children under five years old, who account for more than 85% of all KD cases.

The seasonality of Kawasaki Disease has been extensively studied across various geographic scales, including national and subnational levels in Japan (5), Canada (6) or Spain (7), as well as on a global scale (8). Several hypotheses have been proposed to explain the observed variability in the annual incidence cycles: fluctuations in circulating pathogens, with the prevalence of viral infections typically peaking during the winter months; seasonal modulation of immune system activity and response, characterized by the differential expression of immune-related genes depending on the time of year (9,10); seasonal allergies, which reach their zenith in spring and early summer, likely as a consequence of elevated pollen levels and other airborne allergens (11). Additionally, changes in human behaviour and activities, such as the start of the school year and increased indoor activities during colder months, may also influence the exposure patterns of susceptible populations. Furthermore, as proposed by Rodó et al., atmospheric circulation patterns, particularly tropospheric winds, may play a critical role in the seasonal transport of causative agents from the Asian mainland to Japan (12–14).

Age and sex are critical determinants in the pathogenesis and progression of many diseases, particularly those involving the immune system or immune-mediated conditions (15). Generally, females exhibit stronger immune and inflammatory responses, which result in a higher prevalence of autoimmune diseases and improved survival rates from infections and malignant cancers. However, this pattern appears to be reversed during early life stages, including the in-utero and pre-puberty phases,

where males demonstrate heightened activity in both the innate (inflammatory) and adaptive immune responses (16). This phenomenon might partly explain the consistently higher incidence of KD in males given that 85% of KD cases occur in children under 5 years old, with only 1% of cases presenting in individuals over 10. Apart from the prevalence difference, where there is approximately a 1.5:1 male-to-female ratio, the temporal dynamics of KD are remarkably similar between both sexes. For a detailed visualization of the full age and sex distribution, refer to Figure 2 and Figure 3.

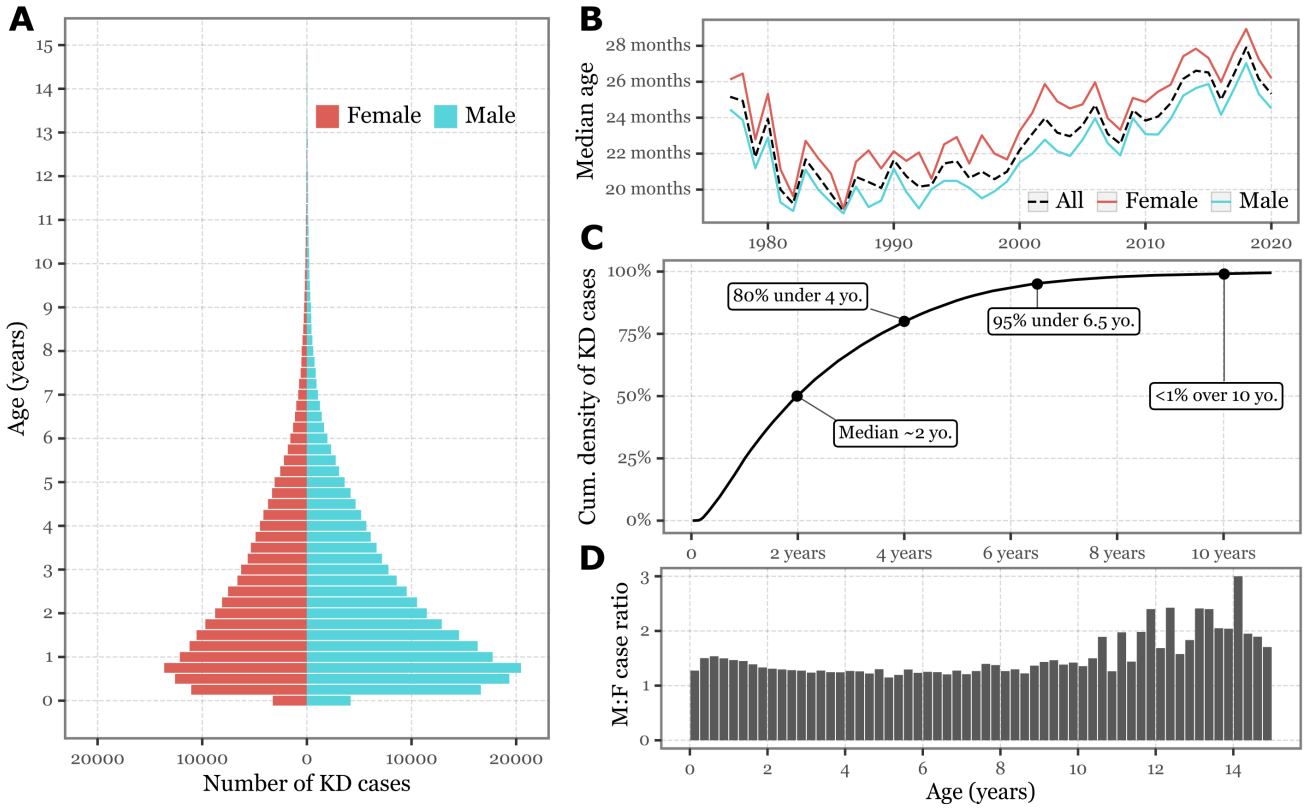


Figure 2. Age and sex distributions of KD patients. Panel A presents the total number of KD cases stratified by sex and organized into 3-month age intervals. Panel B illustrates the median age at KD diagnosis, stratified by sex, over time. Panel C displays the cumulative distribution of KD cases by age at diagnosis, highlighting key age thresholds. Panel D depicts the male-to-female case ratio as a function of age, shown in 3-month intervals.

Age, however, appears to play a more significant role in KD: the interaction between an individual and their environment, including exposure to various external factors, as well as the development and maturation of the immune system, undergo significant changes during the early months and years of life. Consequently, the dynamics of age-stratified disease patterns are likely to reveal differences that may not be apparent when age is not considered. However, stratifying data by age in the context of rare diseases poses challenges, primarily due to the increased noise-to-signal ratio, which can hinder robust statistical analysis. This is one reason why age-stratified analyses of KD have been limited, particularly in studies focusing on spatiotemporal patterns. Nevertheless, the comprehensive registry available for Japan offers a unique resource that could facilitate such analyses. Although a few studies have examined the interplay between the age of patients and the seasonal signals of the disease (17,18), no research, to our knowledge, has investigated the changes in these seasonal cycles over time. In this study, we focus on age-stratified incidences of Kawasaki Disease to examine time-varying synchrony between different age groups and overall changes in the seasonal patterns in Japan and its prefectures.

2. Data and Methods

KD hospital admission data. Data on KD hospital admissions in Japan were sourced from the epidemiological registry established by the team in the Public Health division of Jichi Medical University based on the data collected through the ongoing Japanese nationwide surveys. For the present study, we used data spanning the 16th to the 26th nationwide surveys, encompassing a total of 269,796 cases of KD patients admitted to hospitals throughout Japan from 1999 to 2020. The dataset includes, for each patient, the date of hospital admission, the prefecture of the admitting hospital, the patient's sex, and the precise age in days at the time of admission.

Age-group populations. To calculate nationwide and prefecture-specific KD incidence rates for the studied periods, population estimates were sourced from the Portal Site of Official Statistics of Japan (https://www.e-stat.go.jp). The dataset includes annual population estimates for single-year age groups at the national level, while prefectural estimates are provided in 5-year age brackets. To achieve a granular age-based decomposition at the prefectural level, we applied national-level single-year age group ratios to estimate population distributions within each prefecture based on the 5-year age groups. For higher resolution estimates within a single year, an assumption of homogeneous distribution was made, by which the total population of a specific age (e.g., 1-year-olds) was equally divided into subgroups (e.g., 0–6 months and 6–12 months). The national ratios estimated for the three primary age groups analysed (<6 months, 6–24 months, and 24–60 months) are shown in Supp. Figure 1.

Age-specific KD incidence rates. The age-specific KD incidence rates were then calculated by dividing the number of cases in each prefecture by the estimated population for the respective age group. To prevent artificial spikes or drops in incidence rates caused by the annual frequency of population estimates, we applied a daily interpolation to the population data, smoothing the signal and ensuring a more continuous and accurate representation of the trends.

Japanese prefectures boundaries. We obtained shapefiles delineating the boundaries of Japan's 47 prefectures to facilitate the computation of area-dependent variables and their representation in maps. These shapefiles were sourced from the dataset provided by the United Nations Office for the Coordination of Humanitarian Affairs (OHCA), under a Creative Commons Attribution for Intergovernmental Organizations (CC BY-IGO) license. The dataset can be accessed in the site of the Humanitarian Data Exchange https://data.humdata.org/dataset/cod-xa-jpn.

Meteorological data. We accessed ERA5-land data through the Copernicus Climate Data Store using the `cdsapi` Python library, with a spatial resolution of 0.1° covering the entire Japanese archipelago. Each grid cell was matched to its corresponding prefecture, and daily averages were computed for each variable and prefecture. The variables used included 2m temperature, total precipitation, 10m u-component of wind, and 10m v-component of wind. The u and v wind components were combined to project the observed wind directions onto a unit vector representing the northwest/southeast axis. Positive values of the northwest (NW)-wind variable indicate winds blowing from the north, northwest, or west, while negative values represent winds originating from the south, southeast, or east.

Environmental data. We gathered data from 2,039 Air Pollution Monitoring stations distributed across Japan and calculated prefecture-level averages by weighting the stations according to the relative population of children under 5 years old in each municipality within the prefecture. The comprehensive

dataset and detailed documentation of the data processing steps are available in this GitHub repository: https://github.com/AlFontal/environmental-data-japan. The variables we used included SO_2, NO, NO_2, NO_x, $PM_{2.5}$, CO, CH_4, Total Hydrocarbons (THC), and Non-methane Hydrocarbons (NMHC).

Vegetation indexes. We used Google Earth Engine (GEE) through its Python API to extract Enhanced Vegetation Index (EVI) values from the MOD13A2 data product, which provides 16-day composites of Terra MODIS satellite imagery. The data covers the entire globe at a spatial resolution of 1000m, starting from 2000-02-18. We opted for EVI over NDVI due to its increased resilience to atmospheric conditions and cloud cover. Specifically, we extracted data for the Japanese archipelago (calculating 16-day prefecture-wise averages) and an agricultural region in Northeast China (bounding box with ranging from latitudes 42N-50N and 115E-134E), which aligns with the identified source area of tropospheric winds reaching Japan during periods of high KD incidence (13). To smooth the signals and mitigate variations caused by cloud cover, we applied a Savitzky-Golay filter to the EVI values, using a window size of 7 and a polynomial order of 3.

Air source assignations. To identify the regions contributing aerosols to Japanese prefectures, we employed Lagrangian particle dispersion modelling using HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory Model) version 5.2.0 (19). Simulations were conducted via the R-language API, utilizing the open-source `splitr` library. We implemented backward dispersion models, which encountered issues in the final release of the `splitr` library (https://github.com/rich-iannone/splitr/pull/82). These issues were addressed by patching the code, with the fix made available in a forked public repository (https://github.com/AlFontal/splitr). To obtain a representative image of the air sources across the country, we ran 60 backward dispersion simulations from the centroid of each of the Japanese prefectures (excluding Okinawa) per day, with each simulation spanning 96 hours.

Seasonal signals. To extract seasonal signals, we used a 28-day (4-week) centred moving average to the incidence rates, smoothing out day-to-day variations and the weekly cycle to focus on longer seasonal frequencies. Initially, this transformation was performed separately for male and female patients for the 2000 to 2019 period. After determining that there were no significant differences in seasonal patterns between the sexes (see Figure 3) we combined the data to increase signal strength and focus on age. We then proceeded with the analysis using granular age groups: 6-month intervals for children under 2 years old and 1-year intervals for children over 2 and under 10 years old. Ultimately, this led to the selection of three primary age groups for further analysis: <6 months, 6–24 months, and 24–60 months. Seasonal signals were expressed as relative rates with respect to yearly averages to facilitate comparison across different periods and groups. The same methodology was used to extract seasonal signals for the environmental and meteorological variables.

Long-term Trends. To compare the long-term trends of KD incidence across groups and remove seasonal fluctuations, we focused on interannual trends by applying a 365-day trailing moving average.

Temporal synchronicity and decoupling between age-groups. We employed several statistical methods to evaluate the synchronicity between KD incidences across different age groups. For discrete yearly comparisons, Pearson correlations were calculated on the extracted seasonal signals to assess the correlation in the seasonal behaviour of KD incidences. To continuously assess the degree of coupling and identify any transient decouplings between the long-term seasonal signals, we use scale-dependent

correlation analysis (SDC) (20,21) with a window of 365 days to scan the coherence from the KD incidence series between age groups from 2000 to 2020.

Spatial autocorrelations. We assessed spatial autocorrelation using Global Moran's I, applying Queen's contiguity weights to define spatial relationships between prefectures. Due to Japan's archipelagic nature, some prefectures, such as Okinawa and Hokkaido, are geographically isolated and lack contiguous neighbours. As a result, these prefectures were excluded from the spatial autocorrelation analysis to avoid bias. The spatial weights matrix was row-standardized to ensure equal influence of each prefecture in the analysis.

3. Results

3.1 Sex-stratified dynamics

The stratification of KD incidence data by sex shows a consistent male-to-female case ratio of approximately 1.3 to 1.5 across the entire registry (Figure 3B). Despite this higher incidence in male children compared to females, once we decompose the temporal signal into trend and seasonality, we observe a highly synchronous behaviour in both components. Figure 3A illustrates the strikingly similar patterns followed by male and female incidences on a year-by-year basis, encompassing the three epidemics, the increasing trend, and the decrease observed in 2020 during the COVID-19 pandemic.

Regarding the seasonal signal, focusing on the 2000 to 2019 period, the seasonal cycles appear nearly identical between the sexes. Both show a consistent nadir in the fall, with cases increasing from December and peaking in January (Figure 3C). Yearly Pearson correlation coefficients of these seasonal signals range from 0.78 to 0.98, indicating a strong alignment in the seasonal patterns between males and females (Figure 3D). The only year with a notably lower correlation (0.73) was 2017, which still shows a substantial degree of synchrony. For a visual comparison, the seasonal patterns for the years with the highest (2007) and lowest (2017) correlations are displayed in Supplementary Figure 2.

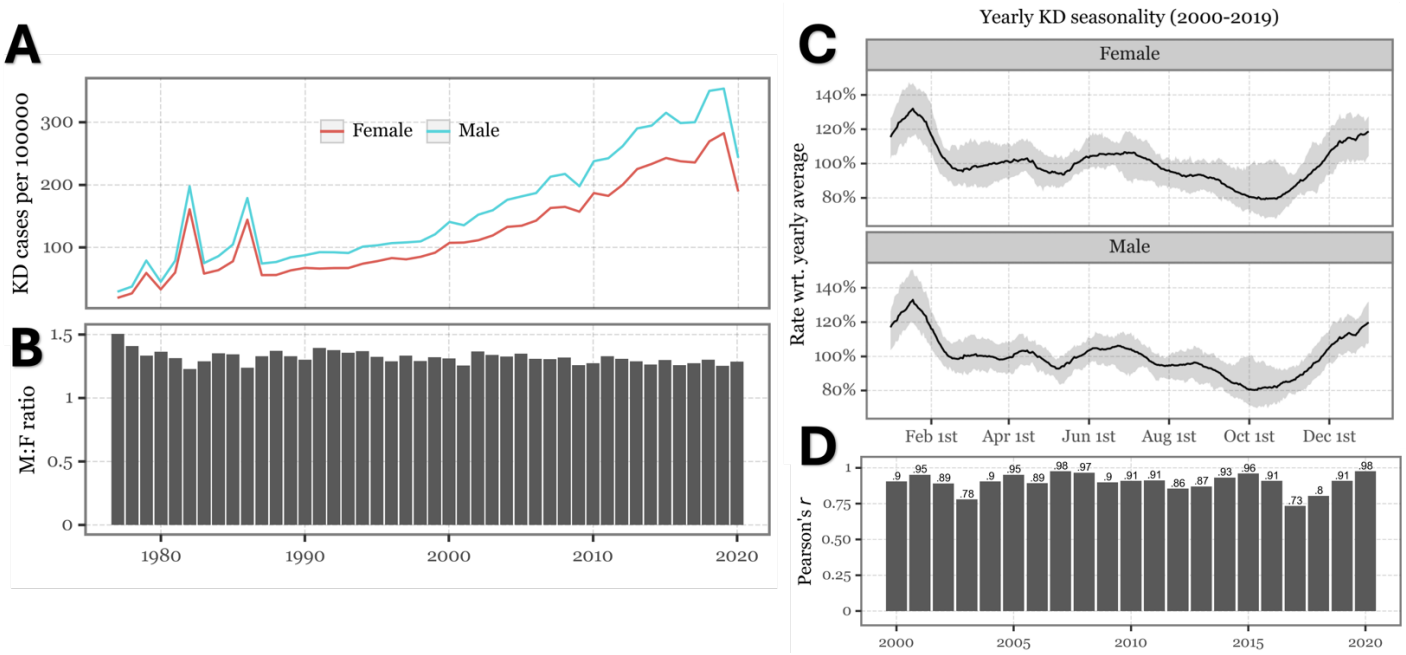


Figure 3. Sex-stratified seasonal and trend patterns of KD. Panel A shows the annual incidence rates of KD for males and females under 5 years old. Panel B presents the male-to-female ratio of KD incidence for each year, providing a comparison of sex-specific prevalence. Panel C illustrates the median and the 10th to 90th percentile range (shaded area) of the normalized seasonal cycle of KD incidence from 2000 to 2019, for both males and females, emphasizing the consistent seasonal peaks. Finally, Panel D depicts the Pearson correlation coefficients (r) between the male and female seasonal signals for each year, indicating the degree of synchrony in the temporal patterns of KD between the sexes.

Given this high synchrony and our objective to maximize the signal in our stratified analyses, we merged the male and female groups in the subsequent analyses, enabling us to concentrate on the distinct signals provided by the different age groups

3.2 Age-stratified dynamics

To enhance the granularity of our analysis, we initially stratified patients under 2 years old into 6-month age intervals and those between 2 and 9 years old into 1-year intervals. The decomposition of temporal dynamics into trend and seasonal components for these more granular age groups is shown in Figures 4 and 5. The trend component clearly showed a greater increase with age, especially among older children (7-9 years), reaching a 4-fold change with respect to the incidence in 1996. The incidence for the youngest subgroups, in comparison, barely stayed the same. The seasonal patterns revealed three distinct subgroups: (1) newborns under 6 months, with a weak winter peak and a higher plateau from late spring to early fall; (2) older children (2-9 years), with a pronounced winter peak, a strong fall nadir, and minor peaks in spring and summer for ages 4-9 years; and (3) toddlers between 6 and 24 months, with a pronounced winter peak but inconsistent summer and fall behaviour. Given these distinctions, we decided to focus on patients under 5 years (constituting over 85% of total cases) and merge the time series in these three distinct subgroups to improve the signal-to-noise ratio and enable more robust comparisons. The percentage of total cases that each of these three groups represent is 7.7% for the <6 months, and 40.4% for each of the 6-24 and 24-60 months subgroups.

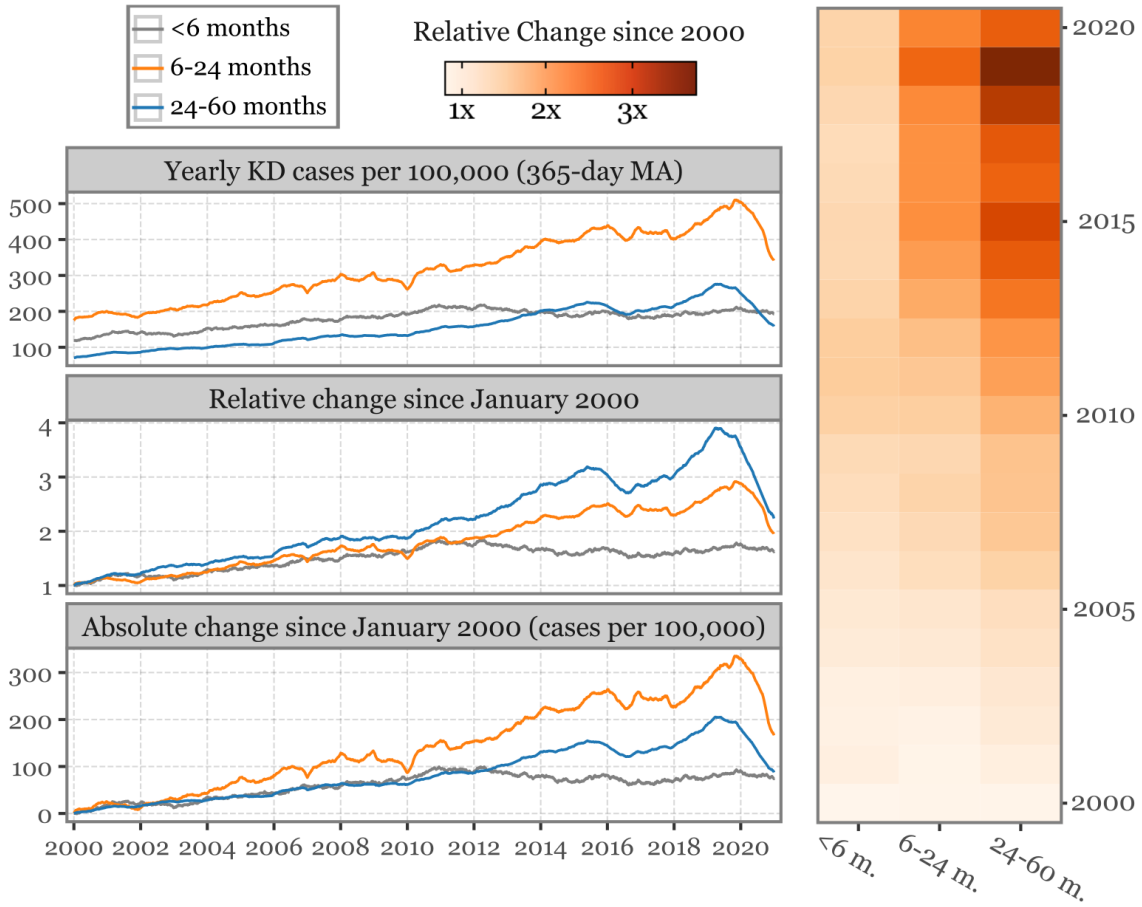


Figure 4. Long-term trends in age-stratified KD incidence since 2000. The left panels show the yearly KD incidence rates per 100,000 children (top), the relative change in incidence since January 2000 (middle), and the absolute change in incidence since January 2000 (bottom) for three age groups: <6 months (grey), 6-24 months (orange), and 24-60 months (blue). The data is smoothed using a 365-day moving average (MA) to highlight long-term trends. The right panel displays a heatmap of the relative change in KD incidence since 2000 for each age group, where darker shades indicate greater increases in incidence rates.

The long-term trends for the three selected age categories are then illustrated in Figure 4. From 2000 to 2019, the yearly incidence rates (cases per 100,000) have increased from 119 to 208 for newborns, from 175 to 504 for toddlers, and from 70 to 251 for young children. While the absolute increase has been the strongest in the toddler group (which consistently shows the highest incidence of the disease), the relative change has been even stronger for the young children, with nearly a 4-fold increase compared to 2000 levels. The interannual changes have remained remarkably consistent across these two groups. In contrast, newborns showed a steady increase in incidence from 2000 to 2011, followed by a plateau, maintaining yearly incidence rates around 200 cases per 100,000. The 2020 data show the impact of the COVID-19 pandemic on KD, with sharp declines in incidence for both the toddler and young children groups, whereas the incidence among newborns barely changed.

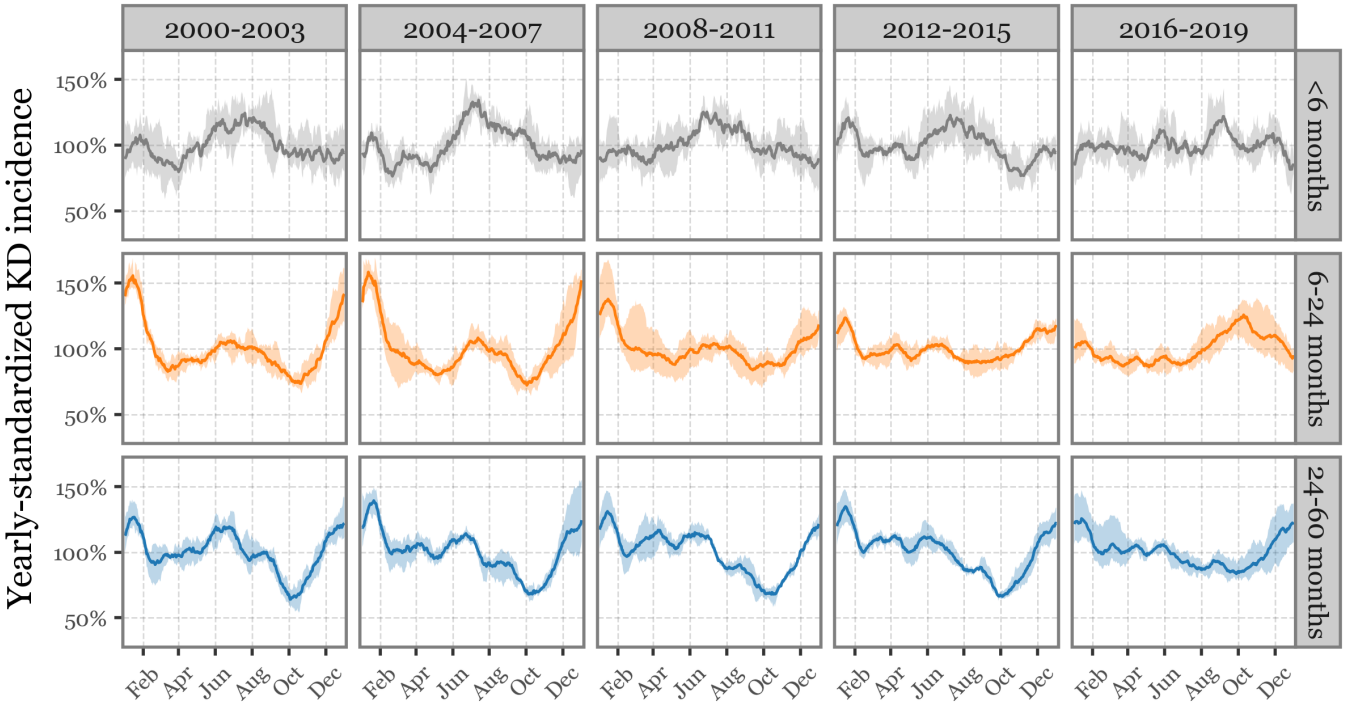


Figure 5. Age-stratified seasonal cycles. The solid lines represent the mean value of the yearly standardized KD incidence as a function of the day of the year for each combination of age group and 4-year periods. The shaded areas indicate the full range (min to max) of values observed during each period.

The seasonal cycles for the age groups are illustrated in Figure 5, divided into five 4-year periods from 2000 to 2019, showing the evolution of seasonal patterns. The year-by-year seasonal signals are detailed in Suppl. Figure 5. The newborns' signal remains consistent across all periods, with no pronounced winter peak and a modest plateau during summer. For the toddlers (6-24 months), the first three periods (2000 to 2011) show a distinct pattern with a strong winter peak, secondary weaker peaks in summer, and a nadir in the fall. This signal diminishes in the following period (2012-2015), where seasonal effects are less pronounced. The final period (2016-2019) reveals a striking shift in the seasonal profile, with a distinct peak in the fall and baseline levels for the rest of the year. The young children group (24-60 months) displays a similar pattern to the toddlers in the earlier periods, with a winter peak and a pronounced fall nadir, although both the peak and nadir are less distinct in the final period.

The synchrony between the two primary age groups (toddlers and young children), which together account for over 80% of the total KD cases, has been a notable feature throughout the study period. The seasonal patterns for these groups closely mirrored each other until 2016, as reflected in the year-by-year Pearson correlation coefficients consistently exceeding 0.7, except for 2012 (Suppl. Figure 5). However, this synchrony was disrupted in 2016, leading to negative correlations during 2017, 2018, and 2019. The two-way SDC analysis presented in Figure 6 further illustrates this disruption, showing that from 2000 to 2008, both age groups exhibited nearly identical seasonal signals, with distinct winter peaks more pronounced in the toddler group. By 2008-2009, a decoupling emerges, particularly evident during the fall and winter months, where the young children experienced a nadir while the toddlers peaked in December. Following 2010, the seasonal patterns realigned, though with noticeable differences, particularly the absence of the fall nadir in the toddler group. The post-2016 period marked a complete breakdown in synchrony, with sustained negative correlations until the COVID-19 pandemic in 2020, when both groups realigned due to a shared decrease in incidence. This sharp change in synchrony post-2016 suggests a significant shift in the factors influencing the epidemiology of KD between the toddler and young children groups, which now appear to be driven by potentially different mechanisms.

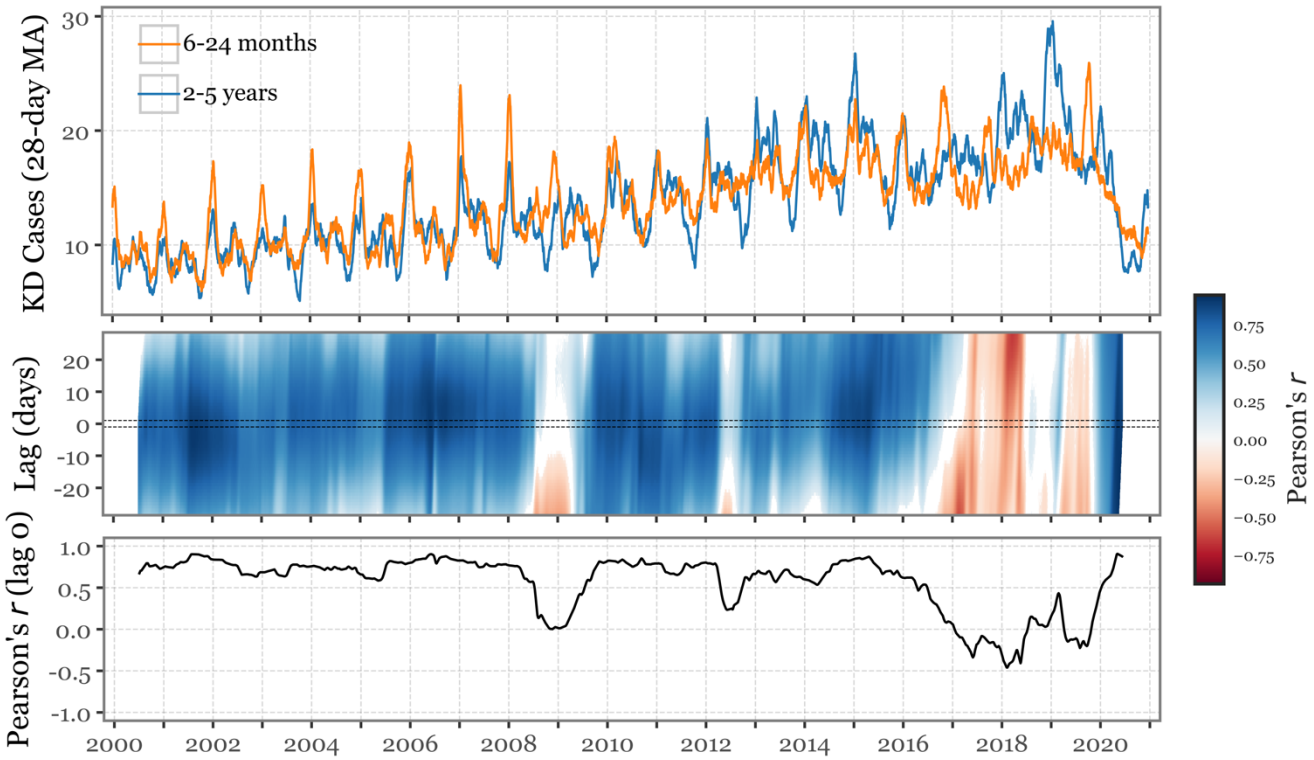


Figure 6. Long-term synchrony between KD cases in toddlers and young children (2000-2020). The top panel shows the 28-day moving average of KD cases for both age groups over the 20-year period. The middle panel is a heatmap depicting the 365-day rolling correlation between the two age groups, with blue shades indicating periods of positive transitory correlations and red shades indicating negative transitory correlations, computed over a range of lags from -30 to +30 days (using the younger group as the reference). The bottom panel illustrates the Pearson correlation coefficient at a lag of 0 days, highlighting periods of strong synchrony and periods of divergence between the two groups.

For comparison, and to highlight the lack of consistent synchrony between the KD incidence in both of these groups and the newborn group, a similar analysis is presented for all combinations of age groups in Suppl. Figure 6.

3.3 Prefectural-Spatial Analysis

Baseline KD incidence

In the preceding sections, KD incidence data was aggregated at the national level, which provided an enhanced signal-to-noise ratio for temporal pattern analysis but inherently obscured spatial variability. To investigate how the incidence of KD varies spatially, we conducted a prefectural-level analysis stratified by the same age groups (<6 months, 6-24 months, and 24-60 months). This approach allows us to assess two key aspects: (1) the degree of spatial clustering in regions of high and low KD incidence, and (2) whether the spatial distribution of KD incidence is consistent across different age groups.

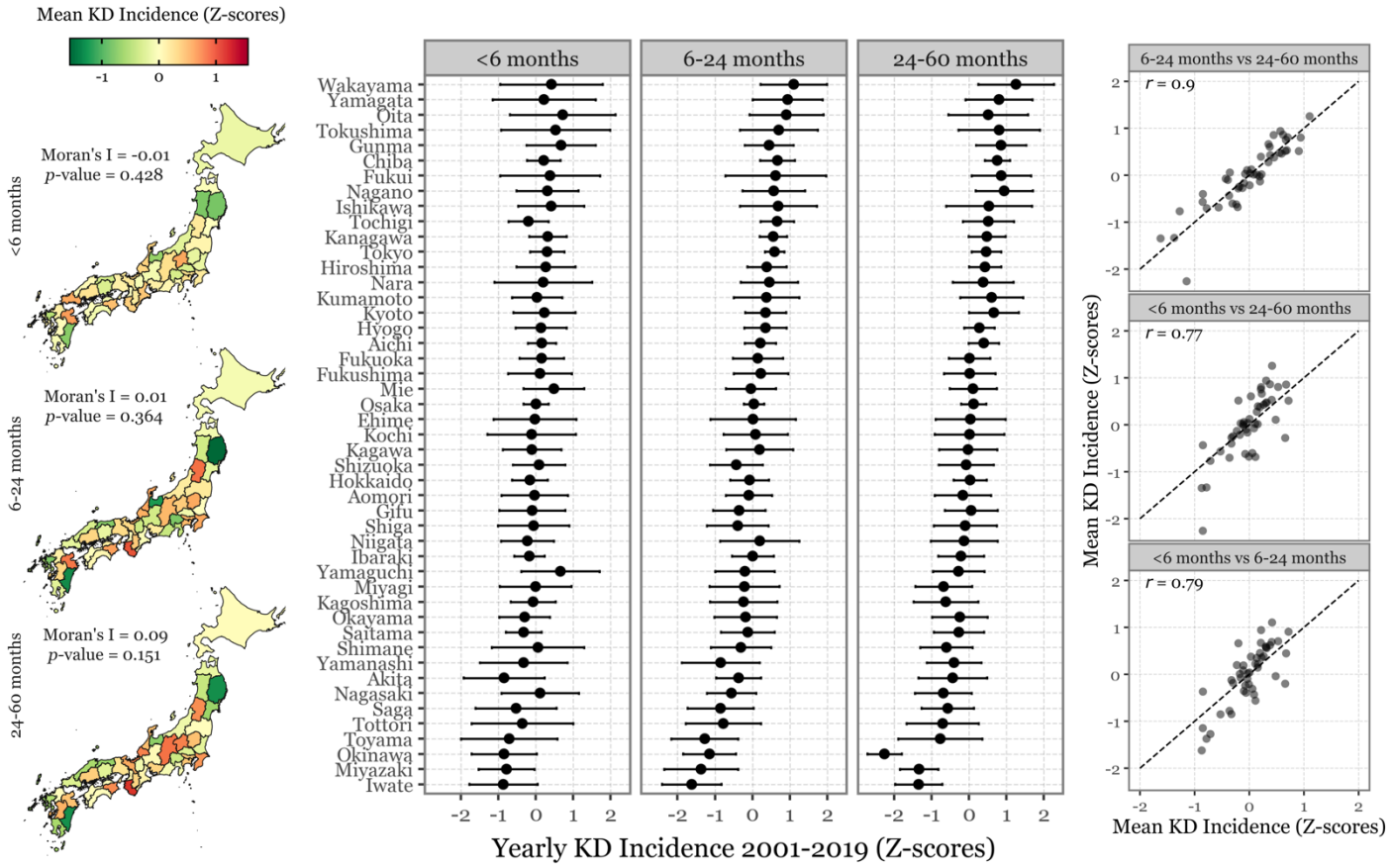


Figure 7. Prefectural distribution of baseline KD incidence stratified by age (2001-2019). On the left panel, maps of Japan illustrate the average standardized KD incidence across prefectures for each age range (<6 months, 6-24 months, and 24-60 months). The spatial autocorrelation for each age group is quantified using Moran's I statistic, with the corresponding p-values indicated in each map. The centre panel shows the average standardized KD incidence for each prefecture within the specified age ranges, accompanied by error bars representing the standard deviation of Z-scores across the study period. On the right panel, scatterplots depict pairwise comparisons of average Z-scores between the three age groups, showcasing the degree of spatial consistency of KD incidence patterns across these groups.

The findings shown in Figure 7 provide answers to both aspects: The analysis of the average standardized yearly incidences for each prefecture during the 2001 to 2019 period indicates no significant spatial autocorrelation (Global Moran's I of -0.01 for newborns, 0.01 for toddlers and 0.09 for young children), suggesting that high or low KD incidence rates in specific prefectures do not cluster spatially in a consistent manner across the country. However, despite the absence of spatial clustering,

there is a strong concordance in the prefectural incidence patterns across the different age groups. Prefectures with high KD incidence in one age group typically exhibit high incidence in the other age groups as well, and vice versa. This spatial consistency is particularly pronounced between toddlers and young children (Pearson's r = 0.9), but also remains notably high when comparing the newborn group with the other two age groups (r = 0.79 with toddlers and r = 0.77 with young children).

This pattern suggests that the factors influencing higher or lower overall KD incidences in each prefecture (which are independent of seasonal and trend-specific effects) consistently impact all age groups similarly. The robustness of these observations is also evident in the year-by-year analysis presented in Supplementary Table 1 and Supplementary Figures 7 to 10. The lack of spatial autocorrelation persists on a yearly basis, with only a few exceptions, and the stability of the spatial distribution of KD incidences across age groups remains consistent throughout the study period. This stability is particularly strong for the toddler and young children groups. However, the incidence of KD in newborns appears to diverge from the other two age groups after 2010, with correlations ranging from 0 to 0.3 in most years after that, compared to correlations above 0.5 before 2010. Given that newborn cases account for only 7.7% of the total cases, greater variability is expected, especially in prefectures with lower populations. Therefore, the global averages likely provide a more accurate estimate of the spatial consistency of KD incidence.

Prefectural KD seasonality

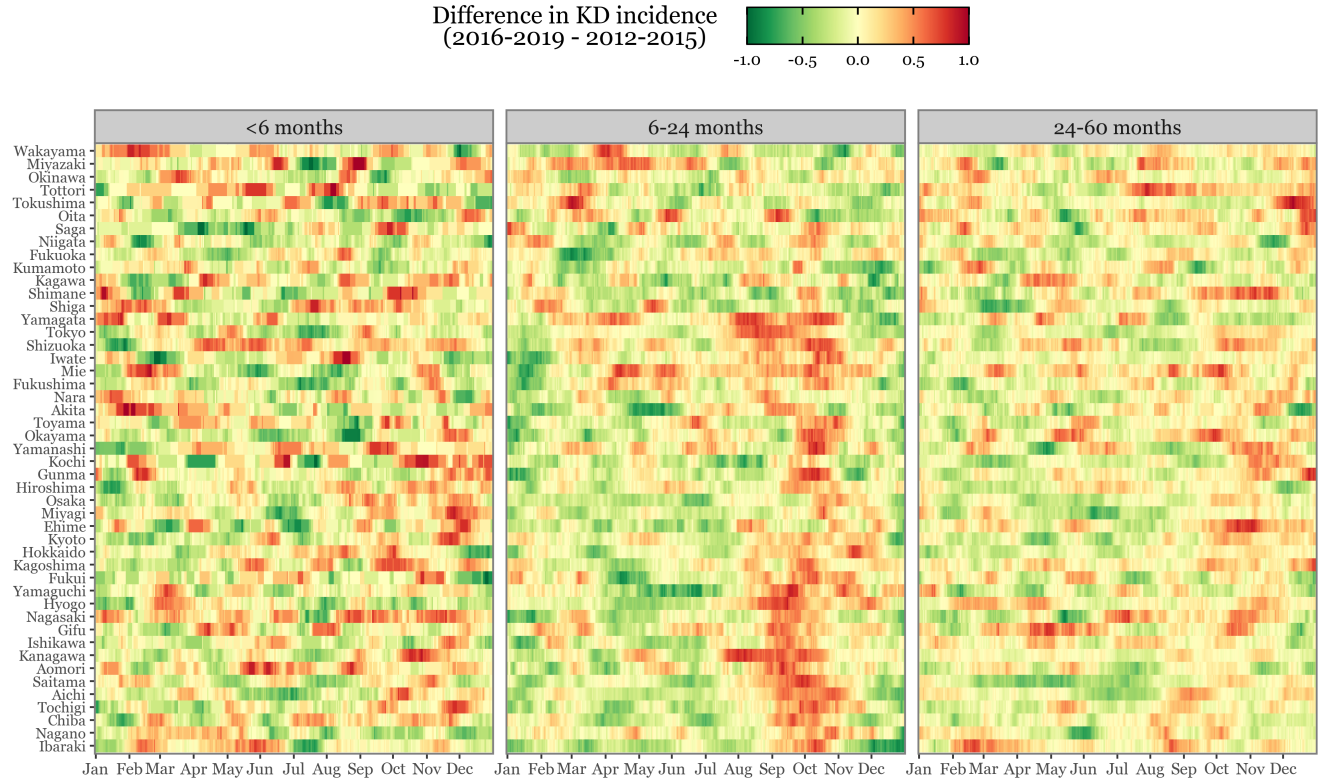


Figure 8. Prefecture and age-stratified difference of seasonal cycles pre and post 2016. The heatmaps show the difference in the average yearly min-max normalized seasonal cycles for each prefecture and age group, comparing the periods 2016-2019 and 2012-2015. Red indicates an increase in KD incidence during that specific time of year in the 2016-2019 period compared to 2012-2015, while green indicates a decrease in incidence for that time of year.

While a clear change in the seasonal pattern of KD incidence in toddlers after 2016 was observed at the national level (Figure 5-6), we expanded the analysis to each of the 47 prefectures to determine whether these changes were consistent across all regions or driven by specific, densely populated urban areas. We conducted an age-stratified analysis of the seasonal cycles, comparing the 2012-2015 period with the 2016-2019 period across all prefectures. As shown in Figure 8, this comparison reveals shifts in the average standardized yearly cycles after 2016, particularly in the toddler group (6-24 months). No consistent pattern emerges across prefectures for the newborns or young children, but most prefectures show a notable increase in incidence for the toddler group from August to November during the 2016-2019 period. Suppl. Figures 11-14 show the actual incidences in both periods for each of the age groups, and Suppl. Figures 15-17 and Suppl. Tables 2-4 show the diffs for the full year. The extent of this change becomes even more evident once mapped (Figure 9). While most prefectures show little to no change during these months for the young children group, the newborns exhibit inconsistent changes (in either direction). However, the toddlers' increase is strong and consistent across the whole map, especially in September and October, with few exceptions in some of Kyushu's prefectures (the southernmost island).

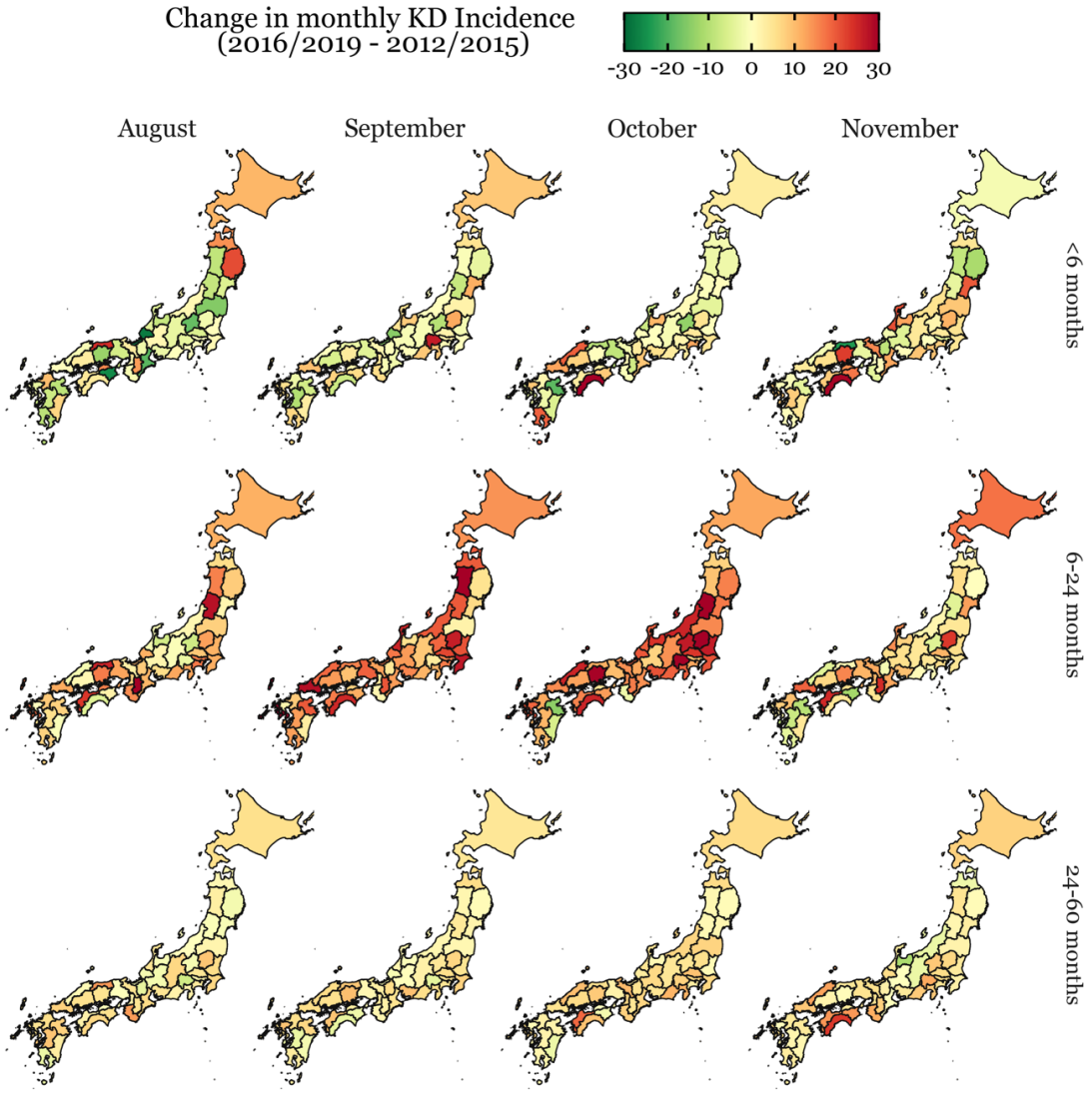


Figure 9. Map of the seasonal change in KD incidence by prefectures and age. Choropleth maps showing either an increase (red) or decrease (green) in the monthly KD incidence in 2016-2019 compared to the 2012-2015 period.

3.4 Environmental factors

Given the substantial seasonal changes observed in a significant proportion of KD patients (specifically toddlers aged 6 to 24 months, comprising over 40% of total cases), we investigated whether these changes might be attributable to shifts in the dynamics of potential environmental triggers. These triggers are hypothesized to influence the onset and variability of KD. Our analysis focused on three key aspects: (a) changes in air mass transport patterns, (b) variations in local environmental factors (such as vegetation, pollution, and meteorology), and (c) alterations in the source regions of air masses at moments of high-KD, particularly cropland areas located in Northeast China as shown in (12,13).

Winds and tropospheric transport sources

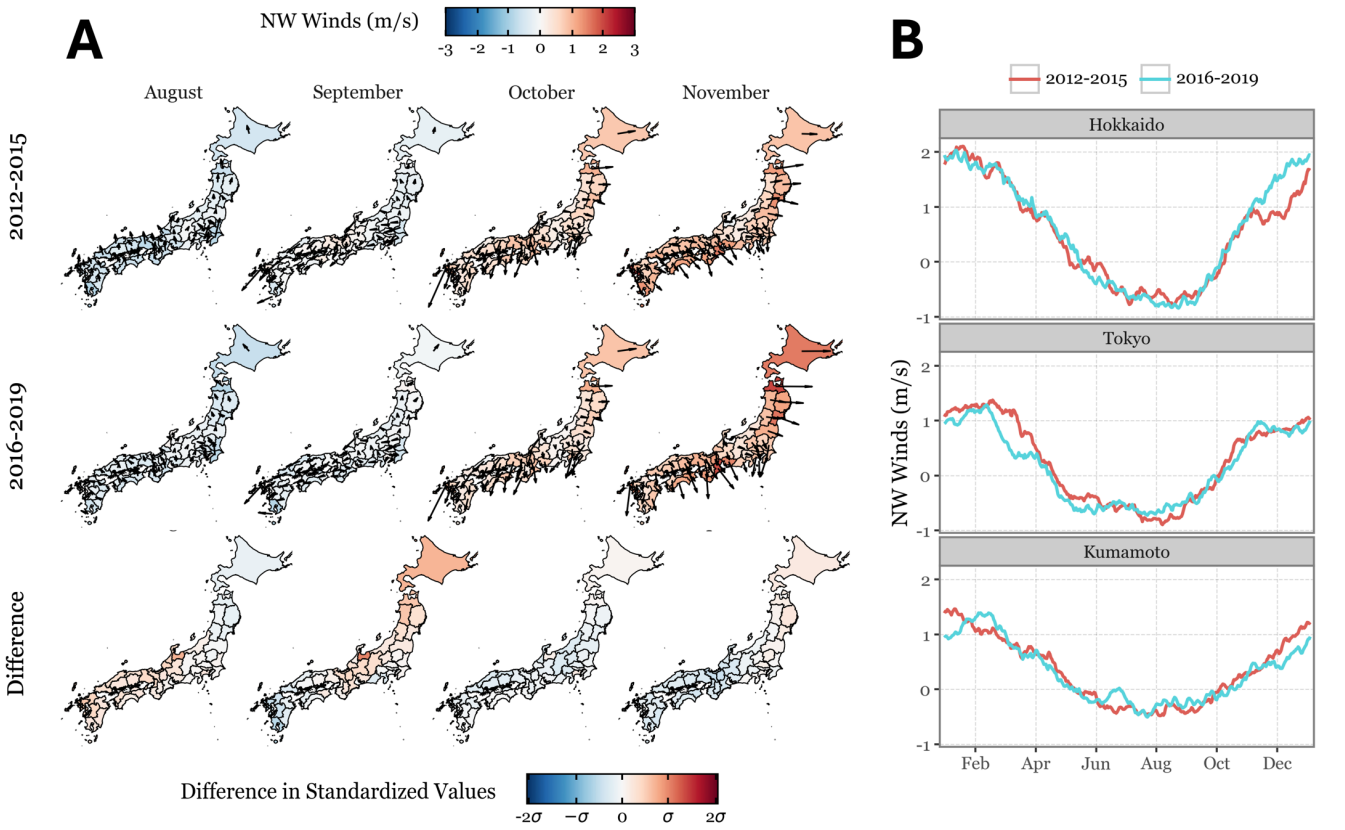


Figure 10. Northwesterly winds by prefecture and period. Panel A presents maps of Japan illustrating the average monthly northwesterly wind component for each prefecture during the periods 2012-2015 (top) and 2016-2019 (middle). Blue shades indicate winds primarily blowing from the southeast, while red shades indicate winds from the northwest. The direction and length of the arrows represent the average wind direction and speed. The bottom row shows the standardized difference in northwesterly wind between the two periods. In panel B, the full seasonal cycle of northwesterly winds is depicted for three prefectures representing the latitudinal span of Japan: Hokkaido (north), Tokyo (central), and Kumamoto (south). The red line corresponds to the average seasonal cycle for the 2012-2015 period, while the blue line represents the same for the 2016-2019 period.

The patterns in the northwesterly wind component (as detailed in the Methods) are illustrated in Figure 10. The seasonal wind patterns remained consistent across the periods of 2012–2015 and 2016–2019, with minimal variation in the direction and strength of winds throughout the year. In the three representative prefectures (Hokkaido in the north, Tokyo in the center, and Kumamoto in the south), the seasonal patterns show little change between periods. The prefecture-wide average wind data

indicate some moderate differences in the standardized values, none of which are significant enough to suggest a meaningful shift in the wind patterns between the two periods.

The composite results of the air source assignment are summarized for the 2012-2015 and 2016-2019 periods in Figure 11, with inter-period differences displayed for the months of August to November, aligned with the emergence of the post-2016 seasonal peak in KD cases within the toddler age group. The complete annual patterns can be found in Suppl. Fig. 18. As illustrated in both the full-year panels and the previously discussed northwesterly wind component, the air masses arriving over the Japanese archipelago exhibit a cyclical pattern. During winter months, air masses predominantly originate from the northwest, i.e., from regions such as NE China and the Korean Peninsula, while southeasterly winds, originating from the Pacific Ocean, dominate during summer. The transition from southeast to northwest winds occurs during the period from August to November.

The comparison between periods shows minimal changes overall, with a slight but observable shift in October. Specifically, there is a reduced frequency of air masses traversing the Korean Peninsula, coupled with an increased frequency of winds originating from Manchuria. Despite these small adjustments, the overall structure of air source regions remains largely consistent across the two periods.

Overall, the comparison of air source regions between the 2012-2015 and 2016-2019 periods suggests that the observed shift in the seasonal cycle of KD cases in the toddler group is unlikely to be driven by any changes in wind transport patterns. As demonstrated in the inter-period analysis, the structure of air mass sources remains largely stable, with only minor variations observed. These small shifts, such as the reduced frequency of winds passing over the Korean Peninsula and the slight increase in winds from Manchuria during October, do not appear sufficient to explain the observed seasonal changes in KD incidence. These minor variations do not appear to be able to account for the seasonal shift in KD incidence, suggesting instead that the change may stem from factors related to the content of the air masses and not to their transport.

Vegetation Index

The Enhanced Vegetation Index (EVI) was employed to track vegetation dynamics in both the local Japanese prefectures and the region of Northeast China, an area that has been shown to be associated with increased KD incidence when winds arriving to the Japanese archipelago are sourced from there (13). Vegetation indexes typically display distinct seasonal patterns, with shifts often indicating changes in land use, such as agricultural expansion, deforestation, or urbanization. These shifts may also reflect alterations in crop type or climatic conditions affecting vegetation health. In this analysis, we evaluated the temporal series of EVI to detect any changes in seasonal cycles between the 2012–2015 and 2016–2019 periods. Specifically, we examined changes in the timing of peak vegetation (to identify possible phase shifts), the total amplitude of the seasonal cycle, and the maximum vegetation index values achieved during each period.

The results of the EVI comparison show mostly negative findings regarding changes in vegetation dynamics between the 2012-2015 and 2016-2019 periods. In the Japanese prefectures (Figure 12), few significant differences are observed. For example, only three prefectures exhibit significant changes in amplitude (Aomori, Gifu, and Hokkaido), one prefecture shows a shift in peak time (Yamanashi), and another (Toyama) presents a significant increase in maximum EVI value. These isolated differences

suggest limited changes in land-use or vegetation health across Japan during the studied periods. Meanwhile, the results for the Northeast China region, shown in Figure 13, indicate no significant changes in any of the metrics (amplitude, peak time, and peak value) over the same periods. This further reinforces the conclusion that there were no major shifts in vegetation dynamics that could explain the observed changes in KD incidence during this time.

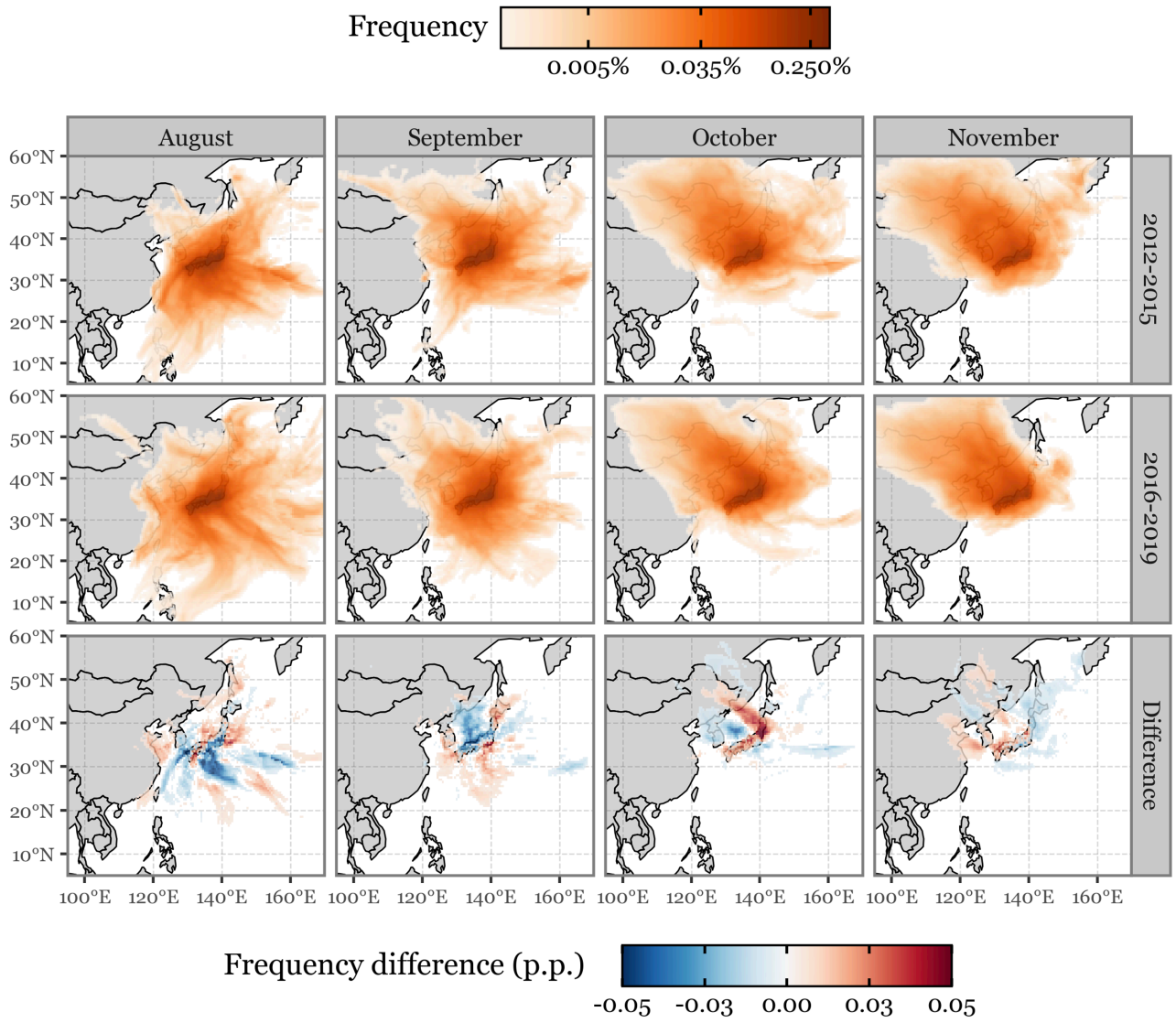


Figure 11. Monthly composite of air source regions for Japanese prefectures by period. The first two rows present composite images of 96-hour backward particle dispersion trajectories originating from the centroids of all Japanese prefectures (excluding Okinawa). Darker shaded areas represent regions with a higher frequency of air parcels arriving at the Japanese prefectures, indicating the likely source regions. Each column corresponds to a specific month (August to November), with the top row showing data from the 2012-2015 period and the middle row from the 2016-2019 period. The bottom row illustrates the difference in source region frequency between these two periods, with blue areas indicating regions where air parcels were more common during the 2012-2015 period, and red areas highlighting regions more prevalent in the 2016-2019 period. The scale of frequency difference is expressed in percentage points (p.p.).

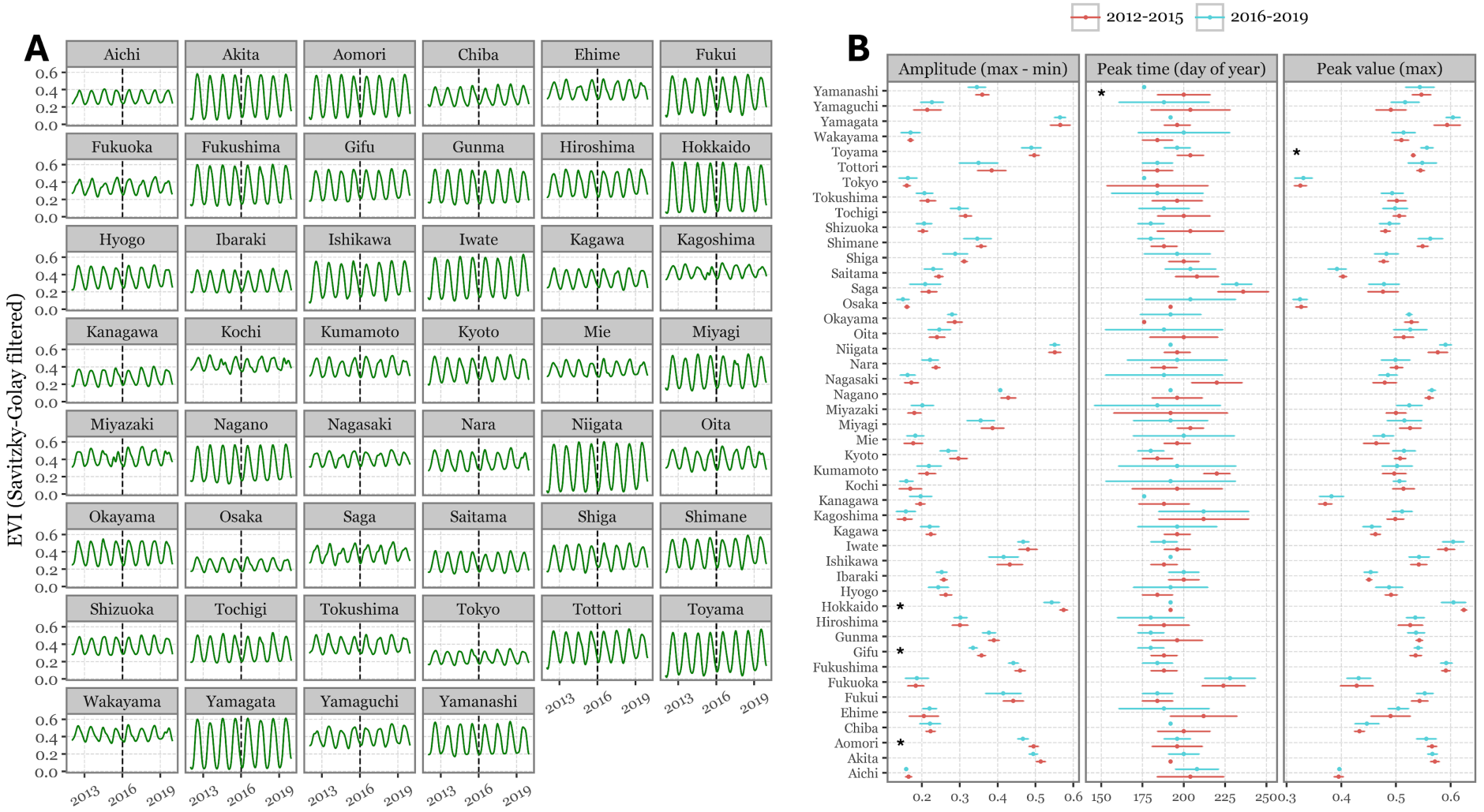


Figure 12. EVI values for Japanese prefectures and changes between the 2012-2019 period. Panel A shows the time series of EVI values averaged for each pixel within each prefecture, derived from the MOD13A2 dataset and smoothed using a Savitzky-Golay filter. Panel B displays error bars representing the mean and standard deviation for three metrics defining the seasonal cycle of vegetation: amplitude (difference between max and min EVI), peak time (day of the year when the EVI peaks), and peak value (maximum EVI). The red error bars correspond to the 2012-2015 period, while the blue error bars represent the 2016-2019 period. Comparisons that showed statistically significant differences are marked with asterisks (see Suppl. Table 5 for statistical test details)

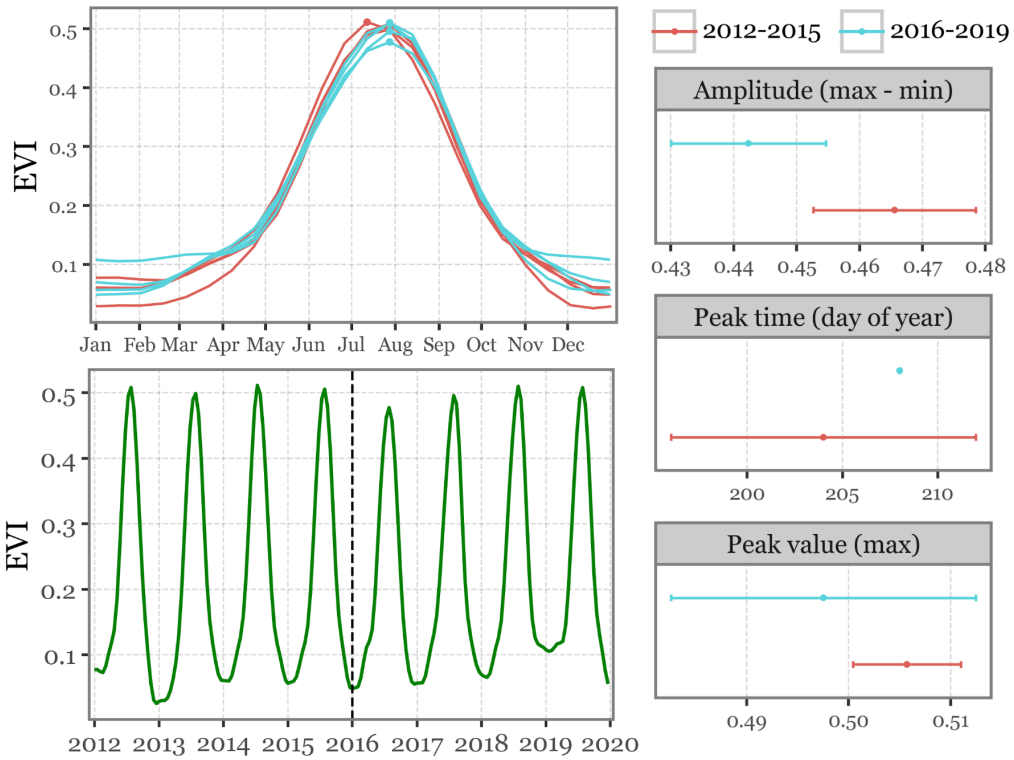


Figure 13. EVI values for Northeast China and changes between the 2012-2019 period. The top left panel displays the daily interpolated EVI values averaged over the Northeast China bounding box (as defined in the methods), plotted as a function of the day of the year. The bottom left panel presents the same data as a continuous time series for the entire period. The right panels show error bars depicting the mean value and standard deviation for three metrics: amplitude (max-min), peak timing (day of year), and peak value (max). Red error bars represent the 2012-2015 period, while blue error bars indicate the 2016-2019 period. Absence of error bars implies identical data points across both periods.

Air Pollution and Meteorology

The results showcasing the monthly differences in the pre-2016 (2012-2015) and post-2016 (2016-2019) periods for the Japanese prefectures are shown in Figure 14 for meteorology and Figure 15 for air pollution.

The temperature changes between the two periods show minimal variation across Japan, with no significant shifts in the spatial patterns. This aligns with the typically low inter-annual variation in temperature observed in the region. The expected seasonal cycles remain stable, with no notable deviations. In contrast, the precipitation changes are more pronounced, particularly in the southwestern regions such as Kyushu during September and October. These changes likely reflect the susceptibility of Japan to seasonal shifts in rainfall patterns, including the influence of typhoons, which typically bring higher levels of precipitation during these months. Although these shifts in precipitation are notable, they are consistent with the region's known climatic instability.

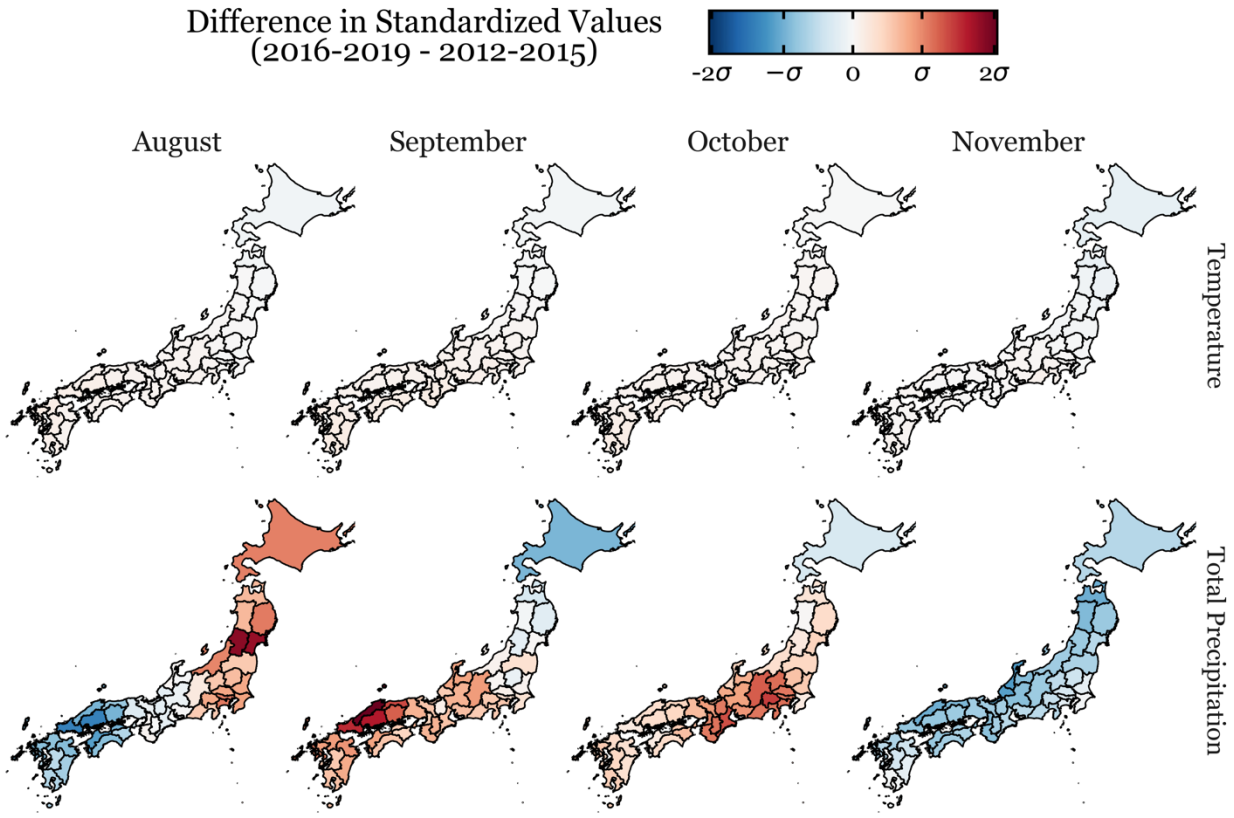


Figure 14. Difference by period and month for meteorological variables. The maps illustrate the standardized differences in temperature (top row) and total precipitation (bottom row) between the 2016-2019 and 2012-2015 periods, for the months of August, September, October, and November. Blue shading indicates regions where the variable was lower in the 2016-2019 period compared to 2012-2015, while red shading represents regions where the variable was higher. The standardization was performed at the prefecture level across the entire study period (2012-2019) to account for local variability.

The changes in air pollution variables between the two periods show a noticeable reduction in most pollutants across Japan (Figure 15). For nitrogen oxides (NO, NO_2, and NO_x), there is a widespread and consistent decrease for all months and regions of the archipelago. $PM_{2.5}$ levels exhibit a decrease across all regions and months, but the most significant changes are concentrated in the northwestern half of Japan in August, with more moderate but consistent decreases throughout the other months. SO_2 also shows notable decreases across most of the country, though some prefectures in the south (Saga, Shimane, Oita) show an increase in August, and Saitama in the northwest shows increases in both October and November. Nevertheless, the general pattern clearly indicates a reduction in SO_2 levels across the entire country.

These reductions in air pollutants, particularly nitrogen oxides and particulate matter, point to an overall improvement in air quality when comparing the 2016-2019 period to the earlier one. However, it is highly unlikely that the shift observed in the seasonal signal of KD after 2016 could be attributed to a decrease in air pollution levels. Instead, it suggests that the changes in air pollution are not directly responsible for the shift in the KD seasonality.

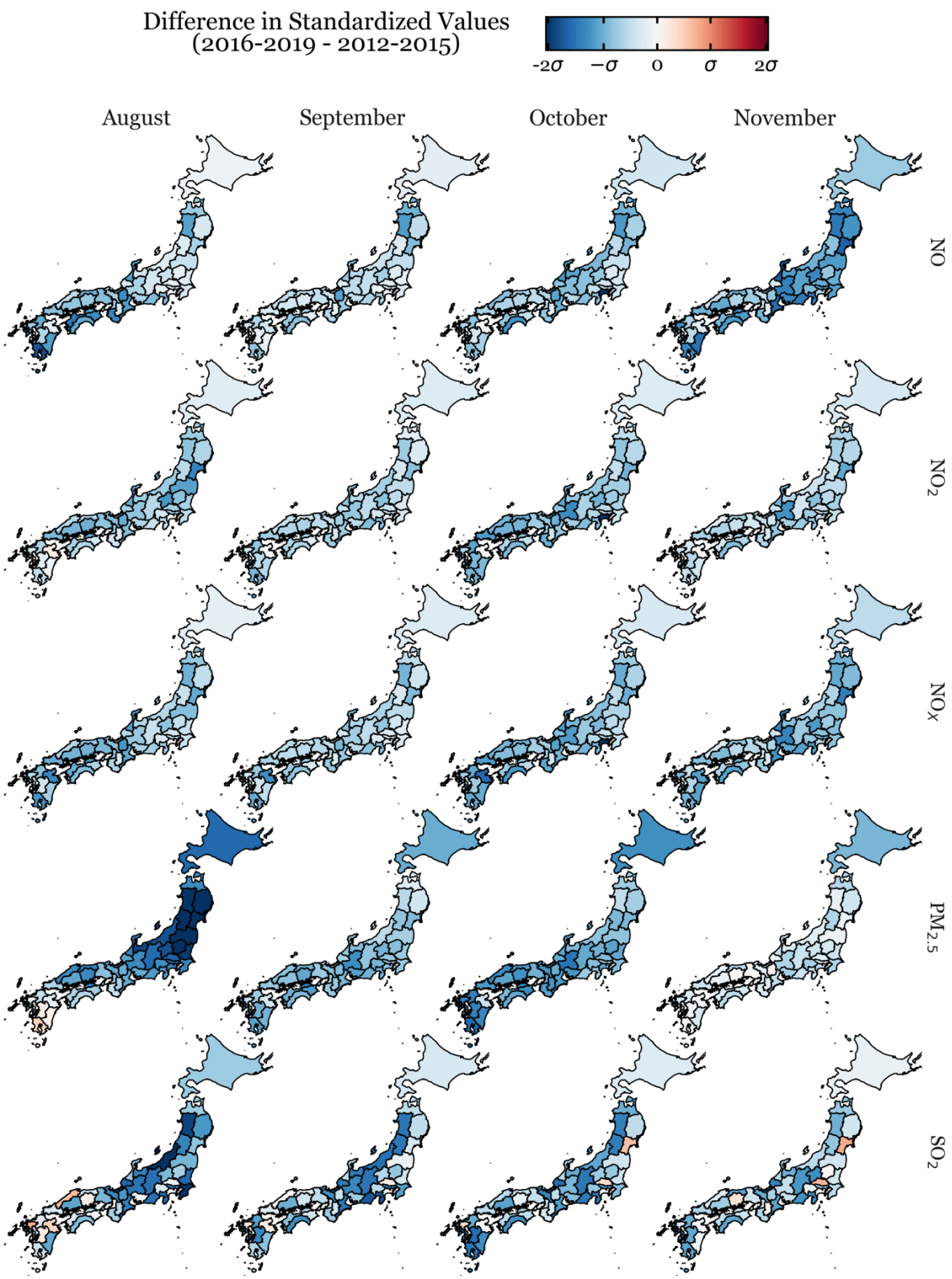


Figure 15. Difference by period and month for air pollution variables. Idem to Figure 14 but for air pollution variables: in order, from top to bottom: nitrogen oxide (NO), nitrogen dioxide (NO_2), nitrogen oxides (NO_x), particulate matter (PM2.5), and sulfur dioxide (SO_2)

4. Discussion

The spatiotemporal and age-stratified analysis of KD incidence across Japan revealed both expected patterns and striking anomalies. The seasonal variation in KD incidence has long been a feature of the disease's epidemiology, with winter peaks commonly observed across all age groups (5,8,17). Our analysis of the 2000–2020 data confirmed the overall consistency of these seasonal cycles for the majority of the child population. Spatially, KD incidence followed a stable geographical distribution, with no significant clustering or shifts across the Japanese archipelago, and similar patterns among age groups. This reinforces the notion that KD's long-term trends and seasonal behaviour have remained stable across both time and regions. This broad consistency contrasts sharply with a dramatic exception: the temporal dynamics of the toddler group (6–24 months).

In this group, a clear shift in KD seasonality emerged after 2016, marked by the appearance of a new peak in cases from August to November and the disappearance of the previously common winter peak. This shift was particularly striking, as it was confined to toddlers, who represent a substantial proportion of KD cases, comprising over 40% of the total. The specificity of this shift raises critical questions: Why did this change occur suddenly, and why only in toddlers? While seasonal shifts are not uncommon in disease epidemiology, the precise and isolated nature of this change suggests a unique driver is involved, one affecting toddlers differently from other age groups.

Several hypotheses could explain this sudden alteration in the seasonal cycle, but our exploration of various environmental factors did not yield a clear answer. Our comparison of pre- and post-2016 environmental data, particularly wind patterns and air source regions, revealed consistent cyclical patterns with minimal variation. This suggests that the mechanisms transporting aerosols and other pollutants did not experience significant shifts between the two periods, ruling out major changes in air mass movement as a driving factor. Similarly, meteorological variables, including temperature and precipitation, exhibited typical variability, with no notable anomalies aligning with the observed shift in KD seasonality. Moreover, vegetation dynamics, as captured by the EVI, remained largely consistent over the study period, with only minor shifts detected in isolated regions.

The examination of air pollution, particularly nitrogen oxides and particulate matter, revealed a general improvement in air quality over time. No distinct changes were observed during the fall season relative to the rest of the year, making it unlikely that pollution played a direct role in driving the post-2016 increase in cases. Although literature has generally found weak or no associations between KD incidence and common air pollutants (22–24) the hypothesis that a reduction in air pollution could paradoxically lead to an increase in KD cases remains difficult to substantiate.

Given the absence of strong evidence from environmental factors, the possibility of a more complex, indirect driver of this seasonal shift must be considered. Potential contributors may include newly circulating viral or bacterial infections, changes in pediatric healthcare practices, or nationwide shifts in childcare patterns that are specific to this age group. To date, no evidence of these factors has been detected. Further studies, particularly those investigating potential biological or epidemiological triggers, are critical to identifying the true cause behind this phenomenon. If an external driver is involved, it likely requires a more nuanced and sophisticated approach to detection—one that goes

beyond the environmental parameters currently explored. This could include behavioural factors or a deeper examination of the composition of the air masses, which may carry various potential allergens, such as pollen, microbial life (25), or pro-inflammatory metals (26), and extend beyond typical air pollution. A more detailed temporal analysis may be necessary to fully capture the complex dynamics of this disease.

The decorrelation of the toddler group from the other age groups is unprecedented in KD epidemiology and adds further weight to the hypothesis of a multifactorial etiology (27). While the general seasonal and spatial patterns of KD have remained consistent across age groups, this sudden divergence highlights the likelihood that different drivers or risk factors may be influencing specific age cohorts at different times. Such a distinct shift in toddlers suggests that age-specific factors, whether environmental, biological, or immunological, could play a pivotal role in shaping the incidence of KD. This finding emphasizes the importance of considering age-stratified differences in future studies to better understand the complex interactions that underlie KD's triggers. Only by accounting for these variations across the population can we begin to unravel the multifactorial nature of the disease.

5. Conclusions

In conclusion, our study was, to our knowledge, the first to explore the dynamics of KD to this level of depth at the spatial, temporal and sex/age stratification. The key results are the following:

- KD incidence consistently shows a higher prevalence in males compared to females, at around a 1.4:1 ratio. This sex difference remained stable across time, regions, and age groups, with both sexes showing identical temporal dynamics.

- The general trend in KD incidence has been increasing over the past two decades, and it is identical for both the toddler and young children groups. The infant age group (<6 months), however, grew until 2010 but its incidence has plateaued since then.

- While the majority of age groups maintained a stable seasonal pattern with a winter peak, the toddler group exhibited a striking shift after 2016, developing a new seasonal peak from August to November. This shift in seasonality suggests age-specific factors that may influence disease onset, which warrants further investigation.

- Despite the increase in incidence, KD's spatial distribution across Japan remained stable, with no significant clustering or hotspots over time. The spatial distribution of the prefectural incidences of KD is consistent across all age groups, pointing toward a common local factor affecting all age groups together, in contrast to the potential drivers of seasonality, which must differ by age-group.

- The analysis of various environmental factors, including wind patterns, temperature, precipitation, and air pollution, did not reveal any significant changes that could explain the observed shift in KD seasonality. Meteorological variables followed typical variability patterns, and air quality improved with reductions in all pollutants studied. These findings suggest that the seasonal shift in KD cannot be directly attributed to large-scale environmental changes, pushing the focus toward more specific drivers of the disease.

6. References

1. Burns JC, Glodé MP. Kawasaki syndrome. The Lancet. 2004 Aug 7;364(9433):533–44.

2. Burns JC, Kushner HI, Bastian JF, Shike H, Shimizu C, Matsubara T, et al. Kawasaki Disease: A Brief History. Pediatrics. 2000 Aug 1;106(2):e27.

3. Yanagawa H, Nakamura Y, Kawasaki T, Shigematsu I. Nationwide epidemic of Kawasaki disease in Japan during winter of 1985-86. Lancet Lond Engl. 1986 Nov 15;2(8516):1138–9.

4. Ae R, Makino N, Kuwabara M, Matsubara Y, Kosami K, Sasahara T, et al. Incidence of Kawasaki Disease Before and After the COVID-19 Pandemic in Japan: Results of the 26th Nationwide Survey, 2019 to 2020. JAMA Pediatr [Internet]. 2022 Oct 17 [cited 2022 Oct 20]; Available from: https://doi.org/10.1001/jamapediatrics.2022.3756

5. Burns JC, Cayan DR, Tong G, Bainto EV, Turner CL, Shike H, et al. Seasonality and Temporal Clustering of Kawasaki Syndrome. Epidemiology. 2005 Mar;16(2):220.

6. Manlhiot C, O'Shea S, Bernknopf B, LaBelle M, Chahal N, Dillenburg RF, et al. Epidemiology of Kawasaki Disease in Canada 2004 to 2014: Comparison of Surveillance Using Administrative Data vs Periodic Medical Record Review. Can J Cardiol. 2018 Mar 1;34(3):303–9.

7. Sánchez-Manubens J, Antón J, Bou R, Iglesias E, Calzada-Hernandez J. Paediatric rheumatology Incidence, epidemiology and clinical features of Kawasaki disease in Catalonia, Spain. Clin Exp Rheumatol. 2016;34(97):S139–44.

8. Burns JC, Herzog L, Fabri O, Tremoulet AH, Rodó X, Uehara R, et al. Seasonality of Kawasaki Disease: A Global Perspective. PLOS ONE. 2013 Sep 18;8(9):e74529.

9. Nelson RJ, Demas GE. Seasonal Changes in Immune Function. Q Rev Biol. 1996 Dec;71(4):511–48.

10. Khoo AL, Chai LYA, Koenen HJPM, Sweep FCGJ, Joosten I, Netea MG, et al. Regulation of cytokine responses by seasonality of vitamin D status in healthy individuals. Clin Exp Immunol. 2011 Apr 1;164(1):72–9.

11. Schmidt CW. Pollen Overload: Seasonal Allergies in a Changing Climate. Environ Health Perspect. 2016 Apr;124(4):A70–5.

12. Rodó X, Ballester J, Cayan D, Melish ME, Nakamura Y, Uehara R, et al. Association of Kawasaki disease with tropospheric wind patterns. Sci Rep. 2011 Nov 10;1(1):152.

13. Rodó X, Curcoll R, Robinson M, Ballester J, Burns JC, Cayan DR, et al. Tropospheric winds from northeastern China carry the etiologic agent of Kawasaki disease from its source to Japan. Proc Natl Acad Sci U S A. 2014 Jun 3;111(22):7952–7.

14. Rodó X, Ballester J, Curcoll R, Boyard-Micheau J, Borràs S, Morguí JA. Revisiting the role of environmental and climate factors on the epidemiology of Kawasaki disease. Ann N Y Acad Sci. 2016 Oct 1;1382(1):84–98.

15. Giefing-Kröll C, Berger P, Lepperdinger G, Grubeck-Loebenstein B. How sex and age affect immune responses, susceptibility to infections, and response to vaccination. Aging Cell. 2015 Jun;14(3):309–21.

16. Klein SL, Flanagan KL. Sex differences in immune responses. Nat Rev Immunol. 2016 Oct;16(10):626–38.

17. Ozeki Y, Yamada F, Saito A, Kishimoto T, Yashiro M, Makino N, et al. Epidemiologic features of Kawasaki disease distinguished by seasonal variation: an age-specific analysis. Ann Epidemiol. 2018 Nov 1;28(11):796–800.

18. DeHaan LL, Copeland CD, Burney JA, Nakamura Y, Yashiro M, Shimizu C, et al. Age-Dependent Variations in Kawasaki Disease Incidence in Japan. JAMA Netw Open. 2024 Feb 6;7(2):e2355001.

19. Stein AF, Draxler RR, Rolph GD, Stunder BJB, Cohen MD, Ngan F. NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. Bull Am Meteorol Soc. 2015 Dec 1;96(12):2059–77.

20. Rodo X, Rodriguez Arias MA. A new method to detect transitory signatures and local time/space variability structures in the climate system: The scale-dependent correlation analysis. Clim Dyn. 2006 Oct 1;27:441–58.

21. Rodríguez-Arias MÀ, Rodó X. A primer on the study of transitory dynamics in ecological series using the scale-dependent correlation analysis. Oecologia. 2004 Mar 1;138(4):485–504.

22. Jung CR, Chen WT, Lin YT, Hwang BF. Ambient Air Pollutant Exposures and Hospitalization for Kawasaki Disease in Taiwan: A Case-Crossover Study (2000–2010). Environ Health Perspect. 2017 Apr;125(4):670–6.

23. Kwon D, Choe YJ, Kim S, Chun BC, Choe S. Ambient Air Pollution and Kawasaki Disease in Korean Children: A Study of the National Health Insurance Claim Data. J Am Heart Assoc. 2022 May 3;11(9):e024092.

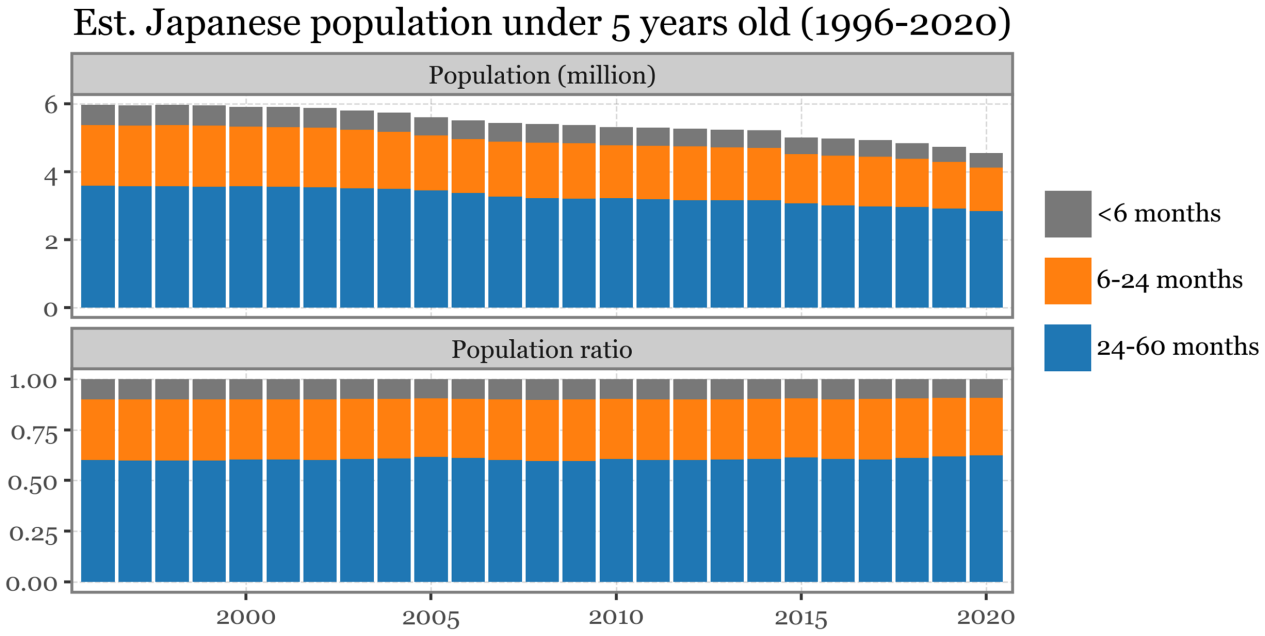
24. Zeft AS, Burns JC, Yeung RS, McCrindle BW, Newburger JW, Dominguez SR, et al. Kawasaki Disease and Exposure to Fine Particulate Air Pollution. J Pediatr. 2016 Oct 1;177:179-183.e1.

25. Rodó X, Pozdniakova S, Borràs S, Matsuki A, Tanimoto H, Armengol MP, et al. Microbial richness and air chemistry in aerosols above the PBL confirm 2,000-km long-distance transport of potential human pathogens. Proc Natl Acad Sci. 2024 Sep 17;121(38):e2404191121.

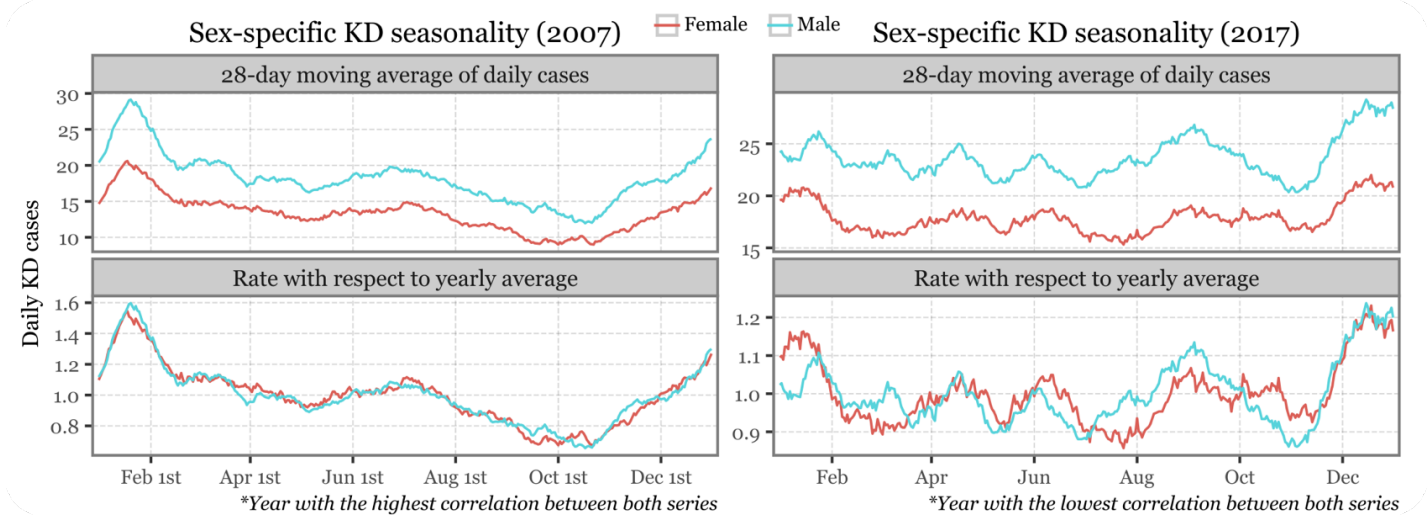
26. Rodó X, Navarro-Gallinad A, Kojima T, Morguí JA, Borràs S, Fontal A. Sub-weekly signatures relate ultrafine aerosols enriched in metals from intensive farming and urban pollution to Kawasaki disease. Environ Res Lett. 2023 Jun;18(7):074011.

27. Burns JC. The etiologies of Kawasaki disease. J Clin Invest [Internet]. 2024 Mar 1 [cited 2024 Sep 28];134(5). Available from: https://www.jci.org/articles/view/176938#B58

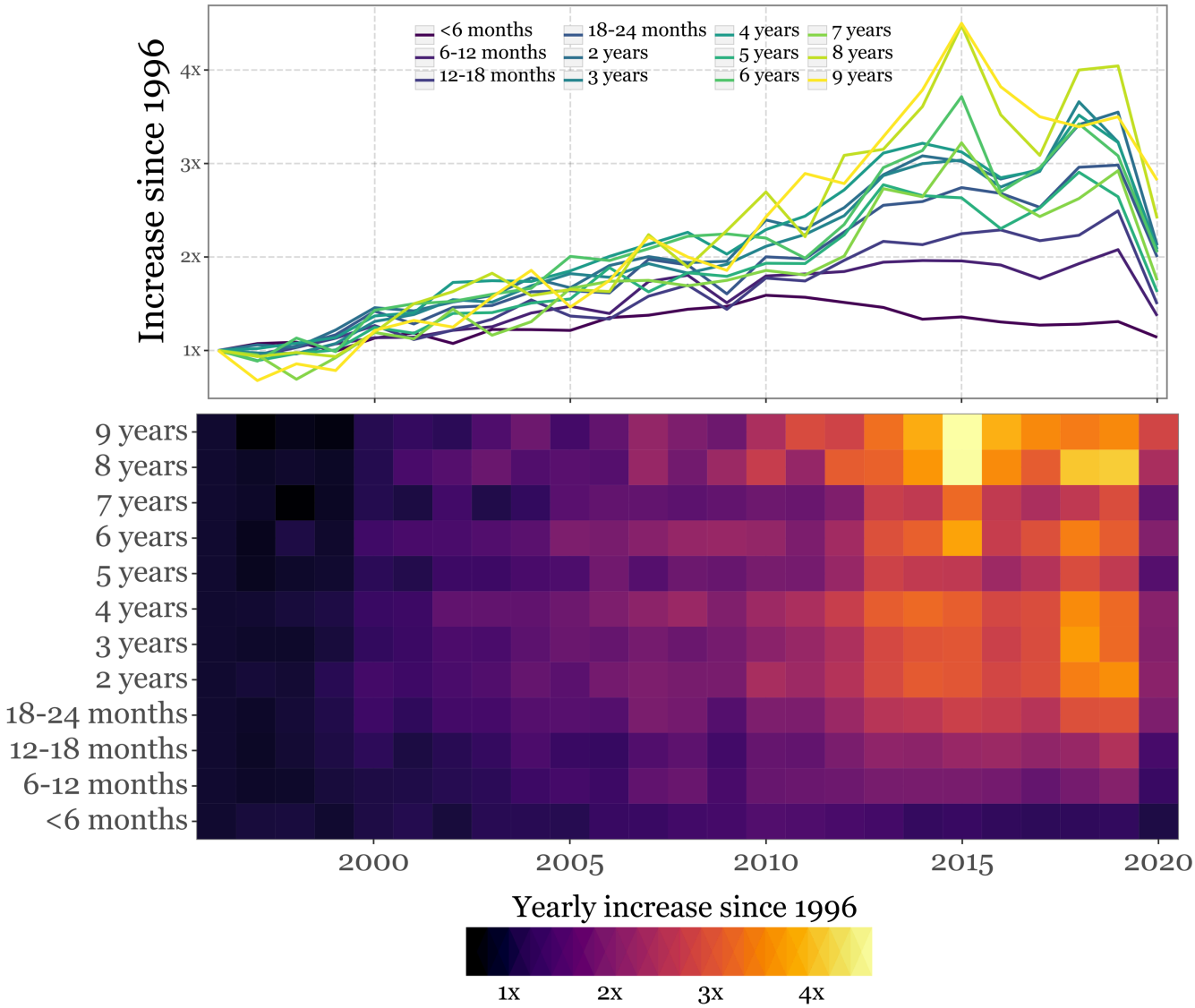
Supplementary Figures and Tables



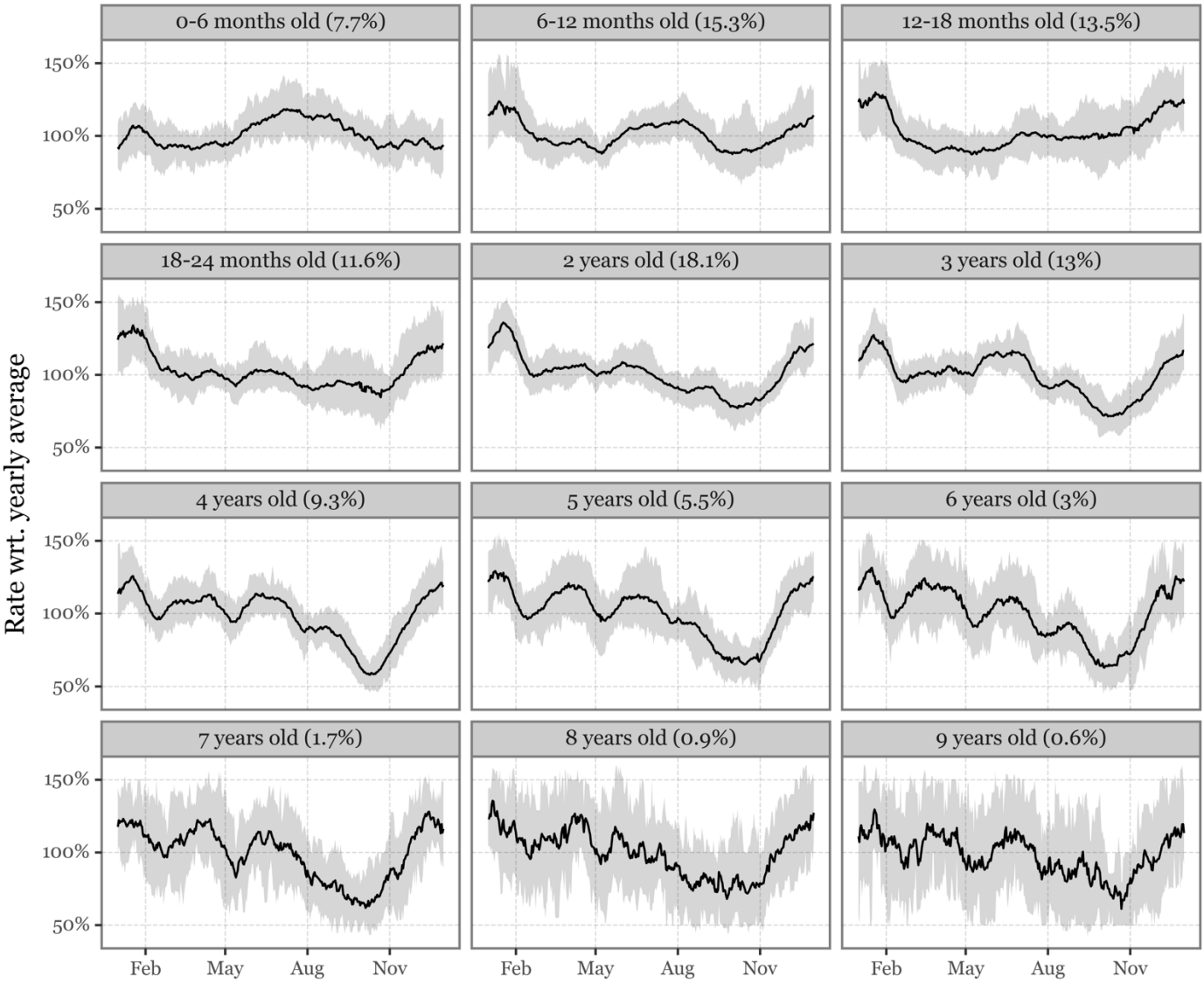
Suppl. Figure 1. Population estimates for the selected under 5 year old age groups. These population ratios are then used to estimate the prefecture-level populations of each of the groups from the 5-year age groups that are provided by the Portal of Official Statistics of Japan.



Suppl. Figure 2. Sex-specific KD seasonality in 2007 and 2017. The top panels show the 28-day moving averages of daily KD cases for male and female patients under 5 years old. The bottom panels display the same data normalized as a rate with respect to the yearly average. The left column represents data from 2007, which had the highest correlation between male and female seasonal patterns, while the right column represents data from 2017, the year with the lowest correlation between the two series. The left column is an example of a near perfect synchrony, while the right column showcases that the year with the lower synchrony still has a high level of coherence between the series.

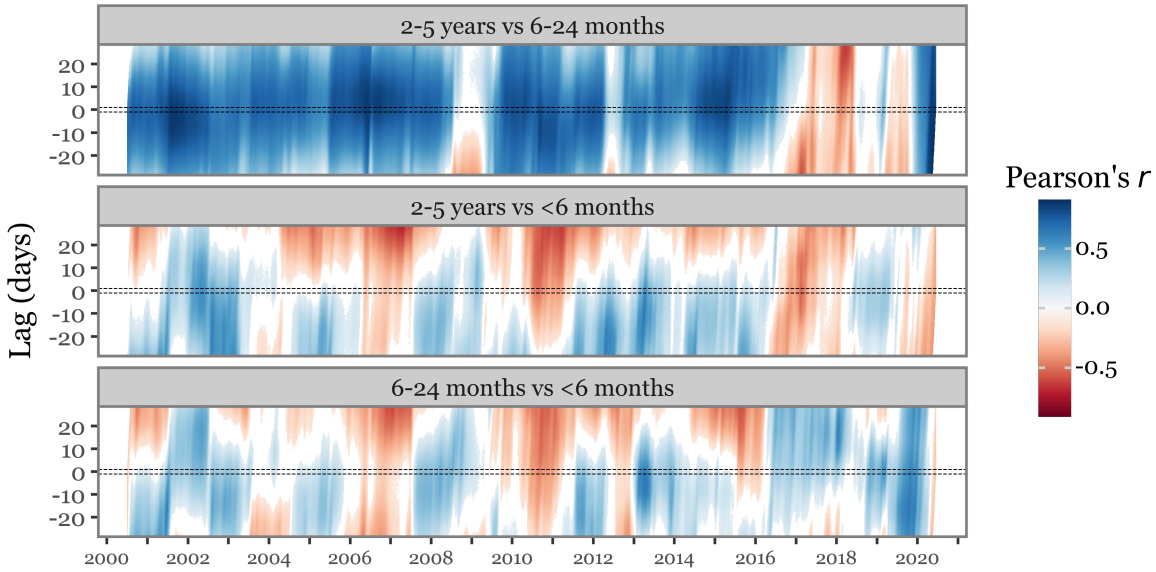


Suppl. Figure 3. Long-term trends in KD incidence since for single year age groups. Similar to Figure 4, but with more granular age groups (6-month intervals for children under 2 years old and 1-year intervals for children aged 2 to 9 years). The top panel shows a line plot depicting the relative increase in yearly incidence since 1996, while the bottom panel presents the same data in a heatmap.



Suppl. Figure 4. Granular age-stratified seasonal patterns of KD incidence (2000-2019). Similar to Figure 5, but with more granular age groups. The panels show the yearly-standardized KD incidence rates as a function of the day of the year for each specific age range. The solid black line represents the mean incidence rate, while the shaded area corresponds to the 10th to 90th percentiles. The percentages in parentheses indicate the proportion of total KD patients within each age range.

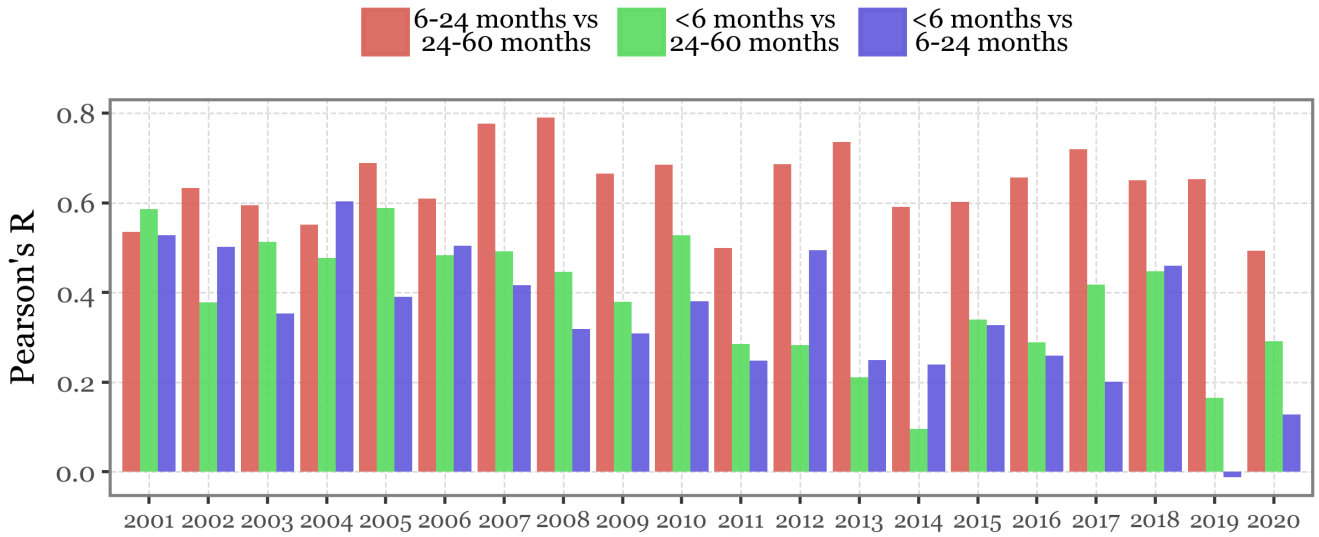
Suppl. Figure 5. Age-stratified comparison of year-by-year KD cases. Each panel displays the daily number of Kawasaki Disease (KD) cases for the three age groups (<6 months, 6-24 months, and 24-60 months). The semi-transparent lines represent the raw daily case data, while the solid lines show the 28-day centered moving average. The Pearson's correlation coefficient (r) for the seasonal signal of the toddler group (6-24 months) relative to the young children group (24-60 months) is indicated for each year.



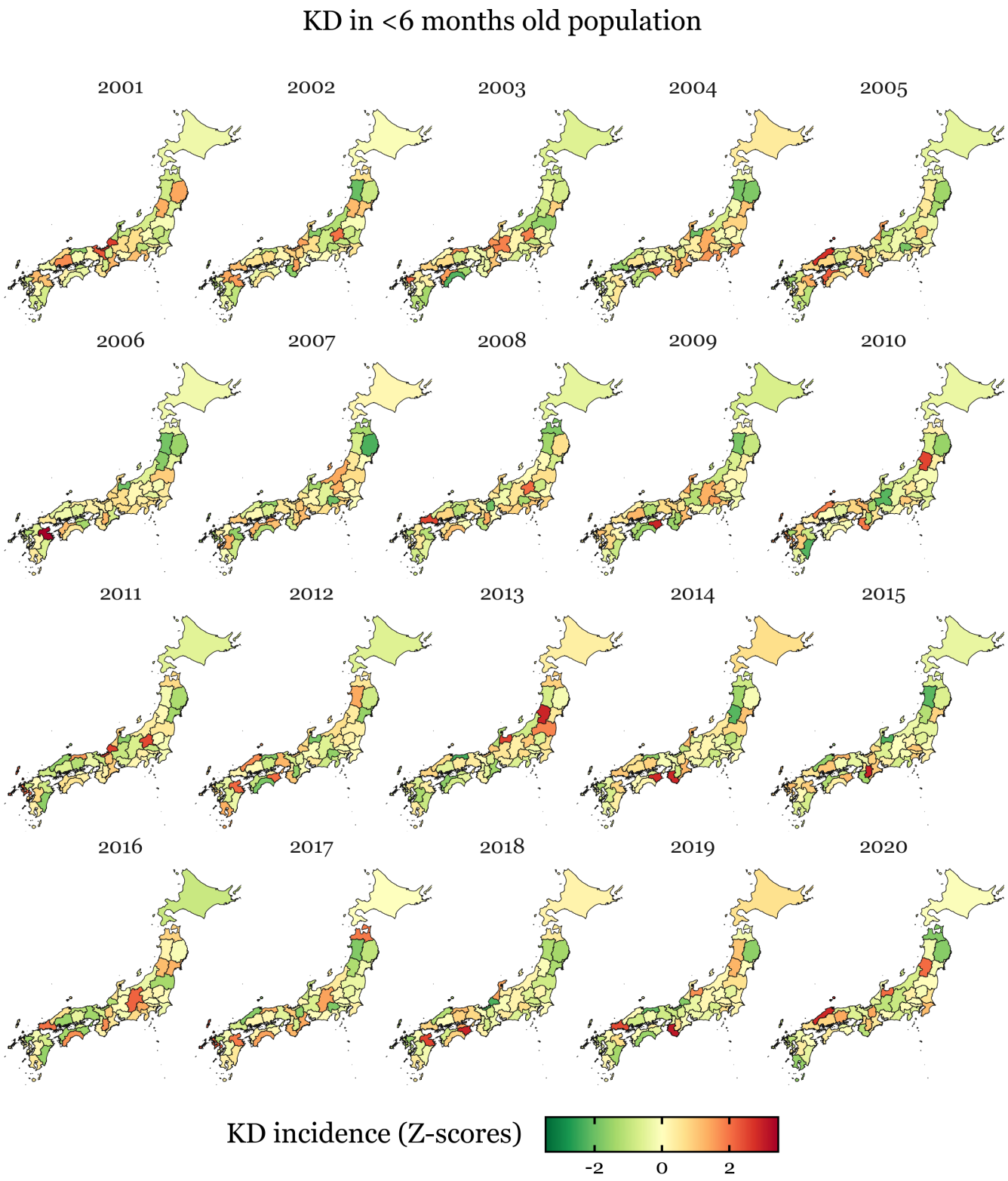
Suppl. Figure 6. Pairwise SDC analysis between age-groups. Same as in Figure 6B, but now showing the three pairwise combinations of ages including the newborns (<6 months old).

Suppl. Table 1. Spatial autocorrelation results for yearly standardized KD incidences across prefectures.

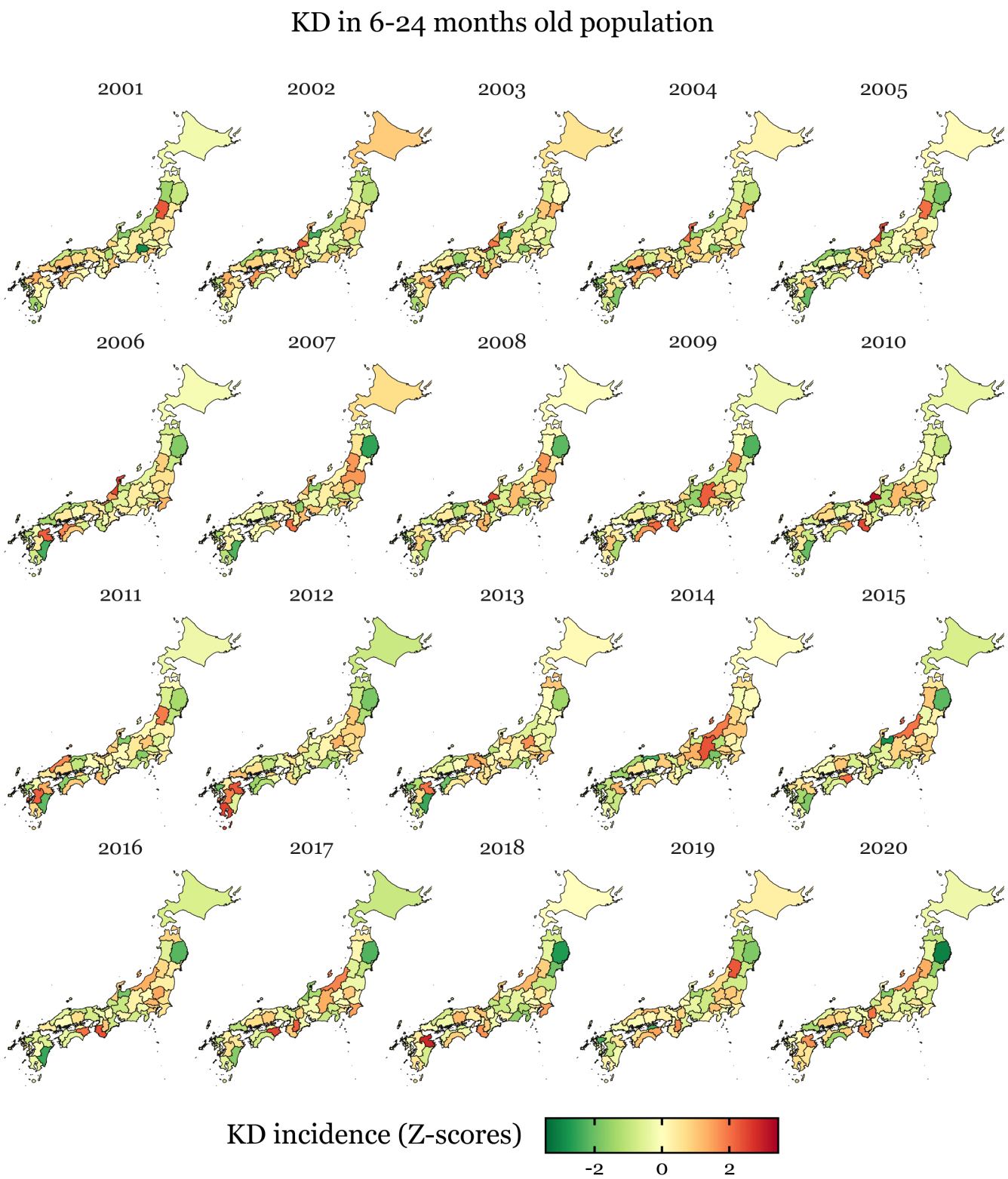
	24-60 months		6-24 months		<6 months	
	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value
2001	0.022	0.340	-0.062	0.377	-0.070	0.316
2002	-0.076	0.336	0.115	0.100	0.089	0.165
2003	-0.020	0.499	-0.204	0.038	-0.035	0.490
2004	-0.078	0.318	0.049	0.264	0.038	0.306
2005	0.061	0.219	-0.005	0.402	0.091	0.159
2006	0.200	0.034	0.056	0.233	0.149	0.043
2007	0.114	0.109	0.167	0.050	0.103	0.147
2008	0.126	0.086	0.030	0.300	0.098	0.149
2009	-0.083	0.291	-0.039	0.444	-0.188	0.059
2010	0.085	0.169	-0.134	0.140	-0.129	0.171
2011	0.159	0.056	0.058	0.218	-0.002	0.405
2012	0.095	0.153	0.095	0.144	0.049	0.254
2013	0.022	0.321	-0.009	0.440	-0.055	0.412
2014	0.201	0.027	0.069	0.209	-0.068	0.359
2015	0.058	0.215	0.061	0.237	0.076	0.192
2016	0.015	0.366	0.017	0.351	-0.040	0.447
2017	0.096	0.160	0.072	0.190	-0.005	0.422
2018	0.093	0.138	0.134	0.070	-0.030	0.497
2019	0.357	0.002	0.264	0.009	-0.135	0.146



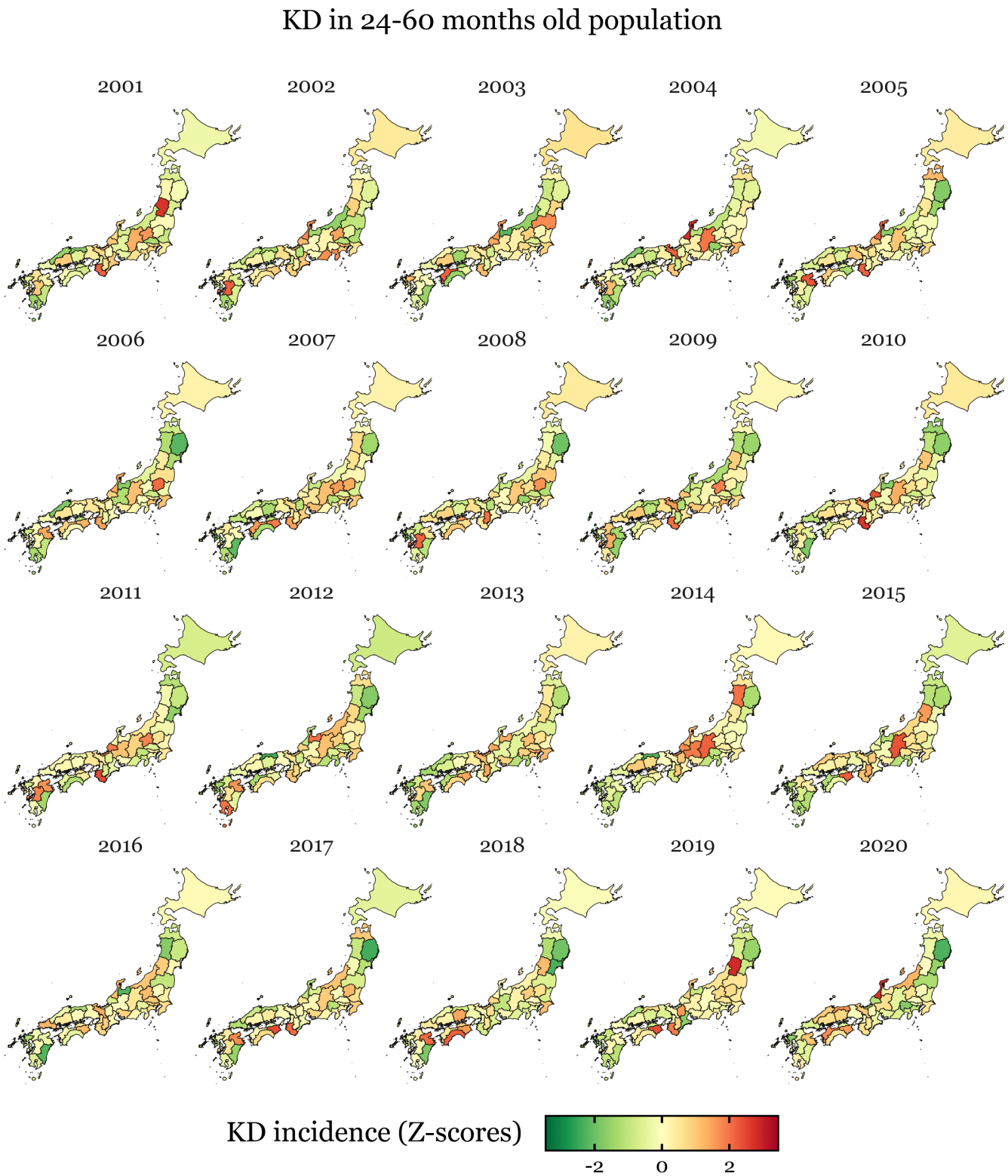
Suppl. Figure 7. Yearly Pearson correlation between yearly standardized KD incidences across prefectures and age groups.



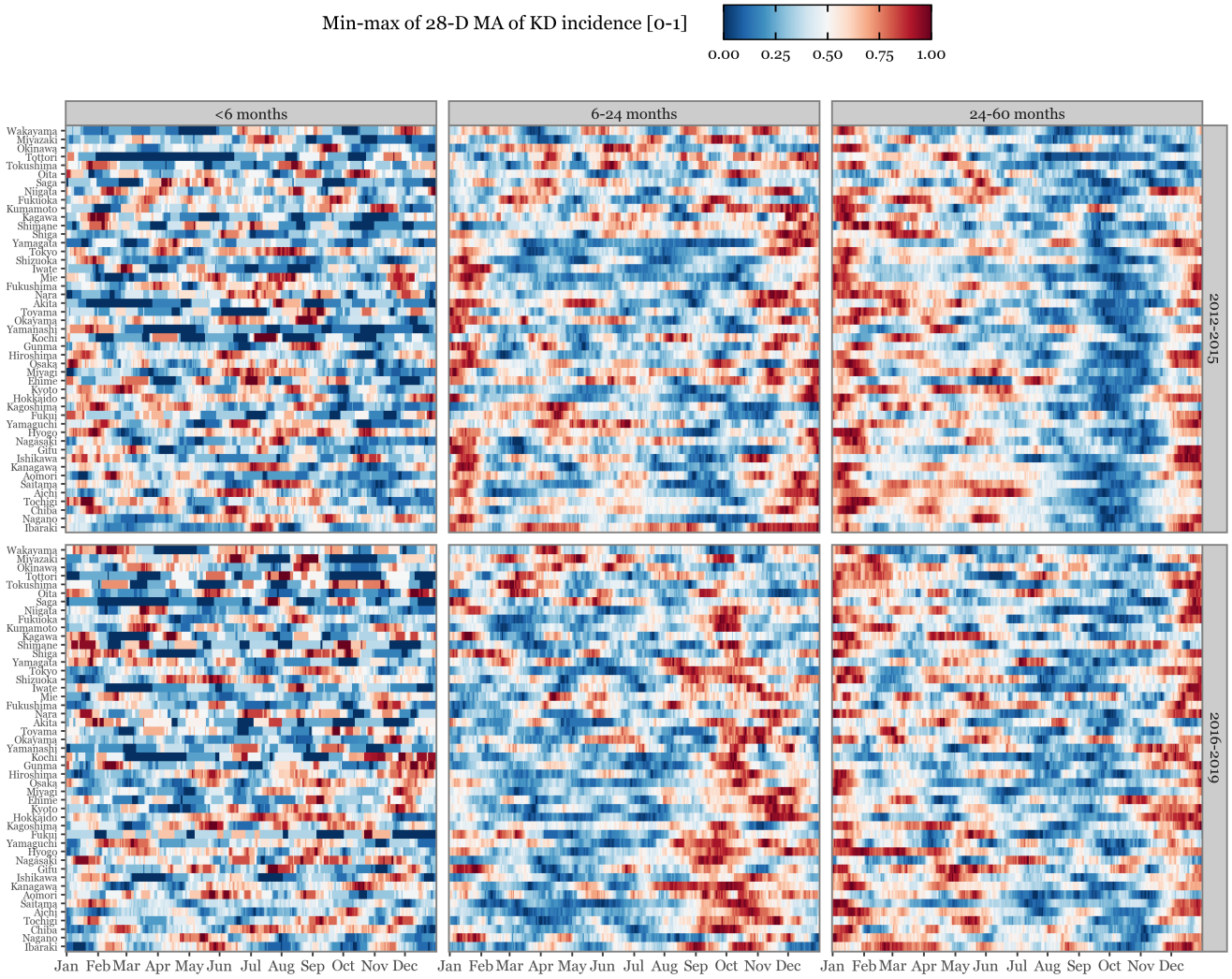
Suppl. Figure 8. Choropleth maps of the yearly-standardized KD incidence in the newborn age group (<6 months old) for the Japanese prefectures.



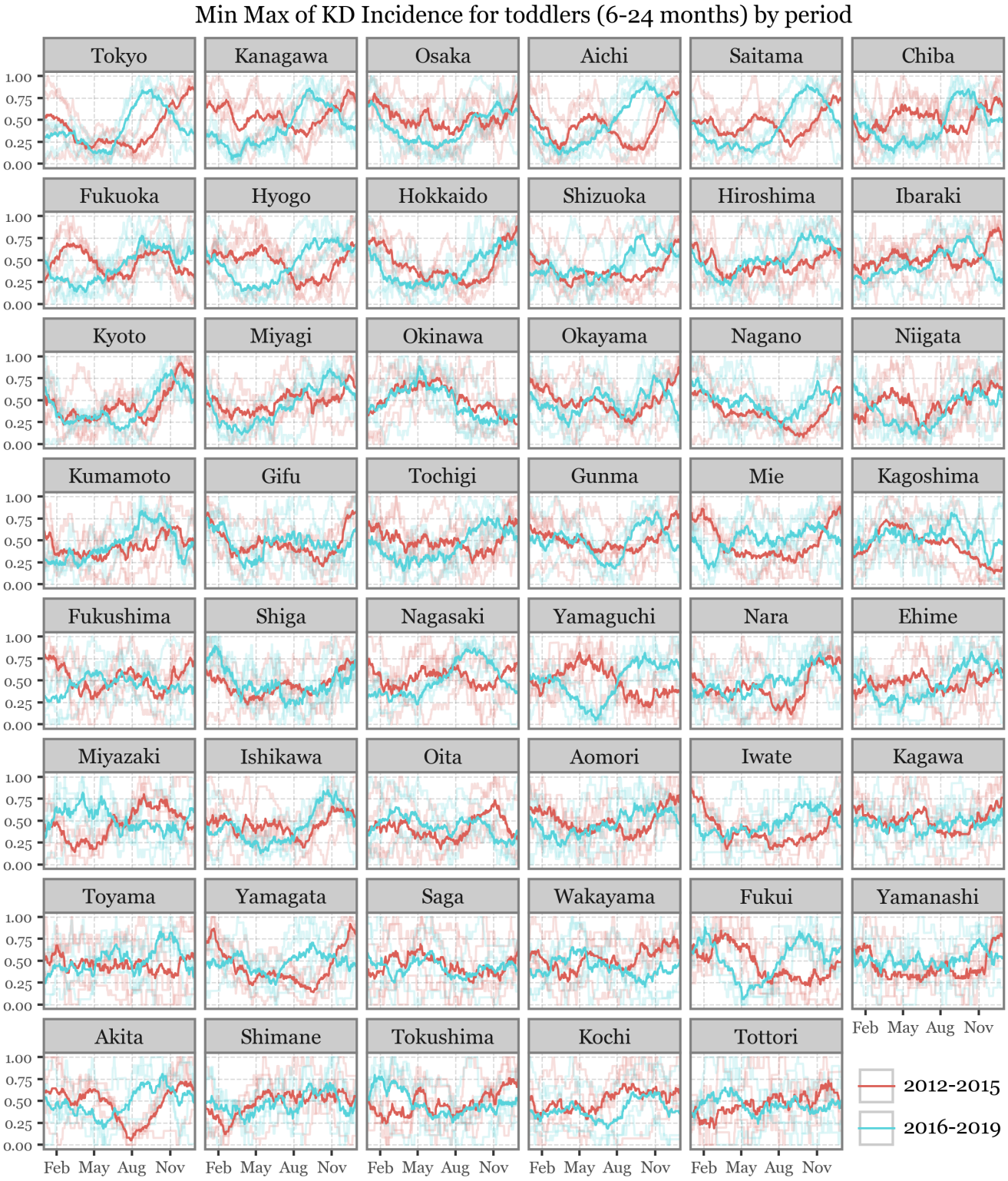
Suppl. Figure 9. Choropleth maps of the yearly-standardized KD incidence in the toddler age group (6-24 months old) for the Japanese prefectures.



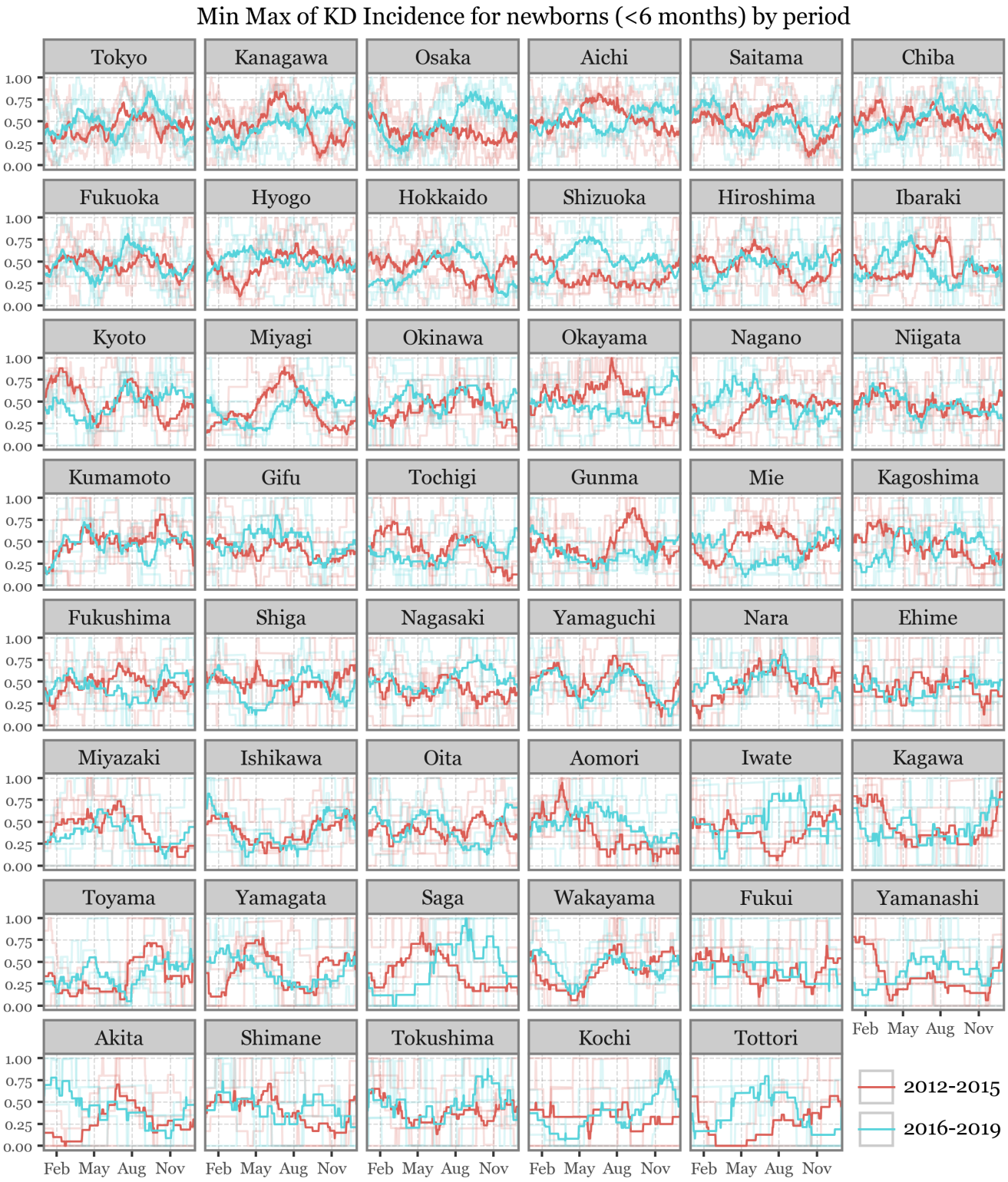
Suppl. Figure 10. Choropleth maps of the yearly-standardized KD incidence in the young children age group (24-60 months old) for the Japanese prefectures.



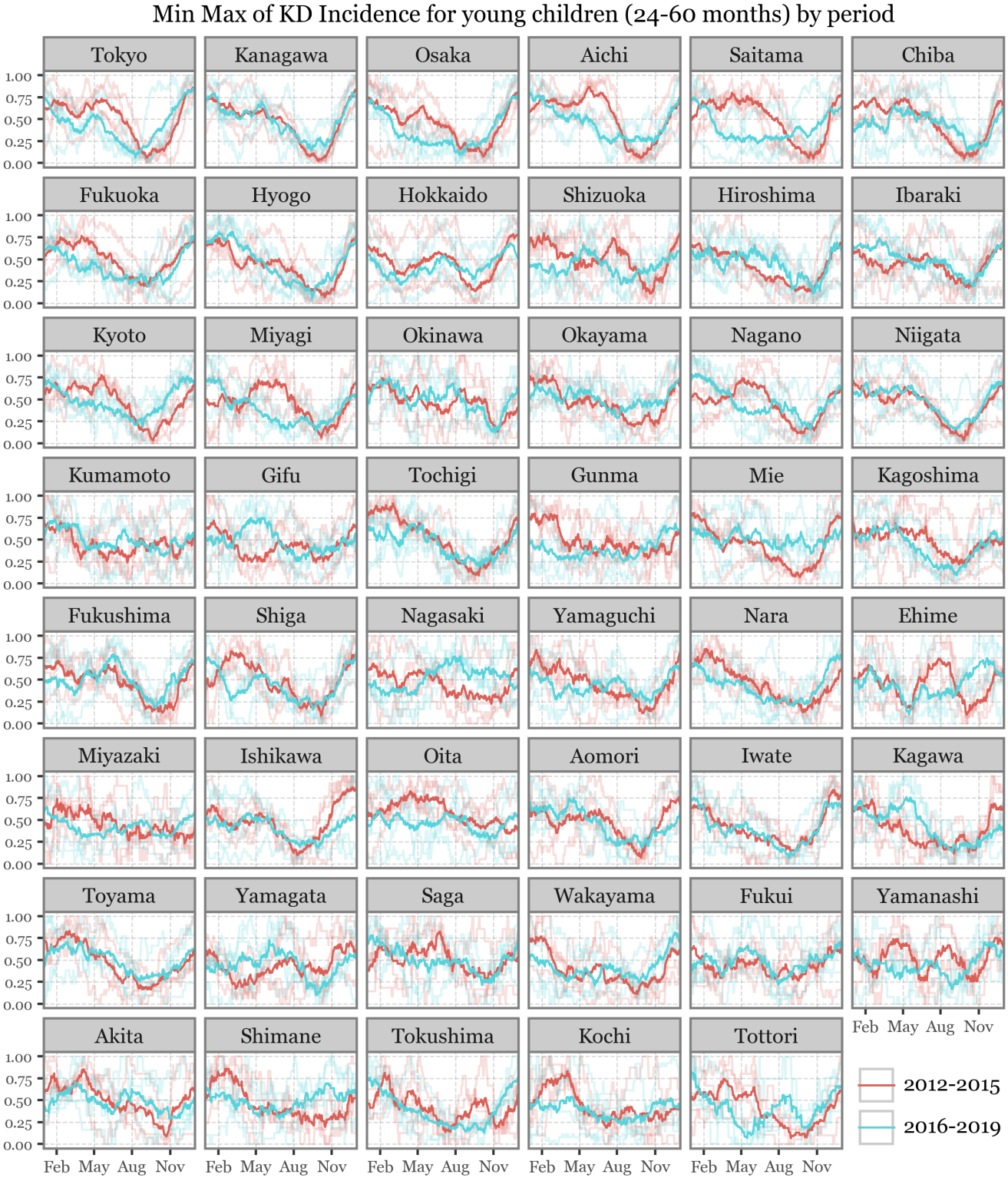
Suppl. Figure 11. Heatmap representing the average yearly seasonal signal (28-days moving average) for the 2012-2015 period (top) and the 2016-2019 period (bottom) for each of the age groups studied (min-max transformed)



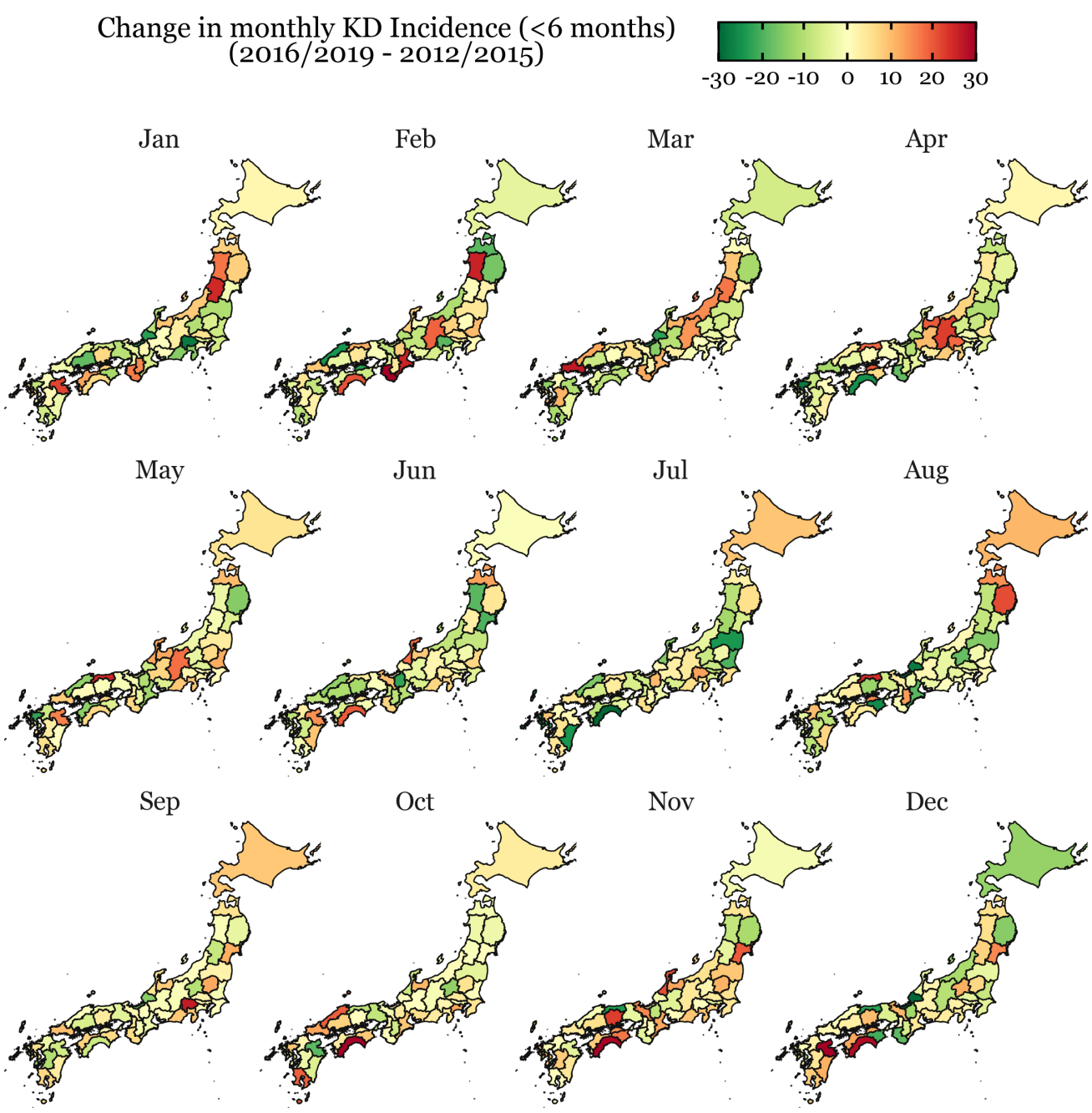
Suppl. Figure 12. Line-plot comparing the min-max transformed seasonal signal (28-day moving average) by period (2012-2015 in red, 2016-2019 in blue) of KD incidence by prefecture for the toddler age group. The bold lines represent the period averages while lighter lines represent the actual values from individual years.



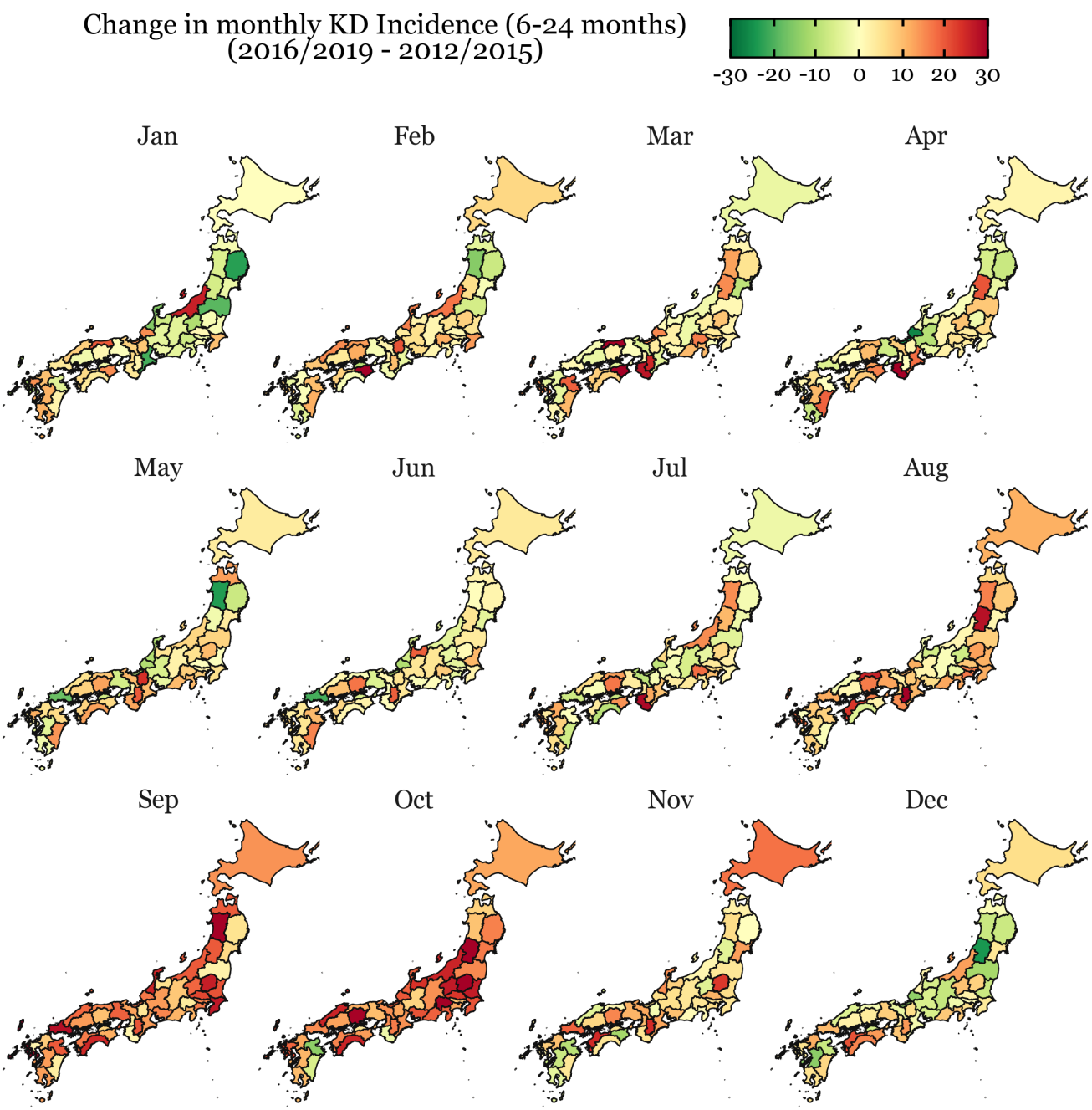
Suppl. Figure 13. Idem as Suppl. Figure 12, but for the newborns age group.



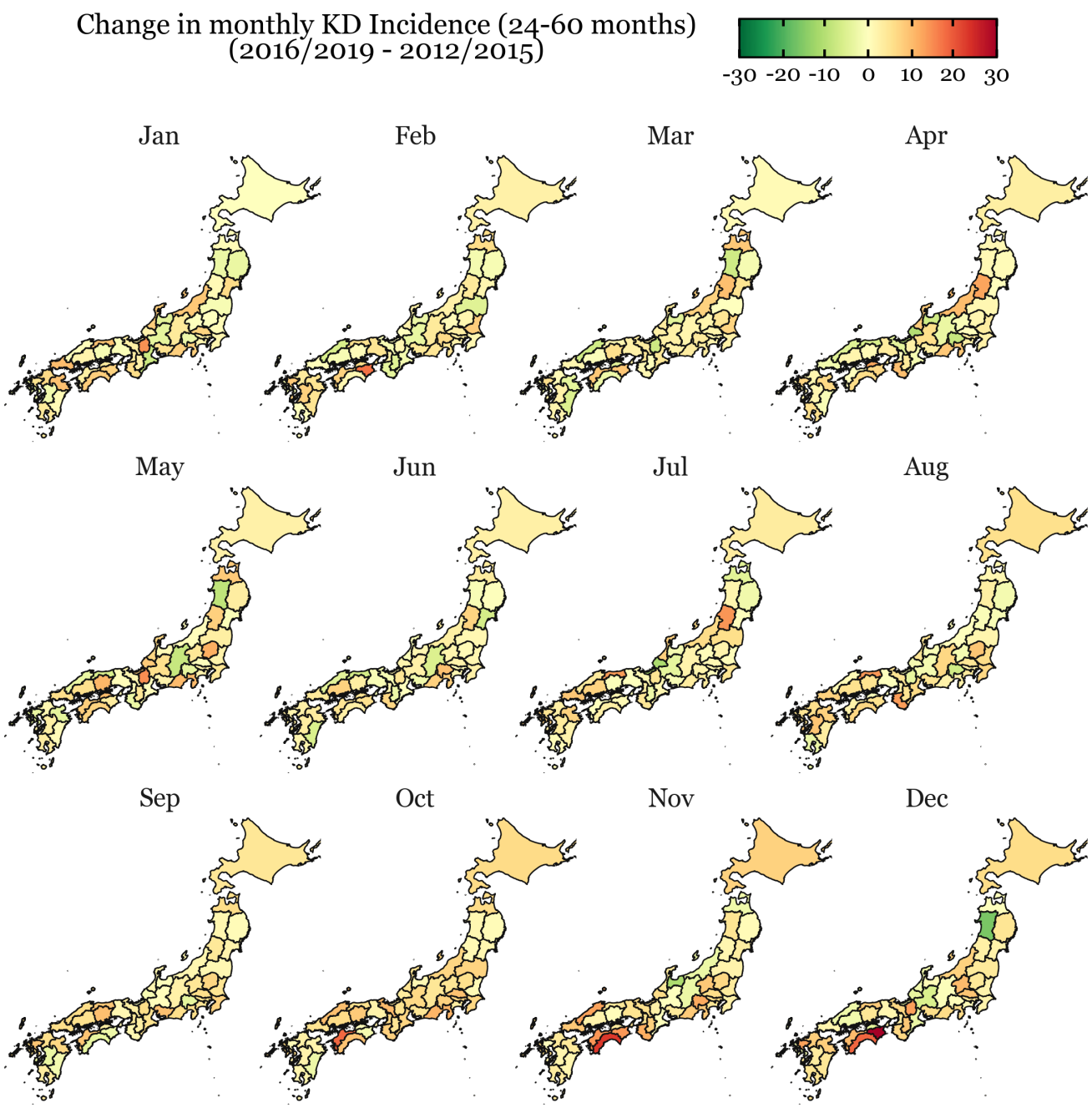
Suppl. Figure 14. Idem as Suppl. Figure 12, but for the young children age group.



Suppl. Figure 15. Choropleth maps showing the difference in KD incidence across prefectures between the 2012-2015 and the 2016-2019 periods for the newborn age groups.



Suppl. Figure 16. Idem as Suppl. Figure 15 but for the toddler age group.



Suppl. Figure 17. Idem as Suppl. Figure 15 but for the young children age group.

Suppl. Table 2. Same data as shown in Suppl. Fig. 15, but in tabular format.

Newborns (<6 months)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Aichi	-0.25	-3.08	3.07	3.91	-5.22	-5.63	-5.79	-2.87	0.5	5.8	0.54	4.11
Akita	17.31	25.73	9.1	4.55	-2.91	-18.87	-9.65	-8.89	-0.77	-2.76	-8.7	5.68
Aomori	7.92	-18.9	-0.69	-8.32	10.34	13.02	2.75	14.66	5.27	-1.65	4.88	5.76
Chiba	-6.68	1.12	-0.45	-5.26	0.52	-4.59	5.79	1.36	-0.03	4.97	4.87	-4.26
Ehime	12.72	-4.12	-4.08	3.19	-14.37	-5.46	-9.44	4.41	7.61	3.48	8.47	12.57
Fukui	-22.87	-10.1	-22.47	-11.38	-7.18	1.46	1.27	-27.54	-14.79	-0.46	-1.77	-31.23
Fukuoka	0.42	4.76	-0.29	-7.89	-7.31	3.21	0.31	10.07	-2.82	1.78	-0.48	2.41
Fukushima	-10.63	4.57	-7.38	-10.56	3.74	-8.31	-24.02	-15.24	2.5	-4.07	9.06	-3.47
Gifu	-0.14	-8.51	-9.97	10.5	7.19	-3.55	2.87	1.53	-0.96	-0.12	-4.89	2.27
Gunma	-10.25	7.54	-5.12	-6.51	-3.37	-3.22	-7.63	-18.03	-9.91	-13.77	1.89	11.74
Hiroshima	-18.26	-3.39	-7.65	-3.37	0.57	-11.25	-6.77	3.73	-5.79	7.58	5.71	3.15
Hokkaido	1.68	-3.81	-6.73	1.81	5.05	-0.57	9.24	10.95	8.87	3.61	-1.54	-13.62
Hyogo	-9.71	-1.03	5.93	-0.52	-1.01	-0.72	-5.98	-6.57	-4.56	-8.73	-3.48	-5.54
Ibaraki	-7.52	10.4	-1.83	4.77	12.19	6.87	-19.86	0.42	-1.63	-0.27	5.19	-8.1
Ishikawa	-8.96	7.94	-8.3	-3.4	12.43	17.58	-15.34	-4.29	4.92	-7.31	20.97	-3.11
Iwate	7.82	-16.29	-11.8	-4.06	-14.98	4.58	5.71	21.29	-3.47	-2.41	-11.4	-14.6
Kagawa	7.75	-22.95	-11.61	20.42	6.94	5.93	-7.22	15.43	1.83	3.16	14.75	12.4
Kagoshima	-6.92	-9.05	-11.45	-1.98	5.8	-1.57	4.49	-9.36	7.06	19.06	-0.36	8.09
Kanagawa	0.58	-3.51	-0.19	-2.81	2.01	2.52	-5.77	-0.1	0.72	12.54	7.89	1.54
Kochi	6.75	20.64	-9.8	-24.77	6.79	19.97	-38.16	3.9	-7.36	34.59	29.63	28.63
Kumamoto	1.5	2.34	11.19	-4.87	3.47	-6.91	5.25	-7.45	-10.5	-2.78	8.42	4.64
Kyoto	-3.45	-8.63	1.42	-2.63	-2.1	10.07	-7.04	3.51	1.78	-1.72	15.3	10.85
Mie	-6.28	25.63	11.07	-3.7	-12.32	-10.23	-2.79	-18.99	-2.16	8.95	12.52	4.26
Miyagi	-1.83	2.43	2.12	-7.35	-5.05	-19.79	-6.22	-5.18	11.79	0.26	19.82	15.38
Miyazaki	-2.4	2.82	-3.61	3.19	0.14	9.19	-24.46	7.17	3.36	-5.35	4.48	11.63
Nagano	3.5	18.94	14.13	22.78	17.76	1.95	4.05	-3.43	-1.04	-1.23	4.32	-10.1
Nagasaki	6.46	-4.12	3.07	-5.56	-4.0	-2.61	-22.74	-0.51	6.08	-4.13	9.62	-7.59
Nara	15.39	4.33	1.36	-5.71	-3.54	-9.63	1.4	13.43	-1.41	-6.89	0.43	-6.05
Niigata	8.58	-10.86	15.25	7.39	-2.8	-8.88	-3.75	1.28	0.13	0.81	5.36	-11.3
Oita	21.94	-8.69	-10.79	0.5	16.11	14.6	1.22	-9.32	-6.7	-18.73	-4.06	34.55
Okayama	6.78	4.66	-0.38	0.91	-0.56	-11.8	3.53	-13.04	3.98	0.13	22.79	9.93
Okinawa	-2.81	-5.52	16.86	5.67	-1.52	-0.89	-3.97	10.05	-7.92	-2.22	2.23	4.6
Osaka	-2.38	-3.92	-4.11	-4.37	2.26	-8.0	-5.51	1.78	4.79	1.46	-0.16	-0.01
Saga	-8.79	-4.12	-13.7	-27.63	-21.9	-9.13	8.71	-3.45	5.5	8.75	-0.87	-0.93
Saitama	5.22	6.07	-2.95	-2.16	-5.13	1.96	-7.74	-5.36	2.39	4.03	4.03	-4.13
Shiga	1.84	8.4	-14.69	-10.24	-10.7	-22.24	10.4	-2.52	-0.28	3.79	-9.14	-12.1
Shimane	-3.43	-24.38	12.29	2.92	-13.77	-13.93	-15.68	-1.13	-0.21	19.49	-1.57	-0.61
Shizuoka	-15.04	-6.25	-1.33	8.39	8.75	7.11	5.79	-0.43	8.79	3.76	9.2	3.3
Tochigi	-5.52	0.93	-6.39	-7.83	6.56	-2.02	-2.76	-1.97	12.24	5.19	11.49	7.05
Tokushima	-11.99	-7.16	-9.63	6.5	-5.6	-1.55	-3.83	-25.87	6.37	8.61	17.0	-20.81
Tokyo	-5.09	3.66	0.97	-1.36	0.11	-3.64	-9.32	0.05	2.74	-1.63	-4.05	-0.6
Tottori	-6.88	11.88	4.54	18.56	25.22	3.75	4.34	24.66	-2.74	-9.52	-23.61	-22.67
Toyama	9.18	-0.87	10.89	19.67	12.45	8.64	2.95	-8.15	8.39	10.88	12.91	2.0
Wakayama	16.54	34.79	12.19	-18.32	6.97	7.21	6.0	-1.27	6.13	-0.82	-4.2	-18.48
Yamagata	25.09	0.52	17.09	-5.6	-1.33	2.98	-12.09	-8.34	-8.13	-0.63	-3.54	7.63
Yamaguchi	-1.97	8.74	26.72	-11.85	8.32	-9.53	5.29	0.08	9.85	12.93	4.2	1.55
Yamanashi	-27.69	-18.86	-0.21	18.44	2.1	4.15	13.1	-3.74	26.58	1.93	-0.06	2.89

Suppl. Table 3. Same data as shown in Suppl. Fig. 16, but in tabular format.

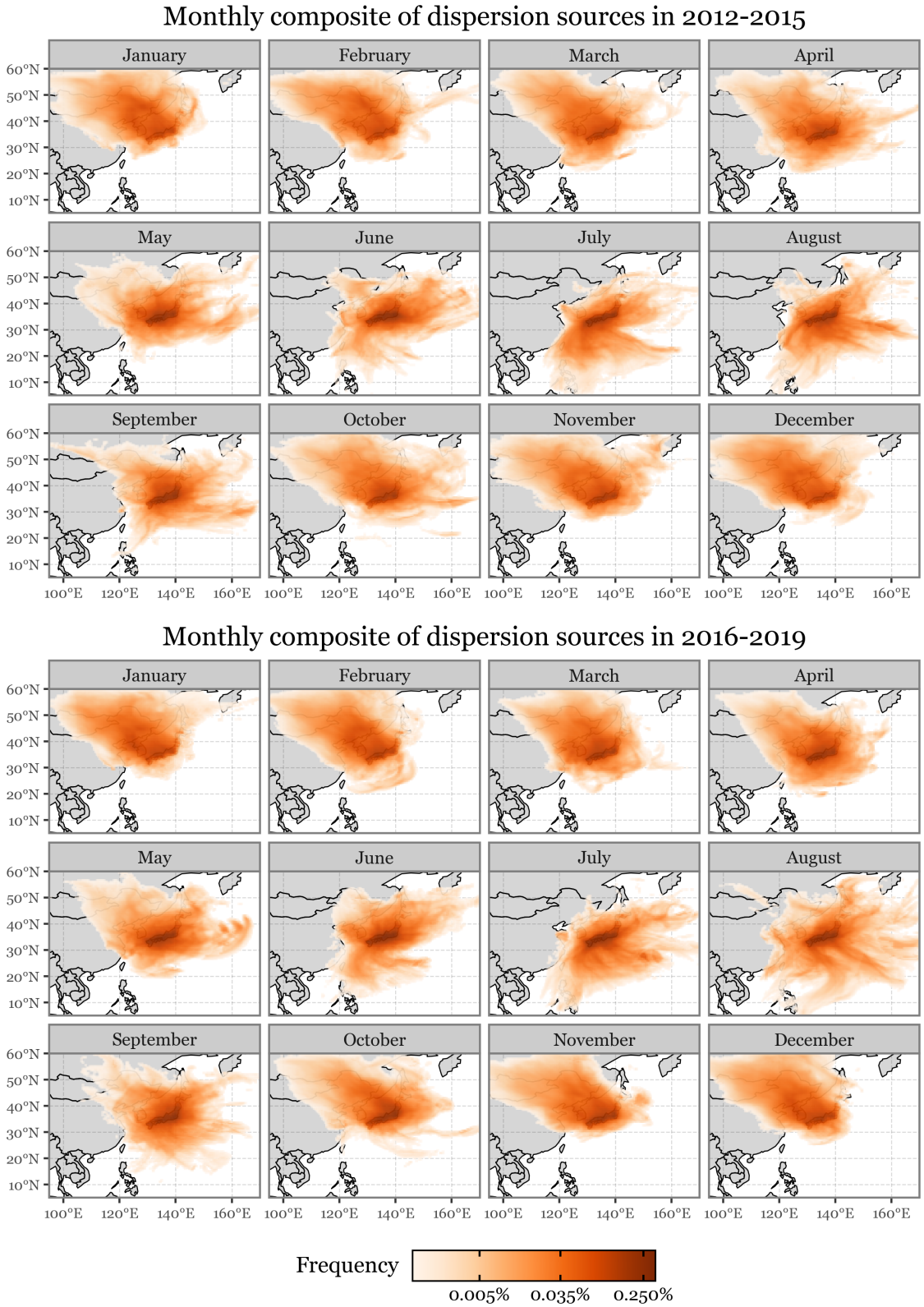
Toddlers (6-24 months)

		Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Aichi	-2.73	6.28	1.31	-3.23	-3.36	-0.65	3.31	6.3	15.72	20.53	9.66	-5.98
Akita	-5.55	-14.86	13.06	-6.35	-23.48	0.33	14.92	16.18	30.07	9.62	6.2	-7.78
Aomori	-1.57	-4.2	1.76	-2.51	13.24	2.37	-0.69	6.18	19.7	7.6	2.62	-0.79
Chiba	9.65	13.84	2.7	0.34	5.66	2.19	2.09	11.8	26.82	20.68	4.91	4.07
Ehime	6.58	7.69	8.13	6.07	7.52	7.18	3.79	23.48	16.63	15.97	24.8	20.63
Fukui	14.99	3.82	15.46	-26.26	-10.52	-10.81	-8.88	11.54	17.05	20.12	5.94	-14.0
Fukuoka	9.83	-0.39	-3.41	5.48	5.76	4.67	12.39	8.58	8.52	12.45	9.76	11.86
Fukushima	-18.96	-6.53	1.38	8.89	7.44	3.81	-4.33	7.99	2.48	15.43	4.48	-11.82
Gifu	-3.83	5.85	2.92	-9.96	-5.8	4.54	-1.53	2.08	13.72	7.51	5.82	-5.74
Gunma	-10.62	5.64	3.58	9.53	5.42	-3.49	-3.49	-7.08	18.6	26.73	-4.81	9.63
Hiroshima	-2.26	4.28	-3.1	-0.31	7.7	9.78	-1.07	1.64	13.79	13.14	6.87	2.75
Hokkaido	0.15	6.87	-3.24	1.9	3.58	4.11	-2.78	11.55	14.52	12.55	17.47	6.07
Hyogo	-0.43	-1.25	-0.38	-7.29	-6.45	-5.28	3.58	12.87	19.17	10.17	11.14	8.18
Ibaraki	-1.09	5.05	2.22	6.62	-0.27	10.67	2.89	11.43	20.68	26.46	5.89	-7.62
Ishikawa	-15.18	16.3	0.91	-1.17	-10.58	-6.68	6.72	-1.22	25.53	19.22	14.97	1.95
Iwate	-22.83	-7.59	5.3	-7.97	-7.71	2.08	-1.5	8.33	5.31	16.15	0.2	-7.61
Kagawa	-2.25	-9.93	-4.25	12.94	-6.6	-5.71	-7.85	16.06	7.49	4.5	5.31	-6.36
Kagoshima	11.57	-4.75	1.22	-7.88	-2.15	4.55	6.04	7.56	13.39	10.23	-3.99	5.33
Kanagawa	-0.74	1.84	-0.64	3.62	7.86	-3.77	9.99	16.77	16.52	12.51	5.98	5.42
Kochi	4.02	-0.97	7.58	8.49	13.06	1.72	-9.61	-4.8	25.07	25.55	4.78	15.81
Kumamoto	11.34	-0.78	-3.9	10.64	-4.54	5.09	7.99	5.66	13.47	10.89	-7.61	-14.4
Kyoto	7.58	2.17	-5.14	8.03	10.6	1.47	-9.0	2.98	13.07	14.6	12.41	-3.09
Mie	-20.71	-3.46	-4.24	17.77	10.52	4.65	9.23	10.85	13.5	14.62	12.57	-1.88
Miyagi	-2.87	-0.98	-9.17	-3.31	3.63	-4.04	-2.79	1.74	6.76	16.77	13.31	4.65
Miyazaki	2.18	11.67	10.7	19.0	14.23	15.83	-6.98	-0.86	3.49	-5.9	4.6	6.86
Nagano	-4.38	1.71	-2.69	2.24	2.49	4.88	-6.05	-0.76	8.9	14.75	4.62	-8.29
Nagasaki	-11.04	7.8	14.81	-1.4	2.75	7.07	13.79	17.78	39.6	20.74	2.18	3.35
Nara	2.05	10.0	25.29	-0.21	19.97	19.41	12.4	30.06	20.39	16.59	24.64	14.03
Niigata	26.05	17.37	-2.32	-0.77	7.27	-4.11	15.08	1.64	19.95	25.62	1.83	13.33
Oita	-4.63	9.95	19.13	-0.08	6.56	10.43	-0.08	7.53	19.88	-14.33	-9.75	-7.96
Okayama	1.48	12.08	2.42	10.24	13.1	17.27	17.08	16.53	9.69	29.67	15.69	3.91
Okinawa	13.21	1.27	-2.08	6.84	3.5	5.64	0.18	6.01	-1.51	5.78	4.57	5.76
Osaka	6.53	2.37	1.82	0.94	1.89	-0.87	3.65	7.49	4.11	15.01	10.46	4.72
Saga	14.63	-3.35	-3.82	1.43	11.91	2.97	3.44	-0.07	2.09	18.68	8.0	5.51
Saitama	1.3	1.46	4.43	3.78	0.01	-3.92	4.02	8.88	21.65	23.5	11.26	2.42
Shiga	9.11	20.63	3.56	1.45	24.16	5.36	-1.61	8.28	4.57	12.6	-1.55	5.89
Shimane	7.78	15.13	-1.34	1.67	5.39	1.65	6.32	-1.48	18.47	25.51	-4.69	-5.01
Shizuoka	-3.86	1.68	10.54	2.83	10.56	0.47	4.16	12.67	11.19	19.18	0.87	8.58
Tochigi	-1.4	7.82	9.39	-2.12	10.4	1.12	6.75	13.32	26.15	33.4	23.23	7.7
Tokushima	15.06	33.07	37.45	16.11	6.46	2.44	14.57	2.49	11.93	13.62	-12.07	13.82
Tokyo	-2.35	2.95	7.0	4.53	2.32	-0.76	6.72	16.31	15.53	9.13	4.73	0.26
Tottori	19.93	13.42	35.39	10.87	8.78	-0.68	13.34	25.86	15.8	11.67	1.39	-1.53
Toyama	-7.13	2.22	-0.93	-3.89	5.07	19.45	0.88	-6.65	5.46	23.9	7.84	-6.24
Wakayama	4.67	-1.65	27.37	39.29	12.58	0.29	29.13	14.83	1.82	-2.55	4.97	9.17
Yamagata	-6.42	6.03	15.09	20.39	-1.97	4.95	10.64	27.92	19.96	31.88	-4.57	-24.15
Yamaguchi	4.1	9.03	1.34	-5.29	-17.49	-20.49	-7.22	12.16	28.2	15.27	18.92	10.56
Yamanashi	-7.12	11.93	16.51	5.36	8.2	8.43	17.96	-3.18	3.93	30.32	4.79	-2.17

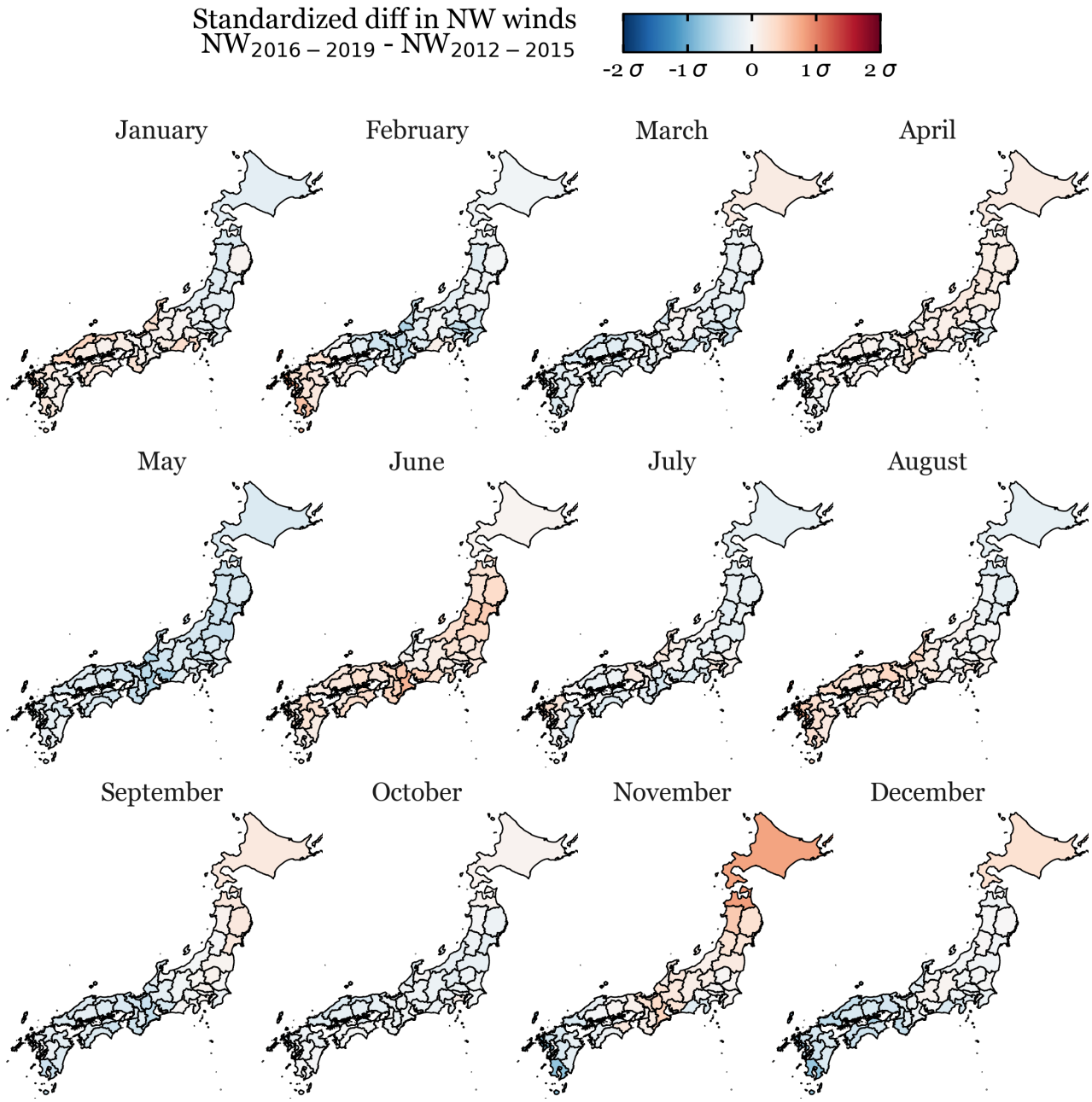
Suppl. Table 4. Same data as shown in Suppl. Fig. 17, but in tabular format.

Young Children (24-60 months)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Aichi	5.44	4.24	2.74	3.13	2.21	-2.26	-1.02	4.95	9.81	4.77	4.09	6.32
Akita	-2.85	-1.85	-7.12	1.36	-9.17	-1.0	1.92	3.22	1.74	1.63	5.11	-16.13
Aomori	0.8	6.15	8.92	1.34	8.47	3.55	-4.64	1.48	6.3	6.38	-2.4	-0.4
Chiba	1.67	1.57	2.6	2.61	4.8	2.65	5.03	6.93	7.53	6.98	2.76	2.34
Ehime	3.11	7.62	8.91	0.42	8.17	4.75	-0.2	2.52	9.29	18.47	14.31	10.43
Fukui	2.54	0.11	2.94	-10.25	8.66	2.07	-11.06	-1.85	-1.23	3.62	6.12	-3.73
Fukuoka	6.73	3.84	0.82	1.88	1.55	2.79	2.89	5.86	5.7	2.88	4.05	8.03
Fukushima	-1.31	-4.9	-0.04	1.63	3.19	1.58	5.96	1.92	1.62	7.33	2.2	4.01
Gifu	-3.77	-3.45	2.9	5.97	5.82	2.76	-3.82	-1.68	0.37	6.73	-0.01	-5.72
Gunma	-0.12	0.93	1.91	-0.97	4.61	2.09	0.68	-1.54	4.49	4.94	8.73	10.47
Hiroshima	2.27	-1.39	-0.48	2.0	5.82	7.73	7.42	5.61	5.57	6.48	2.14	-2.31
Hokkaido	-0.2	2.72	1.1	3.15	2.61	2.96	3.72	5.66	4.48	6.53	7.58	6.17
Hyogo	-0.83	5.0	2.68	3.96	0.48	-0.97	-0.1	-0.35	1.47	2.4	5.41	1.74
Ibaraki	0.73	8.18	7.33	3.48	2.55	1.82	0.55	6.12	5.13	1.16	3.81	2.3
Ishikawa	8.0	2.37	3.64	8.24	6.86	-0.74	7.94	4.83	1.77	4.96	3.96	-2.04
Iwate	-3.24	-1.39	-1.49	0.82	3.39	-0.39	-2.17	-1.73	0.72	0.74	0.79	4.86
Kagawa	3.61	5.99	-4.4	12.59	10.21	5.04	1.44	-0.6	3.25	0.08	3.76	-3.84
Kagoshima	4.06	0.06	2.1	5.84	3.47	2.02	0.89	-2.96	3.59	1.89	3.69	5.11
Kanagawa	1.79	2.32	1.7	2.39	2.31	1.54	0.32	3.96	4.73	3.51	4.72	0.92
Kochi	8.18	1.56	-2.59	6.25	7.42	2.83	2.99	7.04	-3.32	9.42	23.25	18.79
Kumamoto	-2.19	7.82	2.66	1.86	3.21	3.43	3.93	9.26	-3.06	2.22	1.98	3.67
Kyoto	2.99	-1.21	3.71	-4.15	1.84	-1.22	1.22	5.09	3.11	10.55	5.75	6.9
Mie	-7.46	-1.66	4.39	-1.48	0.44	-0.0	1.61	6.58	5.34	7.18	-0.34	-2.87
Miyagi	6.36	1.72	5.22	0.38	-0.76	-6.48	0.33	-1.07	2.28	1.78	3.43	2.37
Miyazaki	1.15	5.02	-5.53	-1.14	0.92	-5.69	2.14	3.4	-1.8	-2.61	1.86	4.66
Nagano	4.11	4.16	2.03	-3.14	-8.21	-5.6	2.14	6.84	2.59	4.53	-2.86	-0.5
Nagasaki	-3.42	-1.79	-1.99	1.88	-3.98	3.37	8.16	5.54	5.83	6.73	6.41	-1.92
Nara	2.59	-4.71	0.4	4.07	4.5	-2.76	-2.67	4.99	2.95	4.89	7.77	9.01
Niigata	8.9	2.75	7.76	10.25	0.27	3.06	6.45	0.19	1.77	7.12	-2.94	7.67
Oita	9.73	5.15	-3.9	6.66	-3.86	3.63	7.58	7.68	6.97	2.28	4.2	8.03
Okayama	-1.23	0.51	5.75	1.1	10.85	3.29	-0.5	5.22	10.07	8.69	0.55	4.51
Okinawa	-0.49	5.8	2.97	2.34	-1.94	4.3	2.42	1.19	0.52	1.4	-1.07	2.59
Osaka	3.5	2.75	4.55	0.21	0.53	1.45	2.9	2.73	4.01	6.83	5.41	3.33
Saga	5.25	8.52	5.96	7.15	-3.19	1.88	4.95	5.06	9.34	0.74	0.04	13.04
Saitama	7.45	3.01	2.01	0.11	-0.46	2.24	1.4	4.49	5.95	9.54	6.23	9.19
Shiga	14.65	1.53	-7.52	2.67	15.05	4.94	5.13	-3.67	6.28	5.27	2.61	13.4
Shimane	0.66	-6.08	-7.0	-4.71	-3.23	-4.4	3.91	1.16	6.21	9.91	13.17	-2.42
Shizuoka	7.77	4.83	3.36	8.86	10.36	7.94	3.54	2.46	6.06	11.07	8.7	6.42
Tochigi	-0.08	1.8	4.87	1.76	11.87	1.09	2.16	9.36	7.27	5.76	6.86	4.39
Tokushima	7.51	17.5	9.04	6.95	3.67	5.42	1.11	5.52	-2.46	-1.76	14.06	36.08
Tokyo	2.91	3.11	1.44	3.73	2.34	0.88	1.43	2.49	5.23	6.86	6.31	4.65
Tottori	9.61	5.43	1.6	-5.73	-5.76	-6.58	13.6	14.05	8.8	5.48	4.37	8.64
Toyama	-4.3	-2.94	-4.33	-6.86	0.97	1.63	-1.76	-0.95	-0.59	1.18	-11.28	0.88
Wakayama	6.27	-3.94	1.49	10.12	-2.98	6.84	4.57	13.54	1.87	7.8	11.84	6.9
Yamagata	1.14	4.16	10.11	12.81	7.78	7.45	14.18	-0.93	4.37	1.27	1.81	2.84
Yamaguchi	10.19	0.59	1.72	-1.96	6.08	4.05	9.09	7.57	5.96	5.26	8.92	1.89
Yamanashi	-2.42	6.5	3.84	-8.23	-3.75	9.85	2.05	-6.86	-3.86	7.63	12.27	0.49



Suppl. Figure 18. Monthly composites of backward dispersion sources by period. Same as Figure 11 but including the full year.



Suppl. Figure 19. Standardized differences in northwesterly winds between the 2012-2015 and the 2016-2019 periods

Suppl. Table 5. Statistical test results from the EVI comparison between the seasonal cycle features of the 2012-2015 period with the 2016-2019 period.

	Wilcoxon Test *p*-value			T-test *p*-value		
	Amplitude	Max value	Peak time	Amplitude	Max value	Peak time
Aichi	0.625	0.875	1.000	0.200	0.777	0.750
Akita	0.250	0.625	0.157	0.052	0.463	0.134
Aomori	0.250	0.625	1.000	**0.023**	0.339	1.000
Chiba	1.000	0.375	0.625	0.925	0.306	0.356
Ehime	0.875	0.625	0.375	0.509	0.512	0.207
Fukui	0.625	0.625	1.000	0.365	0.417	1.000
Fukuoka	1.000	1.000	0.875	0.881	0.854	0.705
Fukushima	0.125	1.000	0.564	0.113	0.927	0.537
Gifu	0.250	0.875	0.625	**0.025**	0.495	0.207
Gunma	0.250	0.625	0.375	0.287	0.506	0.114
Hiroshima	0.875	0.625	0.564	0.937	0.535	0.550
Hokkaido	0.125	0.125	1.000	**0.035**	0.147	1.000
Hyogo	0.250	0.875	0.875	0.250	0.800	0.537
Ibaraki	0.625	0.625	1.000	0.537	0.645	1.000
Ishikawa	0.875	0.875	0.317	0.550	0.970	0.356
Iwate	0.375	0.375	0.157	0.417	0.342	0.207
Kagawa	0.875	0.375	1.000	0.835	0.515	1.000
Kagoshima	0.625	0.625	1.000	0.851	0.345	1.000
Kanagawa	1.000	0.625	0.180	0.951	0.397	0.168
Kochi	0.625	0.875	1.000	0.576	0.555	0.873
Kumamoto	1.000	1.000	0.375	0.805	0.781	0.235
Kyoto	0.375	1.000	0.317	0.148	0.572	0.537
Mie	1.000	0.625	0.875	0.755	0.421	0.809
Miyagi	0.125	0.250	0.414	0.233	0.595	0.356
Miyazaki	0.250	0.125	1.000	0.263	0.152	0.766
Nagano	0.250	0.250	0.564	0.081	0.358	0.620
Nagasaki	0.625	1.000	0.375	0.494	0.687	0.149
Nara	0.375	0.875	1.000	0.251	0.906	0.628
Niigata	0.875	0.375	0.317	0.967	0.226	0.356
Oita	1.000	0.625	0.875	0.750	0.525	0.580
Okayama	0.875	0.875	0.157	0.569	0.566	0.134
Osaka	0.250	0.375	0.414	0.265	0.804	0.414
Saga	0.875	1.000	0.655	0.685	0.933	0.670
Saitama	0.250	0.375	0.785	0.303	0.264	0.705
Shiga	0.250	0.875	0.785	0.210	0.678	0.730
Shimane	0.875	0.625	0.625	0.624	0.301	0.207
Shizuoka	0.875	0.625	0.083	0.797	0.494	0.069
Tochigi	0.125	0.250	0.625	0.294	0.569	0.320
Tokushima	0.875	0.875	0.414	0.583	0.543	0.477
Tokyo	0.875	0.625	0.875	0.823	0.601	0.620
Tottori	0.625	0.875	1.000	0.316	0.824	1.000
Toyama	0.875	0.125	0.625	0.611	**0.006**	0.207
Wakayama	1.000	0.875	0.414	0.995	0.773	0.315
Yamagata	1.000	0.250	0.317	0.962	0.468	0.356
Yamaguchi	0.625	0.375	0.625	0.622	0.220	0.413
Yamanashi	0.375	0.875	0.083	0.366	0.862	0**.024**

PAPER III

Climatic signatures in the different COVID-19 pandemic waves across both hemispheres

Alejandro Fontal[1], Menno J. Bouma[1], Adrià San-José[1], Leonardo López[1], Mercedes Pascual[2] and Xavier Rodó[1,3]

[1]Climate and Health (CLIMA) Program, Barcelona Institute for Global Health (ISGlobal), Barcelona, Catalonia, Spain.

[2]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA.

[3]ICREA, Barcelona, Catalonia, Spain

Published peer-reviewed article:

Nature Computational Science 1, 655–665 (2021)

DOI: 10.1038/s43588-021-00136-6

IF: 12.0

nature computational science

ARTICLES

https://doi.org/10.1038/s43588-021-00136-6

Check for updates

Climatic signatures in the different COVID-19 pandemic waves across both hemispheres

Alejandro Fontal[1], Menno J. Bouma[1], Adrià San-José[1], Leonardo López[1], Mercedes Pascual[2] and Xavier Rodó[1,3] ✉

The roles of climate and true seasonal signatures in the epidemiology of emergent pathogens, and that of SARS-CoV-2 in particular, remain poorly understood. With a statistical method designed to detect transitory associations, we show, for COVID-19 cases, strong consistent negative effects of both temperature and absolute humidity at large spatial scales. At finer spatial resolutions, we substantiate these connections during the seasonal rise and fall of COVID-19. Strong disease responses are identified in the first two waves, suggesting clear ranges for temperature and absolute humidity that are similar to those formerly described for seasonal influenza. For COVID-19, in all studied regions and pandemic waves, a process-based model that incorporates a temperature-dependent transmission rate outperforms baseline formulations with no driver or a sinusoidal seasonality. Our results, so far, classify COVID-19 as a seasonal low-temperature infection and suggest an important contribution of the airborne pathway in the transmission of SARS-CoV-2, with implications for the control measures we discuss.

Given the periodic influenza pandemics in the twentieth century—the so-called Spanish flu in 1918, Asian flu in 1957 and Hong Kong flu in 1968—the medical community has been apprehensive about the emergence of novel pathogens with an extensive reach in a more connected world[1]. Unlike influenza, with its well appreciated ability to mutate[2], the few relatively benign coronaviruses did not originally cause alarm. This was the case until the 2003 outbreak of severe acute respiratory syndrome (SARS) in China[3] and Middle East respiratory syndrome (MERS) in Saudi Arabia in 2012[4], which featured severe symptoms and high mortality. Both outbreaks were short-lived, and isolation of cases proved sufficient to overcome the restricted ability of these pathogens to transmit between humans. The emergence of the novel SARS-CoV-2, with a similar geographic and zoonotic origin as SARS[5], was declared a pandemic by the World Health Organization (WHO) in March 2020[6] and has rapidly spread over extensive regions. So far, over 220 million cases have been recorded, with over 4.7 million deaths[7] and a high mortality, not unlike influenza, in older age groups[8].

Studies of the 2009 H1N1 influenza viruses suggested that low relative humidity (RH) and high population density shaped the early spread of the virus. Those conditions also favored persistence of the virus during the weeks at the beginning of the pandemic, underscoring an increasingly recognized role of humidity in influenza dynamics. Climatic factors were deemed relevant when designing control and prevention measures[9]. Determining the role of climate drivers in the transmission dynamics of emergent pathogens remains fraught with difficulties as a result of limited data, nonlinear responses and multiple potential confounding factors at the early stages of pandemics[10]. Examples of such factors include processes sharing trends, variable intervention measures such as lockdowns, and the rising use of non-pharmacological interventions. An ample availability of susceptible individuals fueling transmission is also thought to lessen the importance of climate conditions[11]. The results of a mathematical model addressing the potential for seasonality in the population dynamics of COVID-19 suggested an inability of climate forcing to establish seasonality given the large number of available susceptible, non-immune individuals. Nevertheless, initial observations suggested that SARS-CoV-2 could be a seasonal disease, a possibility that was not advanced for the short-lived outbreaks of SARS and MERS. Specifically, the initial propagation of COVID-19 emerged in a latitudinal band between 30° N and 50° N, with low humidity levels and temperatures between 5 °C and 11 °C (ref. [12]), and these weather sensitivities were reported from China, with absolute humidity (AH) negatively associated with daily death counts of COVID-19[13]. SARS-CoV-2 further resembled influenza with respect to its winter appearance. Recent perspectives have reviewed and evaluated the evidence for climate-driven seasonality, indicating contradictory results and confounding effects of other factors at these early stages of the pandemic[10,14].

At a more mechanistic level, some observations on the long-range transmission of SARS in 2003[15] raised the scope for aerial transmission and a role of aerosols[3], which would bypass short-range control measures. The intangible connection between airborne and seasonal transmission has been implicit in the scientific interest that followed the SARS 2003 outbreak, particularly for influenza. Both the nature of seasonality and the plausible substantial contribution of airborne transmission were explored[16–18], and attempts were made to mechanistically connect temperature and humidity to the environmental persistence of bio-aerosols and seasonality. Strikingly, the results of these influenza studies on aerosols are similar to those of initial SARS-CoV-2 investigations[19], implying a possible contribution of aerosols to long-range transmission. With the third—and in some cases the fifth—pandemic wave under way in the Northern Hemisphere, and the need to project future dynamics under vaccination and the different duration of immunity, the question of whether COVID-19 is a genuine seasonal disease becomes increasingly central, with implications for determining judicial intervention measures. More generally, the answer can inform our understanding of airborne transmission in other respiratory viruses.

Retrospective consideration of the first and second waves provides an opportunity to address climate drivers influencing the

[1]Climate and Health (CLIMA) Program, Barcelona Institute for Global Health (ISGlobal), Barcelona, Catalonia, Spain. [2]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA. [3]ICREA, Barcelona, Catalonia, Spain. ✉e-mail: xavier.rodo@isglobal.org

rising and waning phases of the epidemic, even in the presence of containment measures. A statistical method for transient correlations (scale-dependent correlation analysis, or SDC) is applied to the time series of reported COVID-19 cases to identify similar temporal variation to that of temperature and humidity over localized windows of time. The consistency of the results across temporal and spatial scales is also examined. Transitory associations can arise from nonlinear responses involving thresholds, whereby a climate factor acts as a more dominant limiting factor of transmission intensity in a given critical range. Because of its ability to detect small localized changes in epidemic shape, the statistical approach is able to uncover climate effects, despite confounding public health interventions and the very limited effect of trends at these scales (Methods)[20–22]. The new multiscale SDC (MSDC) implementation provides an opportunity to further confirm the consistency of associations, assessing their stability across scales. As a comparison, preferential meteorological ranges are also derived with the same approach for seasonal influenza in Japan. Finally, a complementary analysis of COVID-19 is implemented with a process-based epidemiological model for which the transmission rate is driven by climate. Its ability to capture observed cases and deaths is compared to that of alternate formulations with either a constant or a seasonal transmission term. We replicate this analysis at two different spatial scales—regional and city levels.

Results

Global and country-level role of temperature in COVID-19. To address the association of temperature (T) and AH in the initial phase of the invasion by SARS-CoV-2 and before changes in human behavior and public health policies were put into place, we first considered a measure of the initial growth rate of cases across countries. For this, we relied on an estimate of the reproductive number of the disease, R_0, in the 20 days after the first 20 cases of COVID-19 were officially notified for each of 162 countries spanning five continents (Extended Data Fig. 1, Methods and Supplementary Information). The global maps for the climate variables at those corresponding times in each country and for population density are shown in Extended Data Fig. 2. A negative relationship is observed between R_0 and T, as well as AH, at the global scale on the basis of linear pairwise regression models (Extended Data Fig. 1b,c and Extended Data Fig. 2). Although these regressions account for 10–25% of the variance in R_0, the two regressions are consistent and statistically significant ($P<0.001$), irrespective of the COVID-19 measure employed (for example, R_0 or cumulative cases; see also sensitivity tests in Extended Data Fig. 3).

To address whether the negative relationship with both T and AH is merely the result of a coincident timing between trends in epidemic spread and the dominant wintertime conditions in temperate regions, we analyzed the evolution of the disease–climate associations locally in time and at several disaggregated spatial scales for groups of countries in all five continents (Fig. 1). To this end, we relied on a statistical method, SDC, which was specifically developed to identify similar patterns of variation by means of local or transitory (linear) correlations between two time series given a moving window of time (Methods and refs. [20–22]). Application of this analysis to other diseases known to be climate-sensitive (for example, cholera, malaria and Kawasaki disease) has successfully shown that associations can be discontinuous in time, alternating temporal intervals with highly significant correlations and those with low or non-existent ones[23,24]. Such transient coupling can be expected in the response of nonlinear systems to an external driver, especially when the functions relating particular parameters to the driver are themselves nonlinear, so that effects are more evident in particular ranges of the forcing variable. SDC is essentially a pattern-recognition tool that allows one to consider the scale over which to evaluate correlations, as the window size can be systematically varied to focus on increasingly local patterns of variation, for example (Methods)[21]. A new extension of this method, MSDC, is presented in this study to examine the stability of correlations across all scales.

Figure 1 shows the results of SDC analyses for COVID-19 cases and T for time windows of two and a half months (a scale of $s=75$ days) when grouping countries according to region and latitude in the two hemispheres. Similar patterns are found for cases and AH (Extended Data Fig. 4). Strong negative transient associations are obtained for short time lags between the disease and climate time series, with consistent patterns worldwide. Interestingly, the negative relationship is seen during both the first and second waves of the pandemic, and for both the rising and declining phases, with a break during summertime in all continents (Fig. 1 and Extended Data Fig. 4). Transient positive correlations of varying intensity are also detected during the warmer months across locations, varying in intensity and not consistently high. Whether these reflect real patterns linked to mass gatherings of young people in vacation resorts, as in Spain in the summer 2021, is discussed later. To further examine the association patterns, we next considered the reported cases at the smaller spatial scale of individual countries in Europe (Fig. 2, for France, the United Kingdom, Italy, Spain and Germany), the first most affected continent following the emergence of the virus in China. SDC results show similar transient and negative associations for T and AH with COVID-19 cases ($s=75$ days). The negative relationship occurs largely in synchrony across the different countries, for the same time intervals during the waning of the first epidemic wave as T and AH rise, and also during the rise of the second wave in the fall as T and AH fall, with a break in between. For this same temporal window ($s=75$), a similar negative relationship holds also for individual regions within three of these highly affected European countries for which data at the higher spatial resolution of individual provinces are available (namely, Lombardy, Thüringen and Catalonia, Supplementary Figs. 1 and 2). These results are also shown separately for the first and second waves in Lombardy, Thüringen and Catalonia (Supplementary Fig. 2). Locally in time, these associations account for large fractions of the variability in COVID-19 cases in the three regions (over 80% at scales of $s=21$ days; Supplementary Fig. 2). Their discontinuity explains the lower values obtained with the traditional correlation coefficient, since by definition this quantity averages over the whole length of the time series[20].

We can also inspect how these transient couplings change as a function of the temporal lag (in days) between the two time series. As the magnitude of the change in the climate variable T or AH increases, the lag tends to shorten. As shown in Fig. 2b, this behavior is seen for the direction of change that is presumably relevant for influencing transmission (increasing T and AH in the fall of the first wave and decreasing T and AH in the rise of the second wave). The shorter lags may reflect a faster speed of community transmission, concomitant with a more intense effect of the meteorological factors. Zooming in even more locally in time with a window of about two weeks ($s=14$), SDC reveals limited intervals of very highly significant correlations at these regional levels for Lombardy (Italy; Fig. 3) and for Catalonia, Spain (Extended Data Fig. 5). Intervals where the climate covariate accounts for over 80–90% of the variability in cases alternate with complete decoupling between the time series. Interestingly, the strongest negative associations are observed during the rise of the second wave in these European regions. Positive associations are weaker and infrequent at this scale.

We next interrogate our SDC results to address the existence of threshold values in T and AH for the spread of COVID-19. To this end, we examine whether intervals of coupling become more frequent depending on the range of the climate covariates. This behavior would indicate the existence of a critical range within which responses are heightened and, conversely, outside which the forcing becomes ineffective. As an example, a meta-analysis was

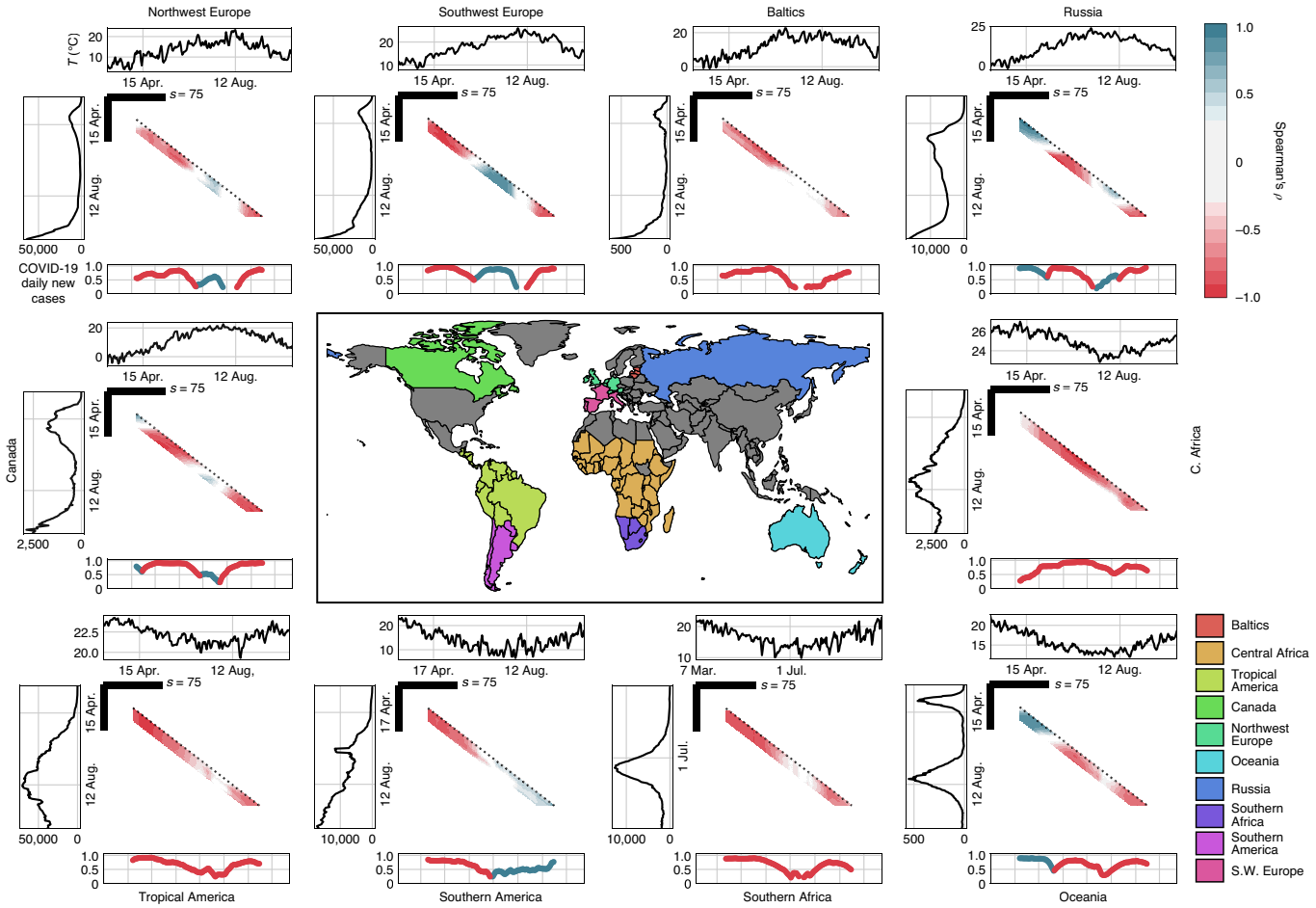


Fig. 1 | SDC analyses for the aggregated COVID-19 cases and temperature for groups of contiguous countries. Two-way SDC analyses for the time series of COVID-19 daily new cases versus mean temperature *T* in groups of contiguous countries. These daily climate conditions are weighted averages based on the relative population of the relevant countries from March to October 2020. The temporal scale or window size, *s*, used for the local correlations is 75 days, and the lag between the placement of the windows in the two time series varies from 0 to 21 days. In the SDC plots, the two time series are shown respectively to the left and top of the central correlation grid, for reported cases (left plots) with time running down and for the climate factor (top plots) with time running right. Each cell of the grid is colored according to the Spearman correlation coefficient, with the row and column corresponding to the position of the two respective time windows of size *s* along each of the time series, with the lag between these positions corresponding to the distances from the diagonal (lag 0 to +21 days). The panel below each plot shows the maximal correlation coefficient obtained at each time point (vertically, and therefore relative to the time of the climate time series). Only correlations found to be significant in a non-parametric randomized test ($\alpha = 0.05$) are shown and colored. Red (blue) points in the grid correspond to negative (positive) correlations, according to the specified scale (top right).

performed for Italy, where detailed regional case data are available. We jointly analyzed all SDC correlations between COVID-19 cases and each of the climate covariates (*T* or AH) by pooling together the results for all the individual regions of Italy (Fig. 3, Supplementary Fig. 5 and Methods; for Spain, results are shown in Extended Data Fig. 6). The resulting distributions, shown in Supplementary Fig. 4 and Extended Data Fig. 6, show the proportion of the total possible comparisons (pairs of time intervals between the two time series) that fall into a given range of the climate covariate, with those proportions subdivided into significant positive and negative correlations (Fig. 3e,f) and including non-significant ones (Supplementary Fig. 4). Distributions are also shown for only the positive and negative correlations along the range of climate intervals (Fig. 3e,f), and for the first two waves separately (Supplementary Figs. 5 and 6 and Extended Data Fig. 6). The mode in these distributions indicates preferential intervals for *T* and AH, respectively, where the climate effects are most evident and beyond which decoupling is likely (roughly 12–18 °C and 4–12 g m^{-3}, respectively; Fig. 3g,h). Although similar results are obtained for the first and second waves separately, these ranges should only be seen as indicative, given the short records yet available (Supplementary Fig. 6 and Extended Data Fig. 6).

To further examine the consistency of the identified relationships with *T* and AH, we applied MSDC—an extension of SDC developed to inspect the evolution of transient correlations at all scales at once in the same graphical display (Methods)—to countries in other continents that experienced a later arrival of SARS-CoV-2 than Europe, namely South Africa, Argentina and Canada, from February to December 2020 (Extended Data Fig. 7). The top and central plots in Extended Data Fig. 7 for each country present, respectively, the most negative and the most positive significant Spearman correlation found for every pair of fragments of a given window size and date. (To visualize timing with respect to the epidemic progression, the bottom plot displays the seven days moving average of daily reported new cases for each of the countries.) Negative correlations exhibit continuity as scale *s* is increased, and also transition to larger values (indications of causal effects). By contrast, positive correlations are much more limited in the range of scales, largely restricted to the very small scales and appearing either at times of minima or maxima of COVID-19 incidence, but not during the rise or the

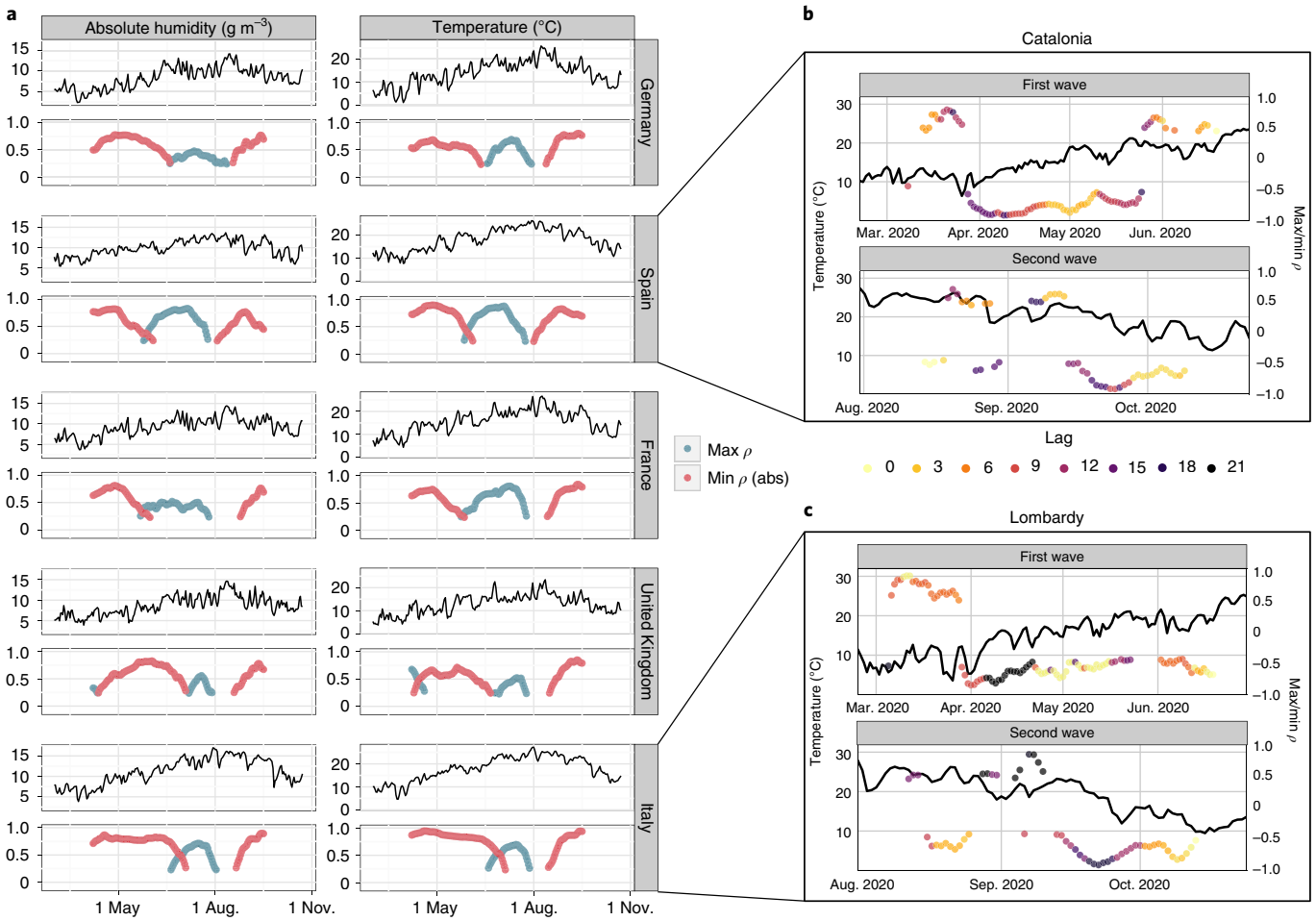


Fig. 2 | Transitory correlations between meteorological factors and COVID-19 incidence for several European countries. a, Panels from SDC analysis for the maximum absolute correlations between daily time series for the meteorological variable (AH, left; *T*, right) and COVID-19 new cases, as a function of time (the position of the local window of time). These panels correspond to those shown below the SDC grid of local correlations in Fig. 1, but here at the country level for five European countries (Germany, France, Italy, Spain and the United Kingdom). As before, $s = 75$, and the maximum correlation is chosen for lags from 0 to 21 days. The colors correspond to maximum negative (red) and positive (blue) correlations for each day. **b**, The resulting maximum correlations for the Spanish region of Catalonia with the corresponding time lag indicated. This dominant lag (the one for the maximum ρ value among the significant comparisons between a lag at 0 and +21 days) is overlaid on the temperature time series for each time point, with shorter lags being represented with lighter colors and longer lags with darker ones. The fragment size (s) used in this panel is 21 days, which allows us to localize and identify multiple shorter transient couplings between COVID-19 incidence and meteorological factors. **c**, As in **b**, but for the Italian region of Lombardy. Note that, in **b** and **c**, when the meteorological factor appears to be more related to the transmission of SARS-CoV-2 and the spread of COVID-19 in the population, the lag shortens, indicating a more intense and fast association.

fall of the pandemic waves. Total attributed variability is therefore large for negative correlations, but much more limited for the positive patterns. When the positive correlations are only present at the smallest scales and do not persist as we move up in scale, they most probably reflect spurious associations arising from random fluctuations (Extended Data Fig. 7).

A mechanistic model incorporating explicit temperature data. To further evaluate the role of climate factors on disease dynamics from a more mechanistic perspective, we also implemented a process-based stochastic model that incorporates an effect of temperature in the transmission rate. The model structure and corresponding parameters are shown in Fig. 4 (see Methods for additional description). The basic model, which divides the human population into seven compartments (susceptible, exposed, confined, infected, quarantined, recovered and dead individuals), already exhibited good performance in terms of predictive skill for the first waves of the COVID-19 pandemic in different countries[25].

Here, we extend its formulation to include temperature as a driver in the transmission rate. For simplicity, T was chosen for this purpose rather than AH based on its stronger relationship with COVID-19 in the above SDC results, and given the similarity of the temporal patterns of the two climate variables. For the purpose of comparison, the baseline formulation omits temperature and includes a constant transmission rate. A seasonally forced version includes a sinusoidal variation of this term with a period of 12 months. The model incorporating T is better able to capture the data for all the different waves and regions (Figs. 4 and 5 and Supplementary Table 1). Figure 4c,d shows model fits for the three different models for recovered cases, active cases and deaths in Catalonia and Lombardy, respectively. Comparisons of the mean squared error for these models are displayed in Fig. 5 and Extended Data Fig. 8. To address the different spatial scales of aggregation, the same mechanistic model was fitted to the epidemic data for the city of Barcelona (Fig. 5a,b and Extended Data Fig. 8). Here, too, the model incorporating T shows better performance overall

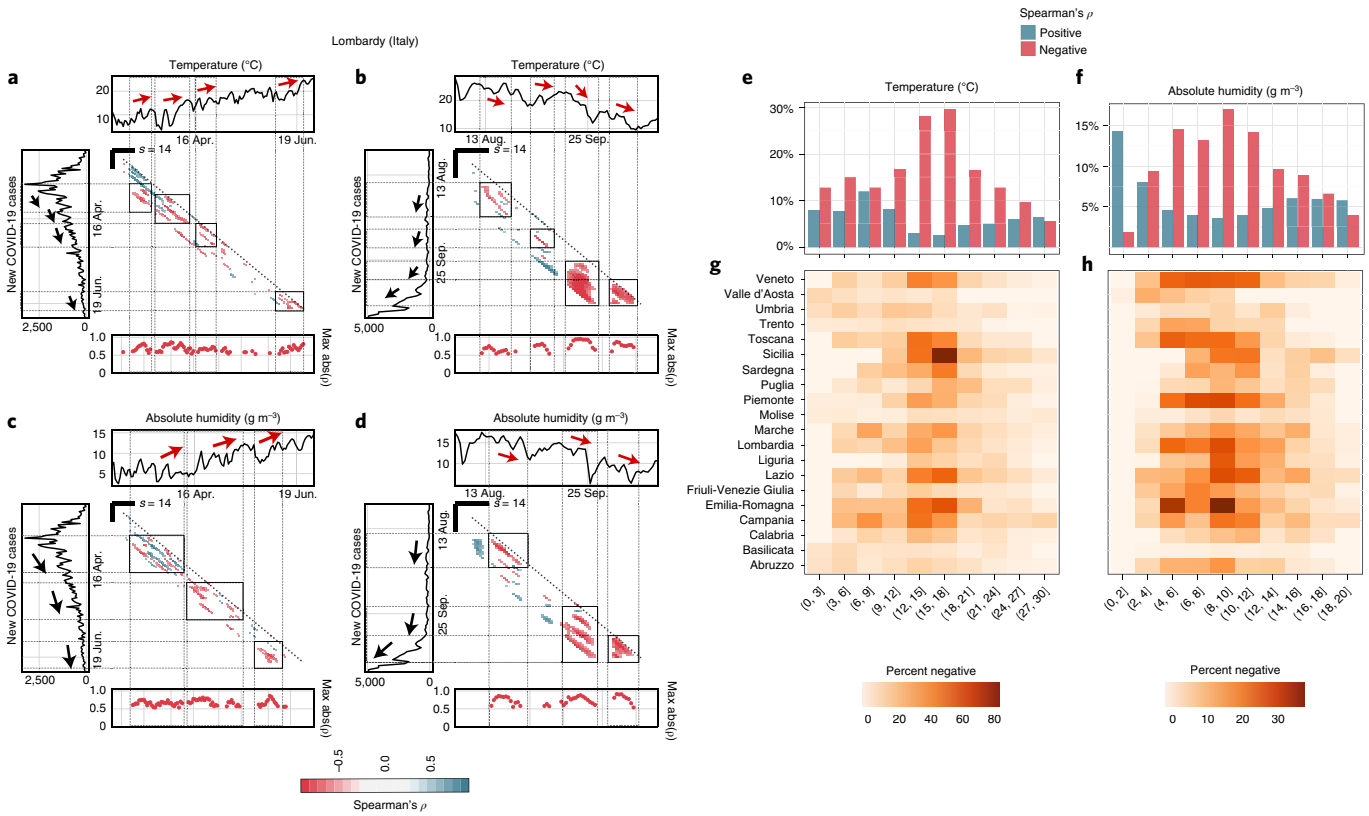


Fig. 3 | Analysis of the effect of *T* and AH in the spread of COVID-19 in Italy. a–d, Short-term transient correlations over time intervals of two weeks at the spatial resolution of region, here for Lombardy in Italy. As for Fig. 1, the graphical results of SDC analysis are shown for the correlations between daily COVID-19 cases and *T* (**a**,**b**) or AH (**c**,**d**) and for the different parts of the pandemic. Here, a smaller window size is used ($s = 14$ days) to examine even shorter-term transient associations. The panels in **a** and **c** showcase the results for the waning phase of the first wave (from late February to July), and the panels in **b** and **d** those for the rising phase of the second wave (from August to late October). The time intervals of significant correlations are indicated with boxes within the central grid, with the corresponding timings in the two time series highlighted with red arrows for the meteorological factor and with black arrows for the response variable, COVID-19 cases. The pairs of arrows for a given box show clear patterns of opposite temporal change (negative correlation) in the climate and epidemiological variables. For clarity, in this case the bottom panels below each plot show only the maximum negative correlation values obtained as the analysis window is moved in time (with red dots indicating negative correlations). **e**,**f**, Frequency distributions of negative and positive correlations as a function of range of the climate covariate (numbers in parentheses), for *T* (**e**) and AH (**f**). The frequency of windows with negative (or positive) correlations is obtained from SDC analysis of COVID-19 cases and the given climate factor in all Italian regions ($n = 20$; for $s = 21$ days). The percentages correspond to the percentages of total comparisons (within each of the ranges of the climate variables) that were statistically significant in either of the directions. The distributions are non-uniform with a mode in a particular range of the climate variables. This is indicative of a nonlinear relationship where the negative associations are most likely to be observed in a particular meteorological range. **g**,**h**, This same frequency distribution but across the different Italian regions. As can be seen, there is an overrepresentation of significant negative monotonic relationships in temperate conditions (AH of 4–12 g m^{-3}, *T* of 12–18 °C).

(Fig. 4c,d) and lower residual estimates than any other model (Fig. 5a,b and Supplementary Table 1). In particular, the fitting to the first wave appears clearly improved with respect to its counterparts, possibly indicating a stronger temperature control for this early epidemic stage. Although notable effects are seen also for the other waves (and are most apparent for the active cases), the improvements are modest in the second wave for Lombardy (Figs. 4c,d and 5a,d), Thüringen (Fig. 5a,b) and in the second and third waves for Catalonia (Figs. 4c and 5a,b) and Barcelona (Fig. 5a,b) (also Extended Data Fig. 8). These results are in strong agreement with those obtained with SDC analysis on the enhanced extraction of a *T* signature in this first COVID-19 pandemic wave.

Finally, the cutoff analysis conducted for COVID-19 (Supplementary Fig. 6) was also applied to records of seasonal influenza (A and B) in the city of Kawasaki in Japan from 2014 to 2021 (Fig. 6b). Seasonal influenza is a better known respiratory disease for which climate effects have been documented, in particular in relation to humidity (Extended Data Fig. 9). Transient negative correlations are clearly evident for influenza A and influenza B with both AH and *T* (Extended Data Fig. 9). As for COVID-19, the number of positive correlations is, by comparison, minimal (Extended Data Fig. 9). The ranges or thresholds for *T* and AH derived from the SDC analysis indicate (1) a slight systematic delay of up to one or two months in the occurrence of influenza B with respect to influenza A (Extended Data Fig. 9) and (2) *T* and AH relevant ranges similar to those obtained for COVID-19 (Fig. 6b). Interestingly, these ranges are slightly different between the two viral strains, suggesting small differences in climatic niches.

Discussion

Taken together, our findings support the existence of strong external drivers of transmission intensity, as suggested by a uniform summer recession despite a variety of intervention measures across countries[25]. They also support the view of COVID-19 as a true seasonal infection, similar in that sense to seasonal influenza and to the more benign circulating coronaviruses[26].

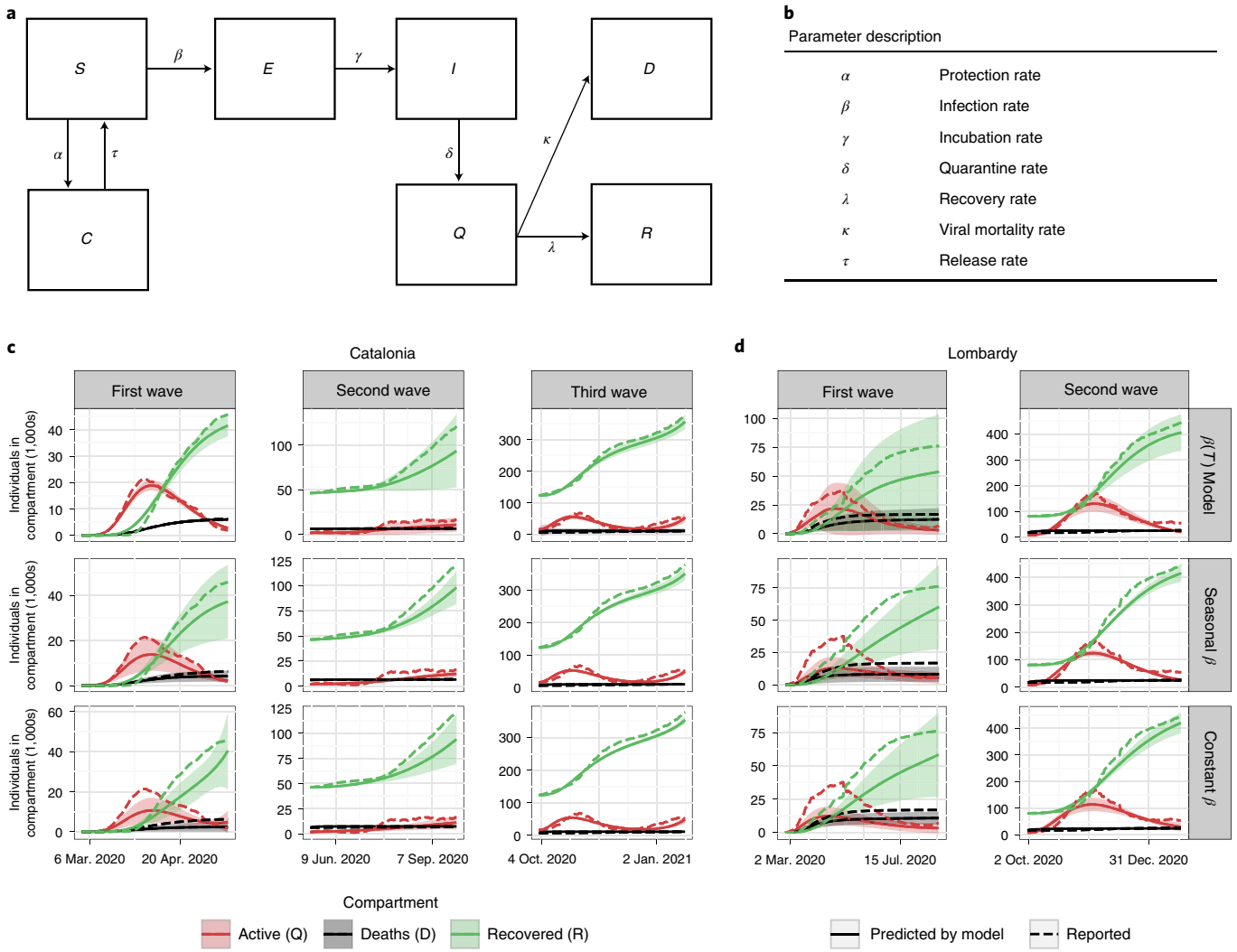


Fig. 4 | Compartmental process-based model diagram and fits. a,**b**, Diagram summarizing the structure of the compartmental process-based model (**a**) and table of the parameters for the transitions between these compartments (**b**). **c**,**d**, Model fitted values (solid lines, with the shaded areas representing the standard deviations (s.d.) of the predictions) compared to the officially reported values (dashed lines) for the active, recovered and deaths compartments in the regions of Catalonia (**c**) and Lombardy (**d**).

Concomitant temporal variation of ecological conditions, human behavior and disease incidence over the annual cycle can produce true but also spurious correlations with climate parameters[21]. Our results are consistent with those of ref. [27] on an effect of temperature and specific humidity on the spread of COVID-19, which is certainly encouraging given that the methods applied in this work (SDC) and their work (distributed-lag nonlinear models) are different. Our analysis is also complementary in considering data from different countries and spatial scales. The analyses in our study were formulated to circumvent the spurious correlations arising from such coincident seasonal variation, as well as effects of potential confounders (Supplementary Discussion). A similar approach could contribute to the further understanding of the complex global seasonality of influenza[28,29].

Further support for a transient but strong role of climate in the modulation of COVID-19 transmission is provided here by the results of the implemented mechanistic model. Model fits and model comparisons consistently indicate a climate effect across the different epidemic waves, for different regions and individual cities (Fig. 5 and Extended Data Fig. 8). The results for Barcelona are fully aligned with those for Catalonia and other regions. When inspecting results by waves, this observation is further reinforced, as the mean squared error for the temperature-driven model is always the lowest, especially for the first epidemic wave. Similarly, increasing the scale resolution does not modify the results by SDC at the largest spatial scales considered (those of region and country).

The problem of co-linearity in climate factors has challenged the search for environmental drivers of influenza in the past, and very similar issues now apply to COVID-19[27]. Both T and AH appear here equally capable of accounting for the observed COVID-19 variation. It is well known that air of rising temperature can contain more water, and for geophysical reasons these two parameters are usually strongly coupled over extensive spatial scales. Also, both parameters can act in the same direction in their biological effects. For example, the environmental persistence of an exhaled infectious virus particle—associated with a small (<5 µm) or a larger (100 µm) droplet[30]—it increases under laboratory conditions with lower temperatures and lower humidity for both influenza[31], SARS[32] and SARS-CoV-2[33]. Although prolonged environmental survival at lower temperatures applies to most of the viruses that have been studied[34], a less humid environment shortens the survival for most upper airway infections such as the rhino- and adenoviruses that cause the common cold[35]. Importantly, however, these viruses lack an envelope and are not seasonally defined[36]. By contrast, enveloped

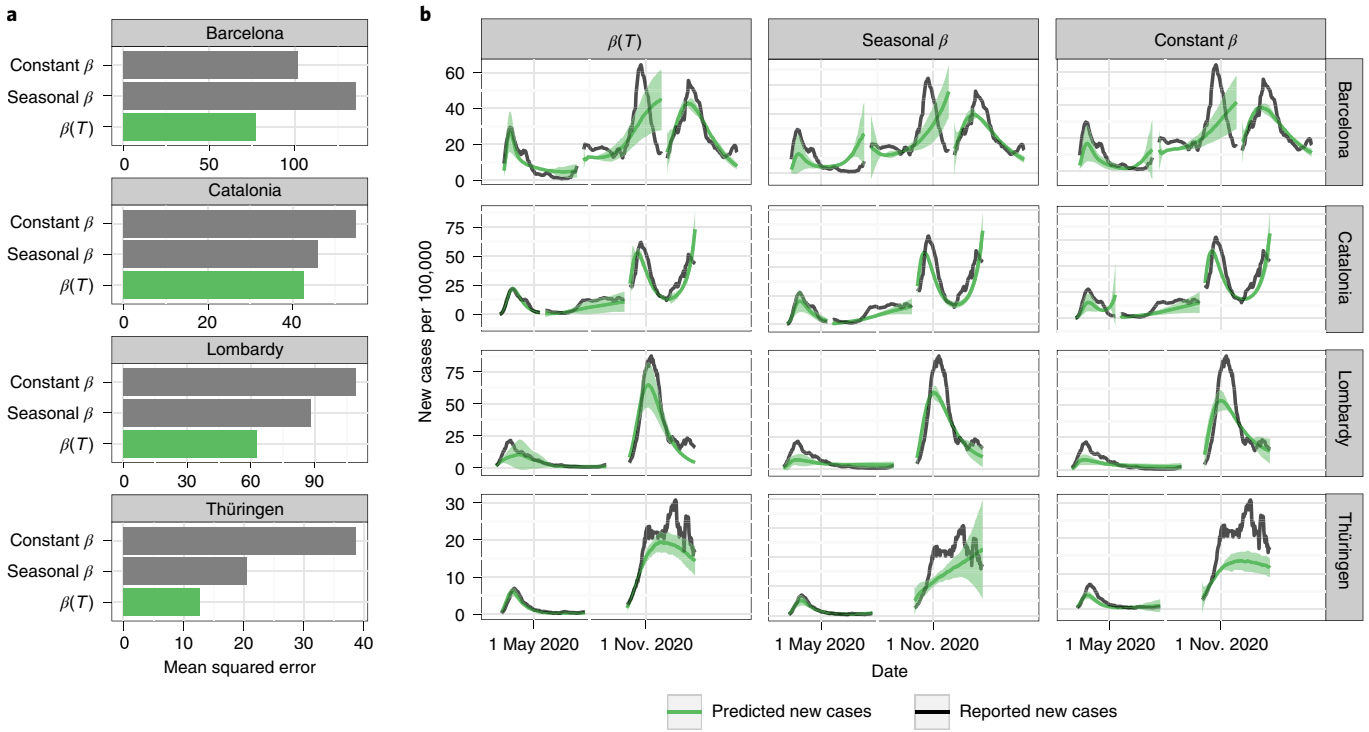


Fig. 5 | Goodness-of-fit comparison among the different model parameters. Comparison of the compartmental process-based model fits for the different regions, periods and parameters used. **a**, The mean squared error for the models fitted to the data across all periods and for each of the regions. The values corresponding to the models with the temperature-dependent transmission rate β are highlighted in green. **b**, The model predicted curves (in green) corresponding to the mean of 100 stochastic simulations are compared to the seven-day rolling mean of reported daily new cases in each region (in gray).

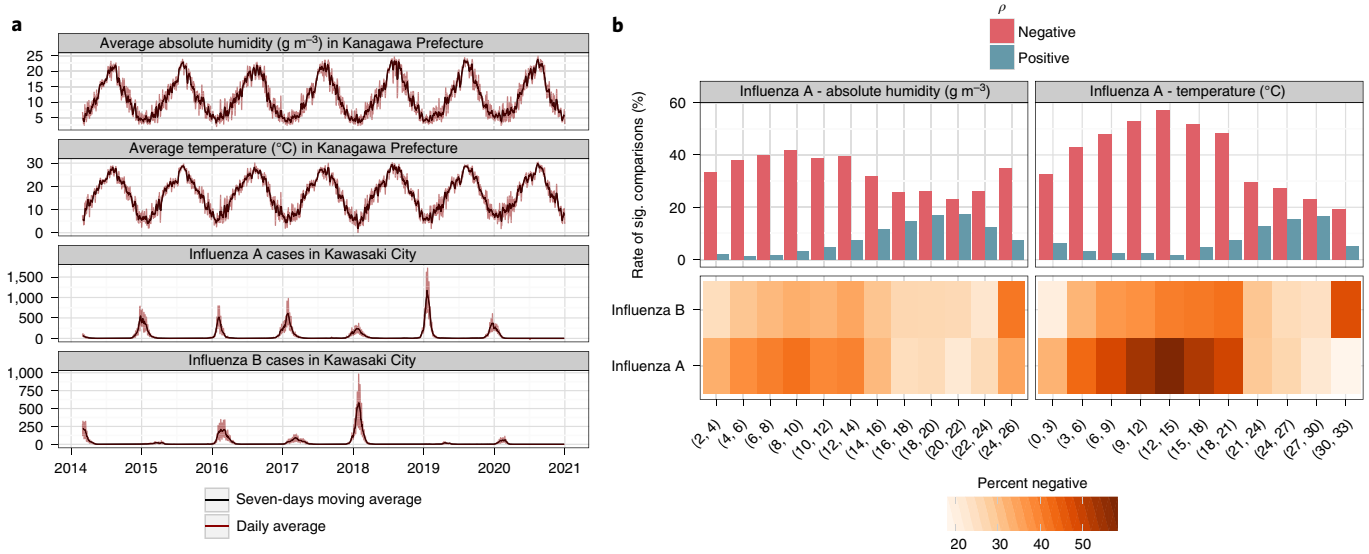


Fig. 6 | Comparison of weather patterns and influenza A and B incidences in Kawasaki City. a, Representation of daily averages of AH and *T* in the Japanese prefecture of Kanagawa, followed by the daily reported cases of influenza A and B in Kawasaki City from 2014 to 2020. **b**, Representation of the critical ranges of *T* and AH when performing an SDC analysis with $s = 30$ days, similar to the one shown in Fig. 3e–h.

viruses, including measles, variola and varicella, are infections reaching the lower airways, are prevalent during the cooler seasons and are predominantly or substantially airborne. It seems plausible that some viruses have adapted to this phase in the seasonal cycle and their concomitant features of temperature and humidity, ruling out the identification of a single or dominant driver of transmission. Our results point to SARS-CoV-2 belonging to this cluster of viruses. Regardless, plausible modes of action of temperature and humidity (Supplementary Discussion) suggest the importance of airborne transmission. Low humidity levels can reduce the size of bio-aerosols by evaporation when these cool-weather viruses are exhaled[37]. 'Droplet nuclei' smaller than 5 µm in size can bypass the nasal zone defenses and enter the deeper parts of the lungs. SARS and SARS-CoV-2 specifically target the angiotensin-converting

enzyme 2 protein for cell entry[38], where this protein is most abundant in alveolar tissue[39]. During the colder winter months in higher latitudes, indoor heating generates a microclimate with low levels of RH, a parameter that better reflects the drying power of air than AH. However, both indoor AH and RH in heated accommodation correspond well with the routinely measured outdoor AH in colder climates[18], and therefore also the gridded climatology used here for studying this parameter.

The permissive role of low temperatures for the transmission of SARS-CoV-2 appears to contradict the warm-weather-related transmission and severity of the second wave in the United States for July–August 2020 (not shown) and the 2021 summer wave, as well as the second or third waves in 2021 in tropical countries such as India and in Latin America. In rich countries of the Northern Hemisphere, the cooled indoor microclimate during these months could accommodate transmission, as in the many outbreaks observed in the meat-processing plants of the United States[40] and Europe[41] where cooled air is mechanically reventilated. In countries such as Argentina, Brazil, Colombia and Peru, the limited vaccine supply with its slow roll-out, weak health systems and fragile economies that make stay-at-home orders difficult to impose or maintain can all combine to explain the recent surges in spring of 2021. The implications, and the possibility that new variants emerging at these latitudes might be better adapted to stable tropical conditions, remain to be thoroughly examined.

The positive correlations we found for the summer months of 2020 (for example, Fig. 1) during the seasonal low of cases may also reflect community transmission in public temperature-controlled buildings such as care homes[41,42]. They may also result from massive social clustering and increased contact during warm-weather leisure activities outdoors[43]. Indeed, air conditioning in the subtropics was also proposed to explain SARS transmission[32], a situation bearing similarities with influenza[44]. Similarly, the Northern Hemisphere experienced an important uptick in cases in summer of 2021 in countries such as Spain and Portugal, the target of massive seaside tourism in Europe. Regardless of explanation, these positive associations were less frequent overall, and less robust and weaker when zooming into smaller temporal window sizes.

Our study has several limitations, the most important one stemming from the limited length of the epidemiological records so far available. Also limiting are the different social and environmental forcing factors influencing the epidemiology of COVID-19, which can be country-dependent. When SDC is employed at a single scale of analysis, it can potentially identify spurious correlations, an aspect largely circumvented by the new multiscale analysis. Interestingly, SDC sensitivity can help track other locally relevant epidemiological dynamics potentially linked to human behavior, such as summer gatherings in 2020 inside building premises or under the effect of air conditioning in hot climates, producing positive associations. Such associations do not arise from direct meteorological preferences of the virus but from effects of human behavior on virus transmission. They would not be detected by fixed regression methods. Data-reporting differences across locations and different sensitivities of virus strains might alter the current results, as data for these new variants (delta, delta-plus, lambda) have not been analyzed here. Variation in reporting errors add uncertainty to the analysis, although SDC appears robust to identifying associations even in the presence of a strong weekly cycle. This robustness opens the door to the development, in the near future, of more curated and tailored climate services and early-warning systems for COVID-19.

The identified climate forcing should persist for novel virus variants with increased transmissibility and an enhanced ability to partially evade protection from previous exposure. It should also persist at lower incidence levels were the disease to become endemic, therefore defining the annual timing for vaccination. Strong current policies to curtail transmission, including lockdowns, where effective, should limit the role of climate drivers. This effect can already be seen when considering the most recent reported cases through the winter in regions of Italy (Extended Data Fig. 6). The frequency of local significant correlations with our method decreases markedly and the critical range of negative correlations is no longer evident (Extended Data Fig. 6b). By contrast, for regions of Spain, where control efforts have been more variable in implementation, this range remains apparent and consistent with the earlier waves (Extended Data Fig. 6b), even though climate conditions are shared by the two countries (Extended Data Fig. 6c). Public health interventions to curb transmission of COVID-19 have focused in this initial phase on reducing the contact rate between people through social distancing, and on hand washing, decontamination of infected surfaces and face covering. With the exception of face covering, these measures have emphasized the importance of short-range transmission, the default assumption on respiratory diseases caused by exhaled droplets[45] with a restricted spatial range[46]. The role of smaller exhaled droplets that become airborne as bio-aerosols or 'droplet nuclei' for extended periods, and may cover long distances in a viable state, has been proposed for COVID-19 but has remained unresolved[47–50]. Seasonality, and the role of low temperatures and the associated low humidity, can be mechanistically linked to viable SARS-CoV-2 aerosols, supporting the relevance of the airborne transmission route implicated in other studies[51]. This link warrants an emphasis on 'air hygiene' through improved indoor ventilation[52] to more effectively intervene in the unfolding pandemic. It also underscores the need to include appropriate meteorological parameters in the evaluation and planning of pharmacological and behavioral control measures.

Methods

Global statistical analysis. Our first attempt to identify plausible effects of meteorological covariates on COVID-19 spread applied a comparative regression analysis. To this end, we focused on the exponential onset of the disease, as it is the epidemic phase that allows for a better comparison between countries or regions, without the confounding effect of intervention policies. We first determined, for each of the spatial units (either countries or NUTS (nomenclature of territorial units for statistics) 2 regions), the day in which 20 or more cumulative cases were officially reported. We then fitted the first-order polynomial function $f(t)=x_0+rt$ for the next 20 days of log-transformed data, where t represents time (in days) and x_0 is the value at initial condition $t=0$. The r parameter can be understood as the exponential growth rate, and is then used to estimate the basic reproduction number (R_0) using the estimated serial interval T for COVID-19 of 4.7 days[53], such that $R_0=1+rT$ (ref. [54]). (We note that we are interested here in the relationship between the reproductive number and not in the actual inference of R_0.) Once R_0 was obtained for all our spatial units, we filtered our meteorological data to match the same fitting period (with a 10-day negative delay to account for an incubation and reporting lapse) for every spatial unit. To compute a single average of the meteorological variables per regional unit, we computed a weighted average on the basis of the population contribution of each grid cell to the total population of the region. We did so to have an aggregated value that would better represent the impact of these factors on the population transmission of COVID-19, as the same variation in weather in a high-density urban area is more likely to contribute to a change in population-level transmission than that of an unpopulated rural area. We then averaged the daily values of temperature and AH for each country and computed univariate linear models for each of these variables as predictors of R_0. Given the somewhat arbitrary criteria to select the dates to estimate the R_0 in each country, a sensitivity analysis was run to test the robustness of the regressions to changes in the related parameters. We tested 70 different combinations of two parameters: the total number of days used for the fit (18–27) and the threshold of cumulative COVID-19 cases used to select the initial day of the fit (15–45). We also calculated the weather averages by shifting the selected dates accordingly. Then, a linear model for each of the estimates was fitted for both T and AH. A summary of the distribution of parameter estimates (the regression slope coefficients and the R^2 of the models) is shown in Extended Data Fig. 3.

Bivariate time-series analysis with scale-dependent correlations. To examine associations between cases and climate factors in more detail, SDC was performed on the daily time series of both COVID-19 incidence and a given meteorological variable. SDC is an optimal method for identifying dynamical couplings in short and noisy time series[20,21]. In general, Spearman correlations between incidence and a meteorological time series assess whether there is a monotonic relation between the variables. SDC analysis was specifically developed to study transitory

associations that are local in time at a specified temporal scale corresponding to the size of the time intervals considered (s). The two-way implementation (TW-SDC) is a bivariate method that computes non-parametric Spearman rank correlations between two time series, for different pairs of time intervals along these series. Different window sizes (s) can be used to examine increasingly finer temporal resolution. The results are sensitive to the value of this window size, s, with expected significant and highest correlation values at the scale of the transient coupling between variables. Correlation values decrease in magnitude as window size increases, and averages are computed over too long a time interval. Values can also decrease and become non-significant for small windows when correlations are spurious. Here, the method was applied for windows of different length (from $s=75$ to 14 days) and, despite a weekly cycle showing up in some cases for small s, results removing this cycle were robust. We therefore did not remove this cycle.

The results are typically displayed in a figure with the following subplots: (1) the two time series, to the left and top of the matrix of correlation values, respectively; (2) the matrix or grid of correlation values itself in the center, with significant correlations colored in blue when positive and in red when negative, with rows and columns corresponding to the temporal localization of the moving window along the time series on the left and top, respectively; (3) a time series at the bottom, below this grid, with the highest significant correlations for a given time (vertically, and therefore for the variable that acts as the driver, here the meteorological time series). To read the results, one starts at the diagonal and moves vertically down from it to identify a given lag for which significant correlations are found (the closest to the main diagonal). In some of the SDC figures, the time intervals with high local correlations are highlighted with boxes. These intervals alternate with other ones (left blank) for which no significant correlation is found. All colored areas correspond to significance levels of at least $P<0.05$. A new presentation of results is also used for the influenza analysis, in which the two time series are superimposed in the same plot with the significant correlations shown in a panel below as a function of time and lag (Extended Data Fig. 9).

Significance is assessed with a non-parametric randomization test (see ref. [20] for further details and for examples illustrating the method). For the baseline test, SDC calculates Spearman correlations (at $\alpha=0.05$) between two white-noise time series at each fragment size s for a non-parametric permutation test (Supplementary Fig. 7). The indices of the series are randomly re-ordered, breaking their temporal shape. This permutation test enables a first estimate of the probability of finding significant spurious correlations, and it can thus be used as a non-parametric significance test for pairs of any length for the time series of interest. As seen in Supplementary Fig. 7, the threshold decays rapidly as the fragment size grows, with high values becoming rarer the longer the time series are compared. As a second test, SDC evaluates the unwanted effects of the internal autocorrelation in a time series. This effect might artificially inflate the correlation obtained between two time series, and should therefore be taken into consideration properly in a significance comparison. To estimate this effect, we generate pairs of autocorrelated (red) noise series of 365 time steps, with a mean $\mu=0$ and standard deviation $\sigma=1$ and varying degrees of autocorrelation (autocorrelation parameter φ in Supplementary Fig. 8) from 0 (equivalent to white noise) to 0.95, in steps of 0.05. We repeated this procedure 20 times for every unique combination of parameters to achieve a robust estimate. The method then searches for significant couplings in either direction. This is carried out for each of these synthetic time-series pairs with an SDC analysis (for example, with $s=30$), yielding, for each of them, the false discovery rate, which accounts for a type I error and provides the rate of significant couplings at $\alpha=0.05$. In Supplementary Fig. 8b, we show the average false discovery rate of the tests for each pair of autocorrelation values. As shown, the chance of finding a spurious coupling increases monotonically as a function of the autocorrelation evaluated.

Our approach focuses on analyzing temporal associations in one location at a time, and comparing across these locations the patterns of association themselves, including their timing (for example, in the waning or rising phase of epidemic waves). This allows comparison of results among distinct regions, despite differences in control measures and disease epidemic state. Time-series analyses applied to each location do not present the typical problems of comparative spatial regression studies, which might be biased by uncontrollable confounding effects across spatial locations.

One reason why couplings between two variables in ecological or epidemiological systems may be transient couplings is the existence of thresholds above or below which responses to forcing are weak or absent. To interrogate our analyses for the existence of critical thresholds/ranges for optimal transmission of the virus (inferred from COVID-19 disease outcomes), we pooled all negative and positive significant correlations performed at a size $s=21$ days between each meteorological variable and COVID-19 cases. We then computed the proportion of those negative correlations among all possible correlations for a given fragment size obtained for each bin of T or AH values and plotted their distribution (Fig. 3 and Supplementary Figs. 4–6) for all individual regions in Italy.

Singular spectrum analysis (SSA) involves the spectral decomposition (eigenvalues and corresponding eigenvectors) of a covariance matrix obtained by lagging the time-series data for a prescribed number of lags M called the embedding dimension. There are two crucial steps in this analysis for which there are no formal results, but useful rules of thumb: (1) the choice of M and (2) the grouping of the eigenvectors to define the specific major components and reconstruct them. Typically, the grouping of the eigencomponents is based on the similarity and magnitudes of the eigenvalues, their power (variance of the data they account for) and the peak frequency of the resulting reconstructed components. For selection of the embedding dimension, one general strategy is to choose it so that at least one period of the lowest-frequency component of interest can be identified, that is, $M>f_s/f_c$, where f_s is the sampling rate and f_c the minimum frequency. Another strategy is that M be large enough that the M-lagged vector incorporates the temporal scale of the time series that is of interest. The larger the M, the more detailed the resulting decomposition of the signal. In particular, the most detailed decomposition is achieved when the embedding dimension is approximately equal to half of the total signal length. A compromise must be reached, however, as a large M implies increased computation, and too large a value may produce mixing of components. SSA is especially well suited for separating components corresponding to different frequencies in nonlinear systems. Here, we applied it to remove the weekly cycle.

MSDC analysis. MSDC provides a scan of the SDC analyses over a range of different scales (here, S from 5 to 100 days at 5-day intervals), by selecting the maximum correlation values (positive or negative) closer to the diagonal. The goal is to consider the evolution of transient correlations at all scales pooled together in a single analysis. The MSDC plot displays time on the x axis and scale (S) on the y axis, and positive and negative correlations either jointly or separately. The rationale behind MSDC is that correlations at very small scales can occur by chance because of coincident similar patterns, but that as one moves up to larger scales (by increasing S), the correlation patterns that are spurious tend to vanish, whereas those reflecting mechanistic links increase in strength. This increase in correlation values should occur up to the real scale of interaction, decreasing afterwards. By 'real', we mean here the temporal scale covering the extent of the interaction between the driver and the response process (in this case, the response of disease transmission to a given climate factor). Thus, continuity of the same sign correlations together with transitions to larger values are indicative of causal effects, whereas the rapid vanishing of small-scale significant correlations signals spurious ones.

Process-based model. *Description.* The dynamical model is a discrete stochastic model that incorporates seven different compartments: S, E, I, C, Q, R and D. The model structure is illustrated in Fig. 4. The transition probabilities of the stochastic model are based on the corresponding rates of the transitions between classes in the deterministic (mean-field) model (specified in Fig. 4b). These probabilities are defined as follows. $P(e)=(1.0-\exp(-\beta\,\mathrm{d}t))$ is the probability of infection exposure of the susceptible class, where $\beta=(1/N)(\beta_I I+\beta_Q Q)$ is the infection rate (of the deterministic model). $P(i)=(1.0-\exp(-\gamma\,\mathrm{d}t))$ is the probability that an new exposed individual becomes infectious, where γ denotes the incubation rate. $P(r)=(1.0-\exp(-\Lambda\,\mathrm{d}t))$ is the recovery probability, where $\lambda_0(1-\exp(\lambda_1 t))$ is the (deterministic) recovery rate. $P(p)=(1.0-\exp(-\alpha\,\mathrm{d}t))$ is the protection probability, where $\alpha=\alpha_0\exp(\alpha_1 t)$. $P(d)=(1.0-\exp(-K\,\mathrm{d}t))$ is the mortality probability, with $K=k_0\exp(k_1 t)$. $P(re)=(1.0-\exp(-\tau\,\mathrm{d}t))$ is the release probability from confinement, where $\tau=\tau_0\exp(\tau_1 t)$. Finally, $P(q)=(1.0-\exp(-\delta\,\mathrm{d}t))$ is the detection probability, where δ is the quarantine rate (for example, at which infected individuals are isolated from the rest of the population).

In the model, both infected non-detected and infected detected individuals can infect susceptible ones. In the model incorporating temperature in the transmission rate, the respective values of β_I and β_Q are calculated as follows:

$$\beta_I(t) = \beta_I\, T_{\mathrm{inv}}(t); \quad \beta_Q(t) = \beta_Q\, T_{\mathrm{inv}}(t)$$

where $T_{\mathrm{inv}} = f\left(\frac{1-T(t)}{\bar{T}}\right)$, with $\bar{T}$ corresponding to the overall mean of the temperature time series and $f(\cdot)$ to a Savitzky–Golay filter, used to smooth the temperature series with a window size of 50 data points and a polynomial order of 3. When the infection rate is constant, we simply omit the temperature term. For further comparison, in a third model, β is specified with a sinusoidal function of period equal to 12 months and an estimated phase.

The number of individuals transitioning from compartment i to j at time t are determined by means of binomial distributions $P(Xi,P(y))$, where Xi corresponds to one of the compartments S, E, I, Q, R, D, C, and $P(y)$ to the respective transition probability defined above. Thus,

- $e(t)=P(S(t), P(e))$, new exposed individuals at time t
- $p(t)=P(S(t), P(p))$, protected individuals at time t
- $i(t)=P(E(t), P(i))$, new infected not detected individuals at time t
- $q(t)=P(I(t), P(q))$, new infected and detected individuals at time t
- $r(t)=P(Q(t), P(r))$, total recovered individuals at time t
- $d(t)=P(Q(t), P(d))$, total dead individuals at time t
- $re(t)=P(C(t), P(re))$, individuals released from confinement at time t

Then, the final dynamics are given by the following equations:

$$S(t) = S(t-\mathrm{d}t) - e(t) - p(t) + re(t)$$

$$E(t) = E(t-\mathrm{d}t) + e(t) - i(t)$$

$$I(t) = I(t-\mathrm{d}t) + i(t) - q(t)$$

$$Q(t) = Q(t - \mathrm{d}t) + q(t) - r(t) - d(t)$$

$$R(t) = R(t - \mathrm{d}t) + r(t)$$

$$D(t) = D(t - \mathrm{d}t) + d(t)$$

$$C(t) = C(t - \mathrm{d}t) + p(t) - re(t)$$

Calibration. The model was implemented using Python and calibrated by means of the least squares algorithm of the scipy library. The error function minimized with this algorithm was obtained from the normalized residuals on the basis of total cases (Q + R + D) and deaths (D).

To search parameter space, we ran 100 calibrations starting from different initial choices of parameter combinations. The tolerance for termination in the change of the cost function was set to 1×10^{-10}. Tolerance for termination by the norm of the gradient was also set to 1×10^{-10}, and the tolerance for termination by the change of the independent variables was set to 1×10^{-10}. The solver was the lsmr method (which is suitable for problems with sparse and large Jacobian matrices) with a differential step of 1×10^{-5}. With this configuration, each fitting run usually converged after ~500 iterations.

Validation. To compare the model including an effect of T in the transmission rate to those without it, we calculated the chi-square, Akaike information criterion (AIC) and Bayesian information criterion (BIC) indices for the residuals obtained from the optimization process. The resulting values are shown in Supplementary Table 1.

Our choice of T to modulate the infection rate (β) instead of AH underlies the fact that the temporal dynamics of both factors roughly follow the same shape, with the advantage that T shows less oscillatory behavior than AH. This fact adds stability to the model when the inverse relationship is used in the calculation of β (Supplementary Information). This selection is further reinforced by the results from the SDC analyses, which yielded larger correlations for temperature, even when penalizing for the larger autocorrelation structure.

Our choice to modulate β using T instead of AH follows from the fact that the temporal dynamics of both climate variables present roughly the same shape, with the advantage that T exhibits weaker oscillations. This less fluctuating pattern provides stability to the model fitting when the inverse relationship is used in the calculation of β (Supplementary Information). Additionally, the transient correlations obtained with SDC yielded higher values for T than for AH (even when accounting for concurrent levels of autoregression in the two variables).

Data availability

Data on COVID-19 daily incidence counts were retrieved from the ECDC dataset[55] for the nationally aggregated data and from the COVID-19 Data Hub[56] for the NUTS 2-disaggregated incidence data. National and regional boundaries were obtained from shapefiles provided by GISCO services by Eurostat[57]. Meteorological data were retrieved from ERA5 reanalysis through API requests to the Copernicus Climate Data Store (CDS)[58]. We obtained temperature (at 2 m), sea level pressure, total precipitation and dew point temperature data for a 0.5° × 0.5° grid for the country-wise analysis and for a 0.25° × 0.25° grid for the rest. From the 2 m temperature and the dew point temperature, we derived relative humidity, and used the atmos Python package[59] to calculate AH. The Gridded Population of the World (GPWv4), developed by the Center for International Earth Science Information Network (CIESIN) at Columbia University, was obtained from ref. [60] at a resolution of 2.5 min. We did not include data from the United States in our analyses given the large geographical extent of this country, which includes different climatic zones, and the largely asynchronous implementation of intervention policies across its different states. Data for influenza A and B incidence in the city of Kawasaki were retrieved from the Real-time Surveillance website (https://kidss.city.kawasaki.jp/en/realsurveillance/opendata) of the Japanese National Epidemiological Surveillance of Infectious Diseases[61] as total daily reported cases from 1 March 2014 to 31 December 2020. Daily temperature averages were obtained by averaging the hourly values for the same period obtained via the ERA5 reanalysis for grid cells spanning the municipality of Kawasaki (coordinates (139.5, 35.5) and (139.75, 35.5) in a 0.25° × 0.25° grid). Source data for Figs. 1 to 5 and Extended Data Figs. 1 to 6 are available in a Code Ocean repository[62]. Source data for Figure 6 are provided with this paper.

Code availability

Code for the SDC analyses has been implemented as a Python package available in PyPi (sdcpy[63]). Code to ensure the reproducibility of the analysis is available in Code Ocean[62] and GitHub (https://github.com/AlFontal/covid_climate_signatures).

Received: 14 May 2021; Accepted: 31 August 2021;
Published online: 21 October 2021

References

1. Fan, V. Y., Jamison, D. T. & Summers, L. H. Pandemic risk: how large are the expected losses? *Bull. World Health Org.* **96**, 129–134 (2018).
2. Shao, W., Li, X., Goraya, M. U., Wang, S. & Chen, J.-L. Evolution of influenza a virus by mutation and re-assortment. *Int. J. Mol. Sci.* **18**, 1650 (2017).
3. Roy, C. & Milton, D. Airborne transmission of communicable infection—the elusive pathway. *N. Engl. J. Med.* **350**, 1710–1712 (2004).
4. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
5. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
6. WHO Director-General's opening remarks at the media briefing on COVID-19. *WHO* (11 March 2020); https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020
7. Worldometer Coronavirus Update (Live): 62,125,065 Cases and 1,451,937 Deaths from COVID-19 Virus Pandemic (accessed September 18 2021) https://www.worldometers.info/coronavirus/#countries
8. Faust, J. S. & del Rio, C. Assessment of deaths from COVID-19 and from seasonal influenza. *JAMA Intern. Med.* **180**, 1045–1046 (2020).
9. Vittecoq, M. et al. Does the weather play a role in the spread of pandemic influenza? A study of H1N1pdm09 infections in France during 2009–2010. *Epidemiol. Infect.* **143**, 3384–3393 (2015).
10. Carlson, C. J., Gomez, A. C. R., Bansal, S. & Ryan, S. J. Misconceptions about weather and seasonality must not misguide COVID-19 response. *Nat. Commun.* **11**, 4312 (2020).
11. Baker, R. E., Yang, W., Vecchi, G. A., Metcalf, C. J. E. & Grenfell, B. T. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* **369**, 315–319 (2020).
12. Sajadi, M.M. et al. Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (COVID-19). *JAMA network open* **3**, e2011834–e2011834 (2020).
13. Ma, Y. et al. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Sci. Total Environ.* **724**, 138226 (2020).
14. Bashir, M. F. et al. Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Sci. Total Environ.* **728**, 138835 (2020).
15. Yu, I. T. et al. Evidence of airborne transmission of the severe acute respiratory syndrome virus. *N. Engl. J. Med.* **350**, 1731–1739 (2004).
16. Shaman, J. & Kohn, M. Absolute humidity modulates influenza survival, transmission and seasonality. *Proc. Natl Acad. Sci. USA* **106**, 3243–3248 (2009).
17. Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T. & Lipsitch, M. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, e1000316 (2010).
18. Marr, L. C., Tang, J. W., Van Mullekom, J. & Lakdawala, S. S. Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. *J. R. Soc. Interface* **16**, 20180298 (2019).
19. Liu, Y. et al. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature* **582**, 557–560 (2020).
20. Rodo, X. & Rodriguez Arias, M. A. A new method to detect transitory signatures and local time/space variability structures in the climate system: the scale-dependent correlation analysis. *Clim. Dyn.* **27**, 441–458 (2006).
21. Rodríguez-Arias, M. N. & Rodó, X. A primer on the study of transitory dynamics in ecological series using the scale-dependent correlation analysis. *Oecologia* **138**, 485–504 (2004).
22. Rodó, X. Reversal of three global atmospheric fields linking changes in SST anomalies in the Pacific, Atlantic and Indian oceans at tropical latitudes and midlatitudes. *Clim. Dyn.* **18**, 203–217 (2001).
23. Rodó, X., Pascual, M., Fuchs, G. & Faruque, A. S. G. ENSO and cholera: a nonstationary link related to climate change? *Proc. Natl Acad. Sci. USA* **99**, 12901–12906 (2002).
24. Rodó, X. et al. Tropospheric winds from northeastern China carry the etiologic agent of Kawasaki disease from its source to Japan. *Proc. Natl Acad. Sci. USA* **111**, 7952–7957 (2014).
25. López, L. & Rodó, X. The end of social confinement and COVID-19 re-emergence risk. *Nat. Hum. Behav.* **4**, 746–755 (2020).
26. Audi, A. et al. Seasonality of respiratory viral infections: will COVID-19 follow suit?. *Front. Public Health* **8**, 567184 (2020).
27. Ma, Y., Pei, S., Shaman, J., Dubrow, R. & Chen, K. Role of meteorological factors in the transmission of SARS-CoV-2 in the United States. *Nat. Commun.* **12**, 3602 (2021).
28. Lipsitch, M. & Viboud, C. Influenza seasonality: lifting the fog. *Proc. Natl Acad. Sci. USA* **106**, 3645–3646 (2009).
29. Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S. & Sugihara, G. Global environmental drivers of influenza. *Proc. Natl Acad. Sci. USA* **113**, 13081–13086 (2016).
30. *Airborne Transmission of SARS CoV 2—A Virtual Workshop* (The National Academies of Science, Engineering and Medicine, 2020); https://www.

nationalacademies.org/event/08-26-2020/airborne-transmission-of-sars-cov-2-a-virtual-workshop
31. Irwin, C. K. et al. Using the systematic review methodology to evaluate factors that influence the persistence of influenza virus in environmental matrices. *Appl. Environ. Microbiol.* **77**, 1049–1060 (2011).
32. Chan, K. H. et al. The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Adv. Virol.* **2011**, 734690 (2011).
33. Morris, D. H. et al. Mechanistic theory predicts the effects of temperature and humidity on inactivation of SARS-CoV-2 and other enveloped viruses. *Elife* **10**, e65902 (2021).
34. Woese, C. Thermal inactivation of animal viruses. *Ann. N. Y. Acad. Sci.* **83**, 741–751 (1960).
35. Boone, S. A. & Gerba, C. P. Significance of fomites in the spread of respiratory and enteric viral disease. *Appl. Environ. Microbiol.* **73**, 1687–1696 (2007).
36. Moriyama, M., Hugentobler, W. J. & Iwasaki, A. Seasonality of respiratory viral infections. *Annu. Rev. Virol.* **7**, 83–101 (2020).
37. Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog.* **3**, 1470–1476 (2007).
38. Zhao, Y. et al. Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. *Am. J. Resp. Crit. Care Med.* **202**, 756–759 (2020).
39. Hamming, I. et al. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol.* **203**, 631–637 (2004).
40. Dyal, J. W. COVID-19 among workers in meat and poultry processing facilities-19 states, April 2020. *Morb. Mortal. Wkly Rep* **69**, 557–561 (2020).
41. ECDC Heating, Ventilation and Air-Conditioning Systems in the Context of COVID-19: First Update. *ECDC* (11 November 2020); https://www.ecdc.europa.eu/en/publications-data/heating-ventilation-air-conditioning-systems-covid-19
42. Lynch, R. M. & Goring, R. Practical steps to improve air flow in long-term care resident rooms to reduce COVID-19 infection risk. *J. Am. Med. Directors Assoc.* **21**, 893–894 (2020).
43. Azuma, K., Kagi, N., Kim, H. & Hayashi, M. Impact of climate and ambient air pollution on the epidemic growth during COVID-19 outbreak in Japan. *Environ. Res.* **190**, 110042 (2020).
44. Moriyama, M. & Ichinohe, T. High ambient temperature dampens adaptive immune responses to Influenza A virus infection. *Proc. Natl Acad. Sci. USA* **116**, 3118–3125 (2019).
45. Chapin, C. V. *The Sources and Modes of Infection* (Wiley, 1912).
46. Keene, C. H. Airborne contagion and air hygiene. William Firth Wells. *J. School Health* **25**, 249–249 (1955).
47. Lednicky, J. A. et al. Viable SARS-CoV-2 in the air of a hospital room with COVID-19 patients. *Int. J. Infect. Dis.* **100**, 476–482 (2020).
48. Somsen, G. A., Rijn, C. V., Kooij, S., Bem, R. A. & Bonn, D. Small droplet aerosols in poorly ventilated spaces and SARS-CoV-2 transmission. *Lancet Resp. Med.* **8**, 658–659 (2020).
49. Morawska, L. & Milton, D. K. It is time to address airborne transmission of coronavirus disease 2019 (COVID-19). *Clin. Infect. Dis.* **71**, 2311–2313 (2020).
50. Tanne, J. H. Covid-19: CDC publishes then withdraws information on aerosol transmission. *BMJ* (24 September 2020) https://doi.org/10.1136/bmj.m3739
51. Zhang, R., Li, Y., Zhang, A. L., Wang, Y. & Molina, M. J. Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proc. Natl Acad. Sci. USA* **117**, 14857–14863 (2020).
52. Jayaweera, M., Perera, H., Gunawardana, B. & Manatunge, J. Transmission of COVID-19 virus by droplets and aerosols: a critical review on the unresolved dichotomy. *Environ. Res.* **188**, 109819 (2020).
53. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284–286 (2020).
54. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B Biol. Sci.* **274**, 599–604 (2007).
55. ECDC Download historical data (to 14 December 2020) on the daily number of new reported COVID-19 cases and deaths worldwide. *ECDC* (14 December 2020); https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide
56. Guidotti, E. & Ardia, D. COVID-19 data hub. *J. Open Source Softw.* **5**, 2376 (2020).
57. *NUTS - GISCO* (Eurostat) https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts#nuts21/
58. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
59. atmos-python: atmospheric sciences utility library. *GitHub* https://github.com/atmos-python/atmos (2021).
60. SEDAC Gridded population of the World, Version 4 (GPWv4): Population Count, Revision 11. *SEDAC* (Columbia University, 2018); https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11
61. KIDSS Kawasaki City Infectious Disease Surveillance System. *KIDSS* (October 2021); https://kidss.city.kawasaki.jp/en/realsurveillance/opendata
62. Fontal, A., Bouma, M.J., San Jose, A., Pascual, M. and Rodo, X. Climatic signatures in the different COVID-19 pandemic waves across both hemispheres https://doi.org/10.24433/CO.5600300.v1 (2021).
63. Fontal, A. *AlFontal/sdcpy* (Zenodo, 2021); https://doi.org/10.5281/zenodo.4949813

Acknowledgements

We acknowledge the support of A. Navarro in the development of the SDC apps. A.F. acknowledges financial support from HELICAL as part of the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement no. 81354. X.R. acknowledges support from the Spanish Ministry of Science and Innovation through the 'Centro de Excelencia Severo Ochoa 2019 2023' program (CEX2018 000806S) and support from the Generalitat de Catalunya through the CERCA program. A.S.J. was supported by a fellowship from 'la Caixa' Foundation, Spain (ID 100010434, fellowship code LCF/BQ/DR19/11740017).

Author contributions

A.F. participated in methodology, software, formal analysis, data curation, visualization and writing, reviewing and editing the manuscript. M.J.B. contributed to writing the original draft, and reviewing and editing the manuscript. A.S.J. contributed to software, formal analysis and data curation. L.L. contributed to software, formal analysis and data curation. M.P. contributed to writing the original draft, and reviewing and editing the manuscript. X.R. contributed to conceptualization, methodology, supervision, software, writing the original draft, and reviewing and editing the manuscript. All authors participated in the discussion and interpretation of results.

Competing interests

The authors declare no competing interests.

Additional information

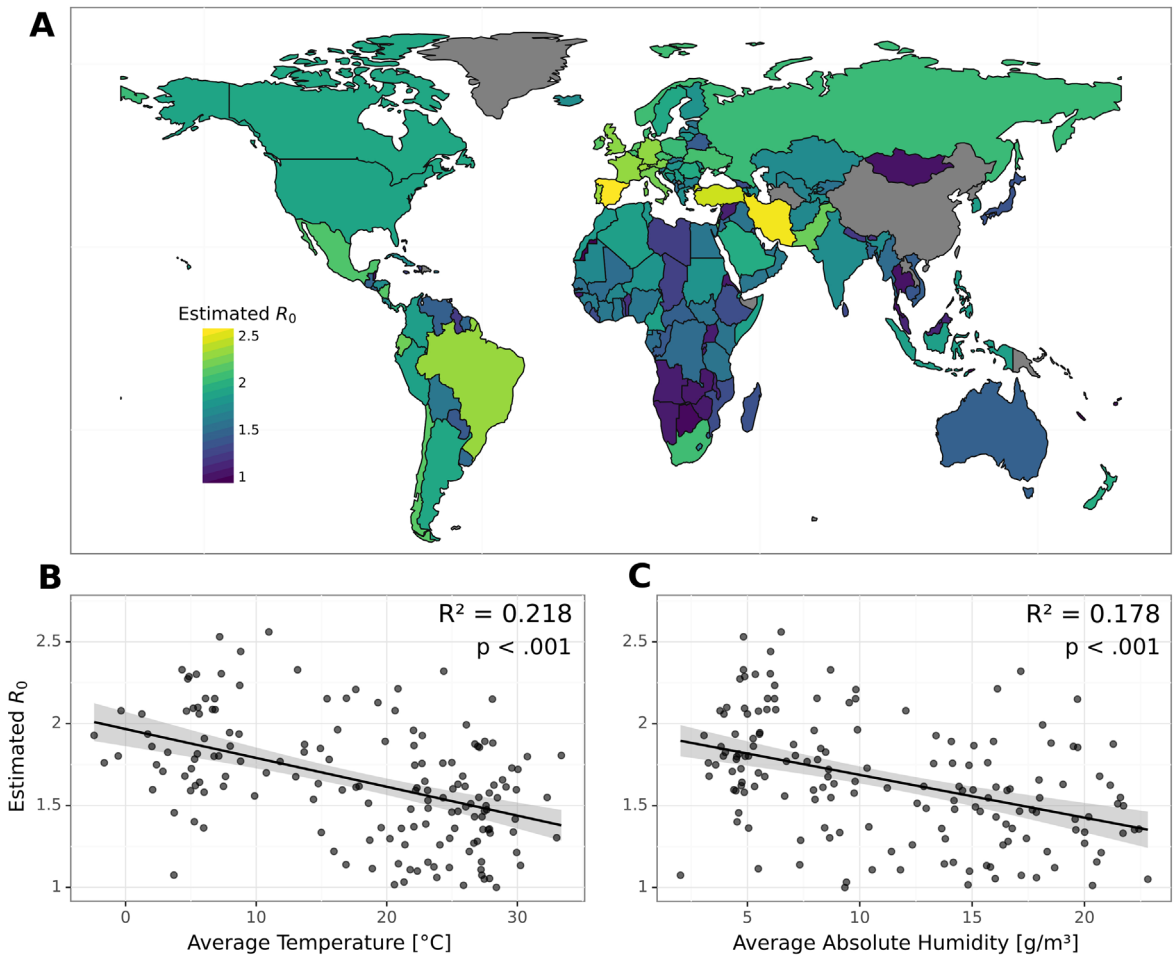
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-021-00136-6.

Correspondence and requests for materials should be addressed to Xavier Rodó.

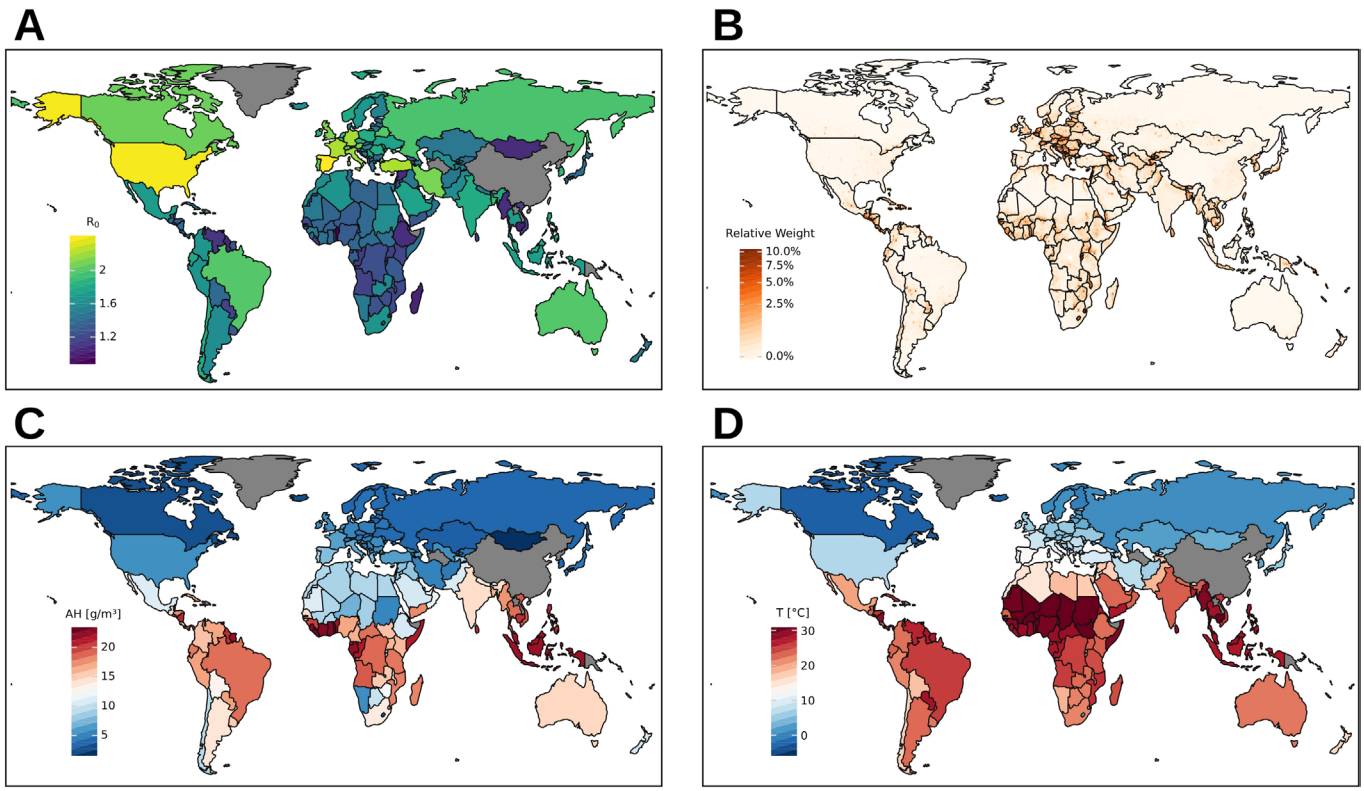
Peer review information *Nature Computational Science* thanks Muhammad Farhan Bashir and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Handling editor: Fernando Chirigati, in collaboration with the *Nature Computational Science* team.

Reprints and permissions information is available at www.nature.com/reprints.

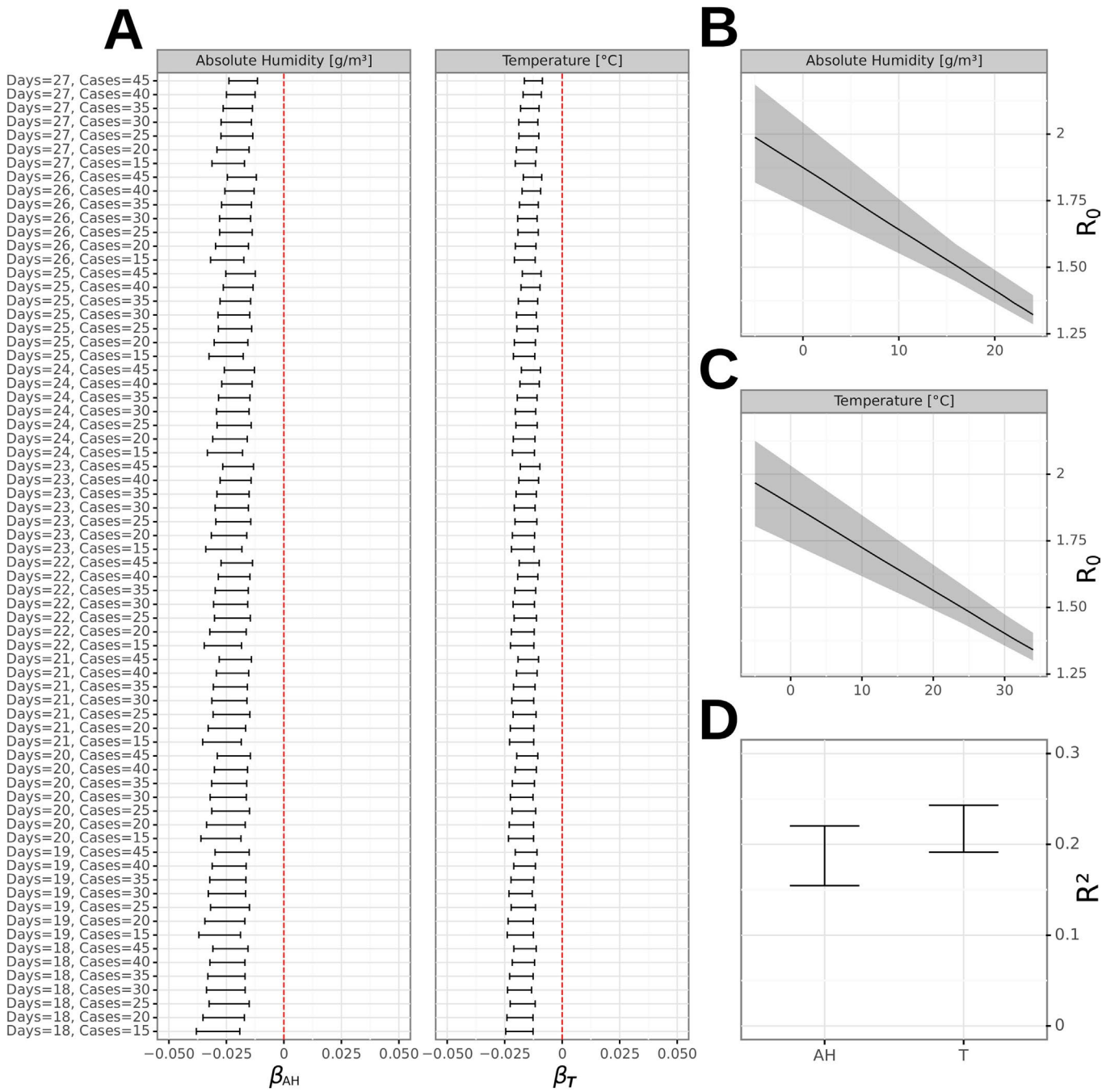
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



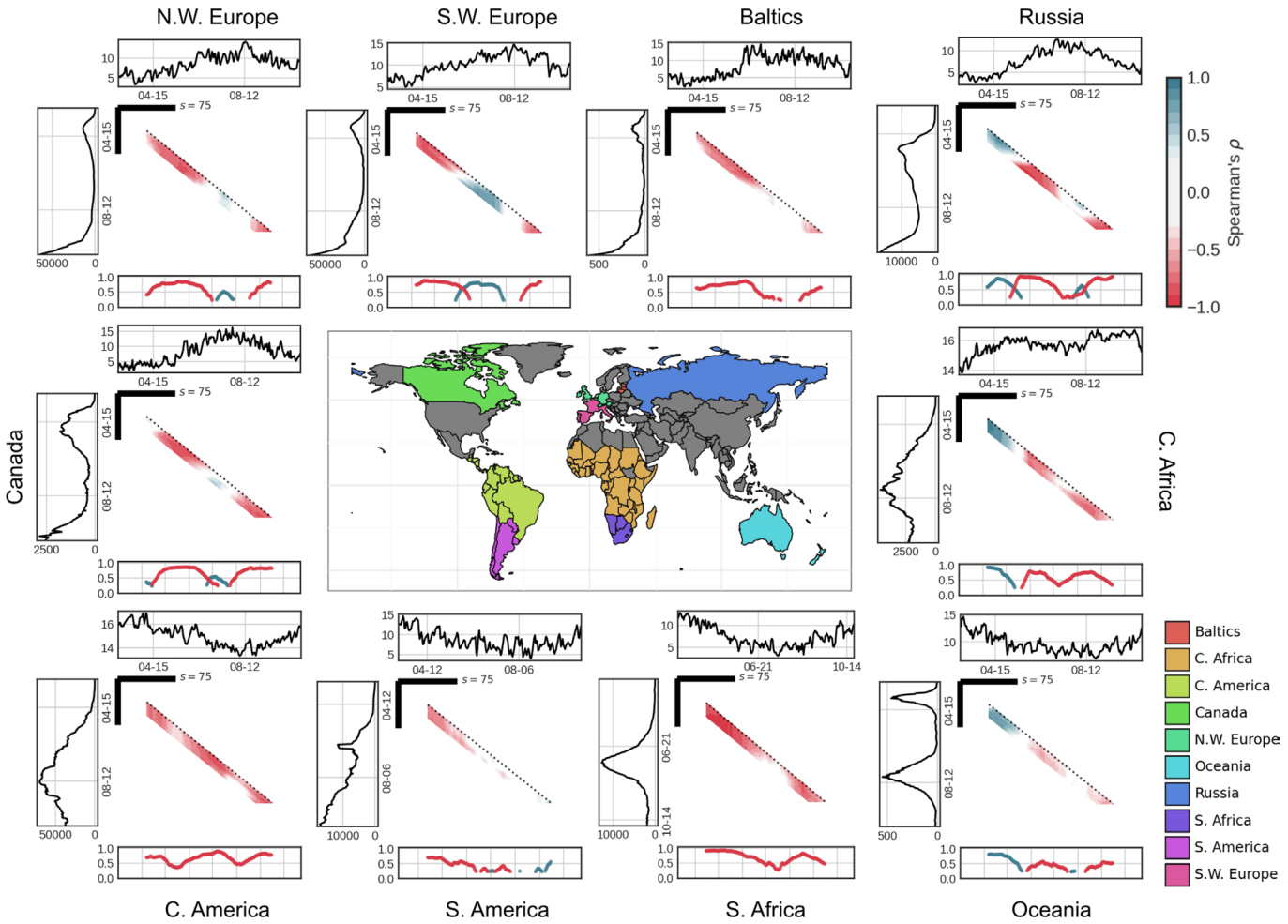
Extended Data Fig. 1 | Comparison of the country-wise estimated R0 during the initial phases of the pandemic in contrast with their meteorological conditions. Estimated basic reproduction number (R0) during the initial phases of the pandemic for all countries included in the study (n=162). (A) To focus on the initial rise of the first wave, these estimates are based on the cumulative COVID-19 cases during the first 20 days following the 20th confirmed case in each country (see Methods). Scatterplots showing the variation of the estimated R0 for the different countries as a function of their average temperatures [°C] (B) and absolute humidities [g/m^3] (C) during the initial phase of the pandemic. Black lines and R^2 values correspond to the estimates of an univariate linear model in both cases, fitted by Ordinary Least Squares (with $p < 0.001$ in both cases in a two-tailed T-test). Of the 162 countries included globally, the numbers corresponding to each continent are the following: Africa (n=50), Asia & Oceania (n=45), Europe (n=38), North America (n=16), and South America (n=12). For comparison, the R0 estimates obtained after the first 100 reported cases (rather than 20) are shown in Extended Data Fig. 2(A). Similar spatial patterns are observed. The spatial distribution of average temperatures and absolute humidity, as well as the national weights based on relative population density, used to compute these averages, are shown in Extended Data Fig. 2(B-D).



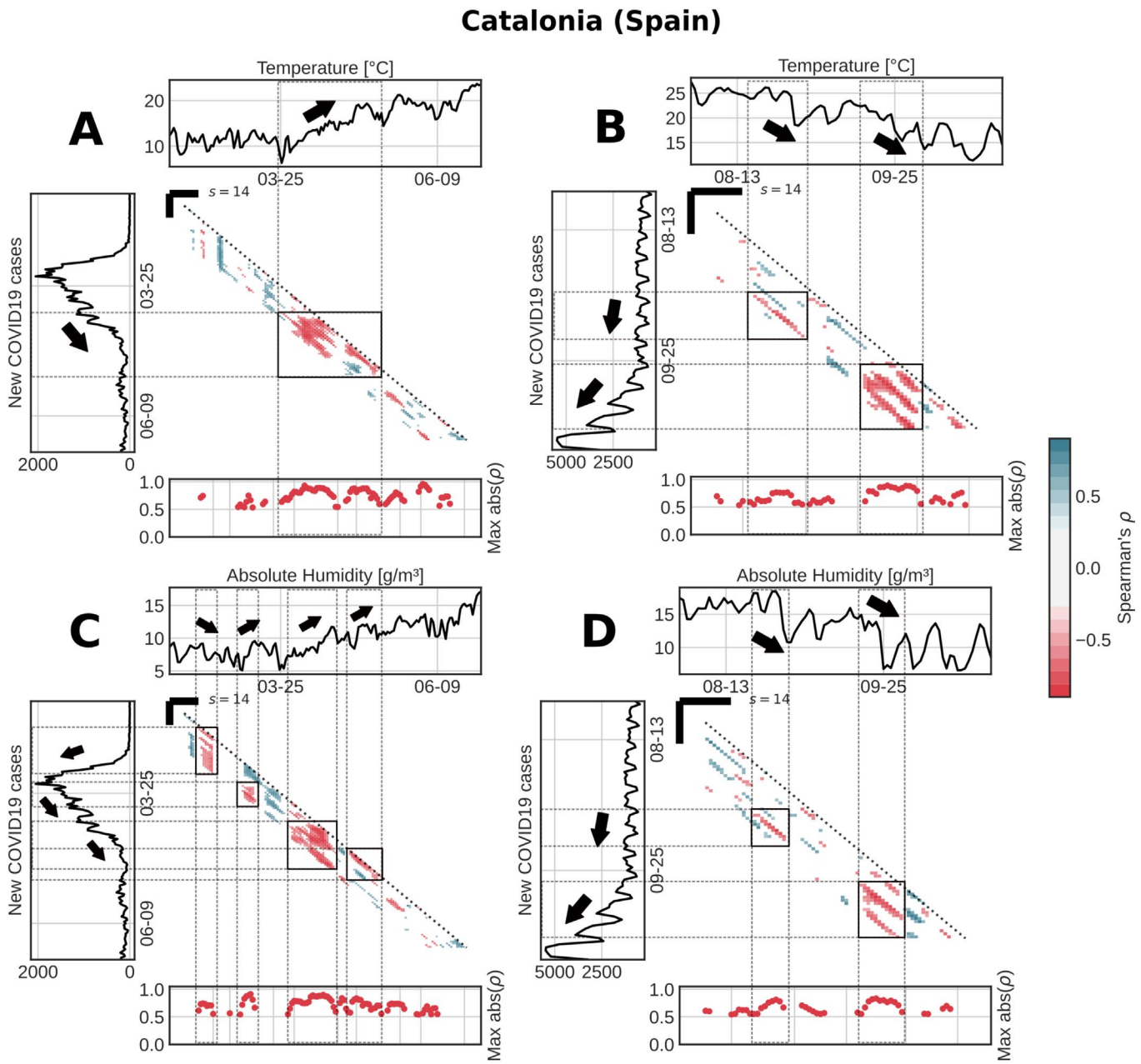
Extended Data Fig. 2 | Worldwide maps for estimated initial R0, weather variables, and population density. (A) Geographical distribution for R0 when it is estimated using the period of 20 days after the 100th COVID19 case (for comparison with Extended Data Fig. 1). (B) Geographical distribution of relative weights given to each cell in a 1×1° grid. Weights represent the fraction of the population living in each cell for every country. Maps in C and D correspond to the averages of AH and T +/-10 days before/after the notification day of the 20th COVID19 case in each country, respectively.



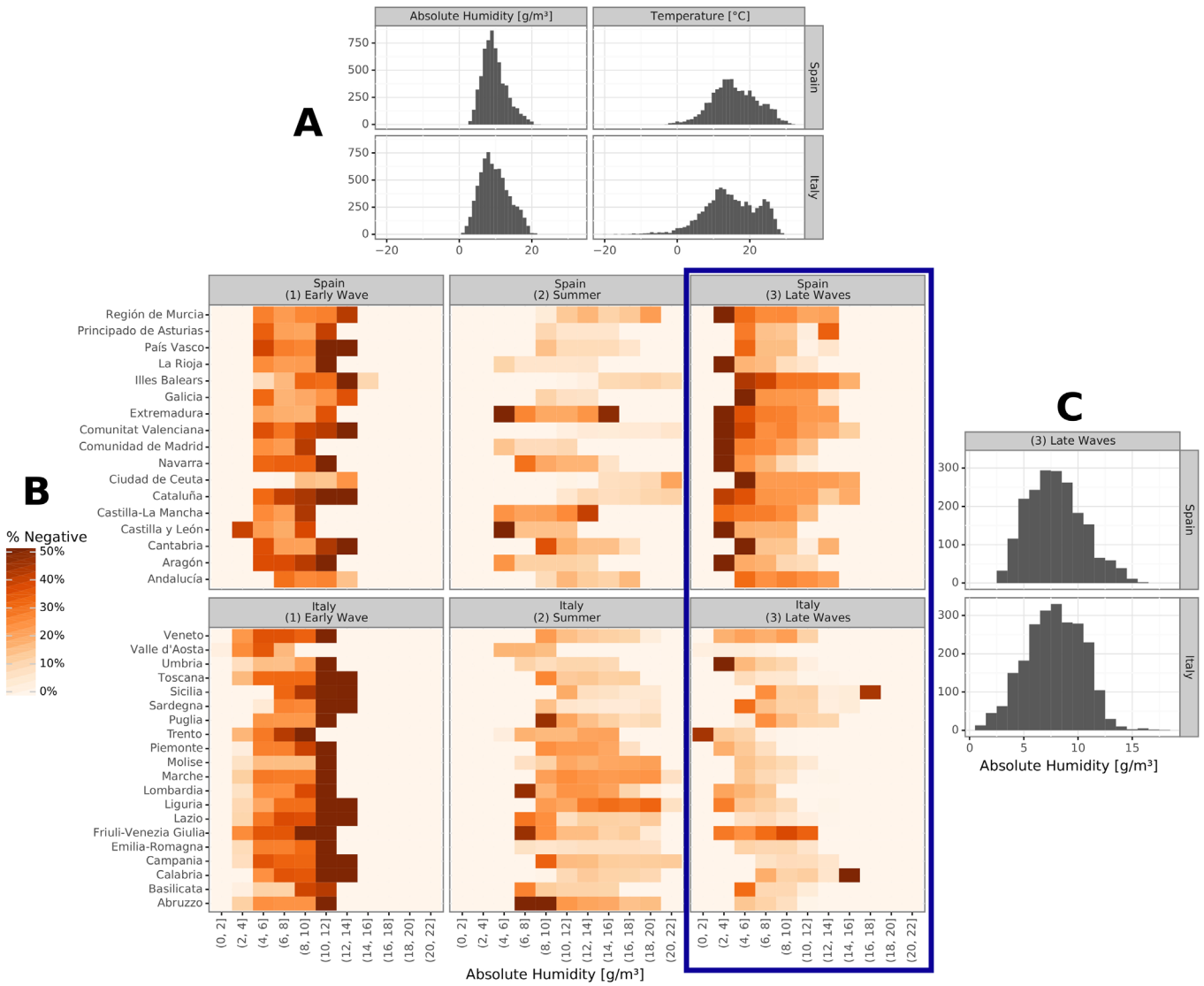
Extended Data Fig. 3 | Regression sensitivity analysis. Sensitivity tests for the OLS regression models shown in Extended Data Fig. 1. (A) 95% CI of the estimated slopes of the OLS regressions (β_{AH} and β_T) for different values of the number of days (18 to 27) and initial number of cases (15 to 45) used for estimating the R0 (see Methods). B and C show the distribution of all the regression lines obtained when considering the variation in days and cases shown in A, for AH and T, respectively. Solid line is the median regression line and the shaded interval corresponds to the area occupied by the most extreme regression lines among the 70 sets of parameters tested. (D) Full range of R^2 estimates obtained as a result of the sensitivity analysis. Results support the robustness and stability of results in Extended Data Fig. 1.



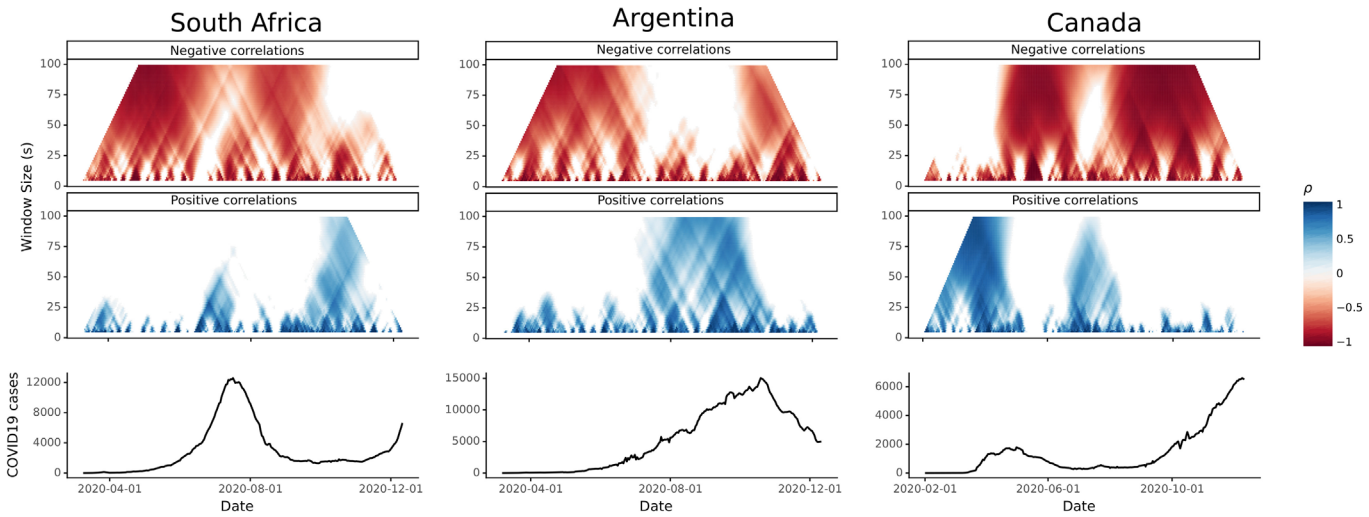
Extended Data Fig. 4 | SDC analyses for the aggregated COVID-19 cases and absolute humidity for groups of contiguous countries. Idem as Figure 1 but for Absolute Humidity (g/m³).



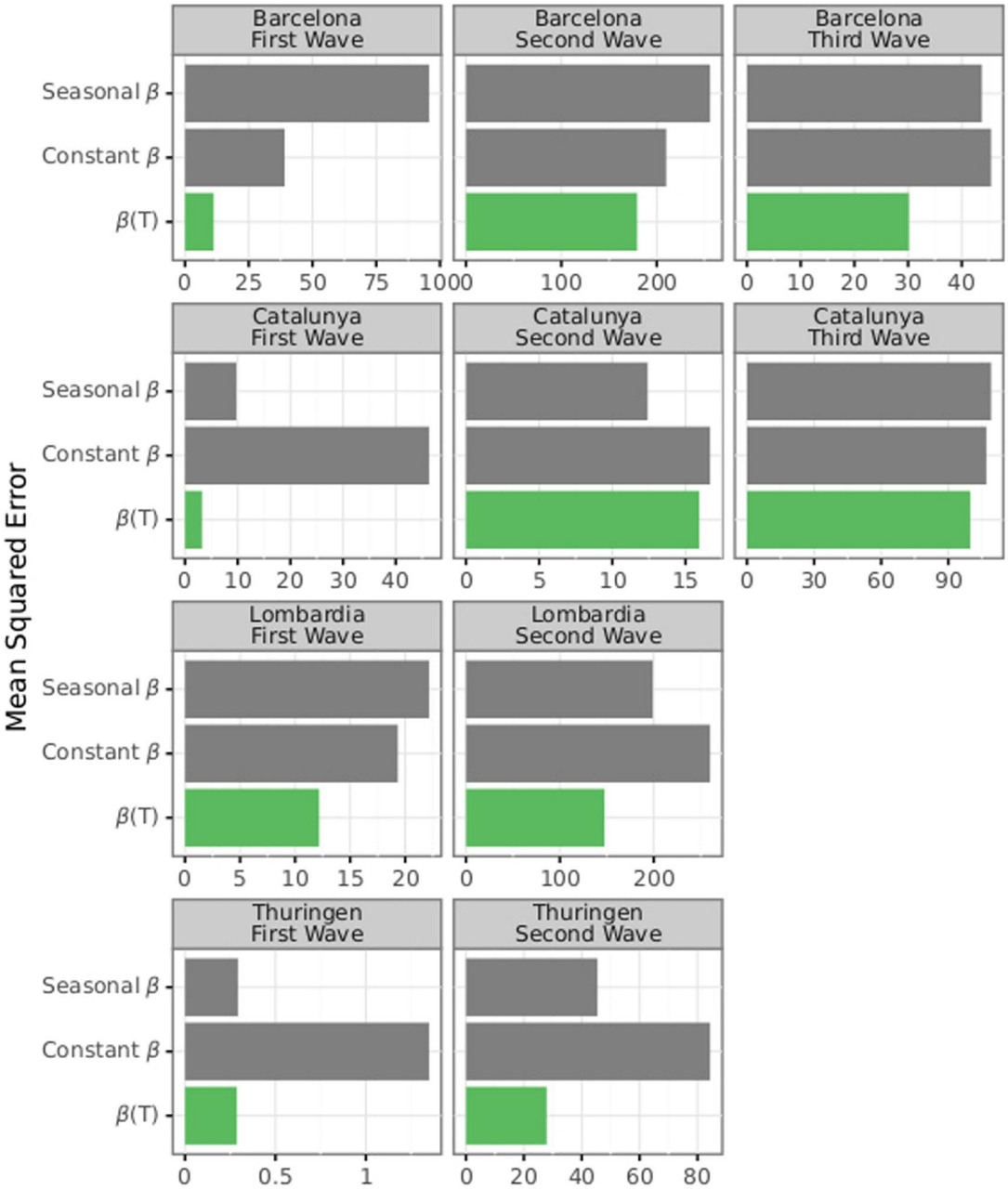
Extended Data Fig. 5 | Scale Dependent Correlation analyses of T/AH with COVID-19 cases during the initial wave and summer in Catalonia, Spain. Idem as Fig. 3 (A-D) but for Catalonia (Spain).



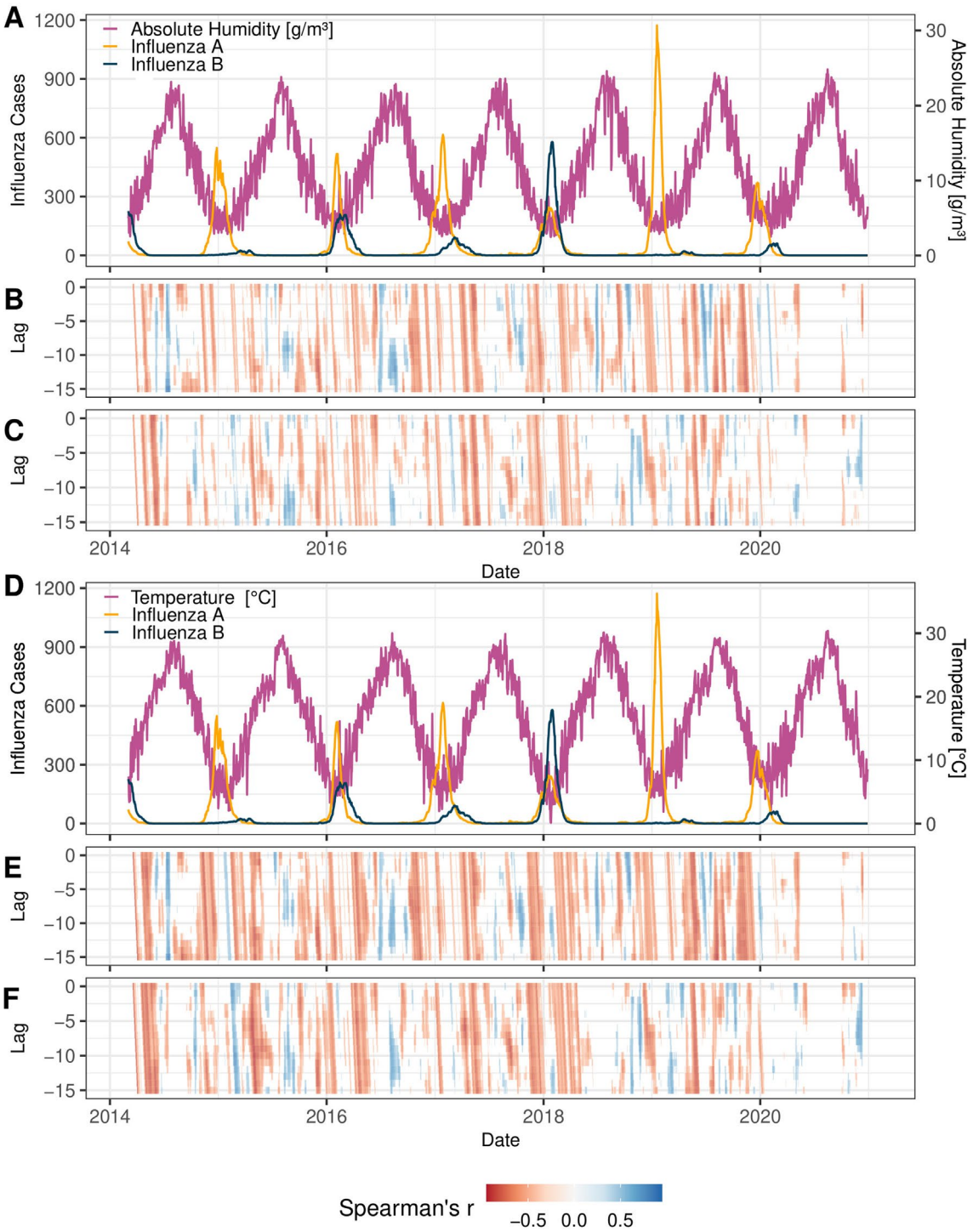
Extended Data Fig. 6 | Distribution of weather conditions and its effect on COVID-19 spread across waves in Spain and Italy. (A) Distribution of daily absolute humidity and temperature averages for Italy and Spain during the whole considered period (15 February 2020 to 31 January 2021). (B) Comparison of the frequency of significant negative local correlations for different ranges of absolute humidity found in SDC analyses (s=21) for Italian and Spanish regions across the three defined periods of the COVID19 pandemic (see methods). (C) Distribution of daily absolute humidity averages in the same regions during the late waves of the pandemic (October 2020 to January 2021).



Extended Data Fig. 7 | Multiscale SDC analysis of COVID-19 and temperature in South Africa, Argentina and Canada. Representation of the results of running multiple two-way SDCs at scales ranging from 5 to 100 days for the average daily temperatures against the number of reported daily new cases of COVID19 in South Africa, Argentina in Canada, spanning from February to December 2020. The top panels (red) represent the minimal significant Spearman correlation found for every pair of window size and date, while the central panels (blue) represent the maximal significant correlation for the same pair of variables. The bottom panels display the 7-days moving average of daily reported new cases for each of the regions.



Extended Data Fig. 8 | Mathematical model's goodness of fit. Mean squared error of each of the models, regions and periods in the study fitted with the process-based model.



Extended Data Fig. 9 | Comparison of historical Influenza incidence and weather conditions in Kawasaki City. A and D show the daily averages for absolute humidity and temperature in the Kanagawa prefecture against the daily reported cases of Influenza A and B in Kawasaki City. Below, the results of an SDC analysis performed at s=30 for absolute humidity against Influenza A (B) and Influenza B (C), and temperature against Influenza A (E) and Influenza B (F).

PAPER IV

Microbial richness and air chemistry in aerosols above the PBL confirm 2,000-km long-distance transport of potential human pathogens

Xavier Rodó[a,b,1,2] , Sofya Pozdniakova[b,1] , Sílvia Borràs[b] , Atsushi Matsuki[c] , Hiroshi Tanimoto[d] , Maria- Pilar Armengol[e], Irina Pey[e], Jordi Vila[f,g], Laura Muñoz[f,g] Samuel Santamaria[b] , Lidia Cañas[b] , Josep-Anton Morguí[b], **Alejandro Fontal**[b,h,3] , and Roger Curcoll[i,3]

[1] X.R. and S.P. contributed equally to this work.

[2] To whom correspondence may be addressed.

[3] A.F. and R.C. contributed equally to this work.

[a] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain;

[b] Climate and Health Group, Climate, Air pollution, Nature and Urban Health, Barcelona Institute for Global Health, Barcelona 08003, Spain;

[c] Kanazawa University, Kanazawa 920- 1164, Japan;

[d] Earth System Division, NIES, Tsukuba 305- 8506, Japan;

[e] Translational Genomics Facility, Fundació Institut de Recerca en Ciències de la Salut Germans Trias i Pujol, Badalona 08916, Spain;

[f] Department of Clinical Microbiology, Biomedical Diagnostic Center, Hospital Clinic School of Medicine, University of Barcelona, Barcelona 08036, Spain;

[g] CIBER, Instituto de Salud Carlos III, Madrid 28029, Spain;

[h] Department of Microbiology, Genetics and Statistics, Faculty of Biology, University of Barcelona, Barcelona 08028, Spain;

[i] Ionising Radiation, Health and Environment, Institute of Energy Technologies, Universitat Politècnica de Catalunya, Barcelona 08028, Spain

Published peer-reviewed article:

Proceedings of the National Academy of Sciences, 121, 38, (2024).

IF: 9.4

DOI: 10.1073/pnas.2404191121

PNAS RESEARCH ARTICLE | ECOLOGY
ENVIRONMENTAL SCIENCES

Microbial richness and air chemistry in aerosols above the PBL confirm 2,000-km long-distance transport of potential human pathogens

Xavier Rodó[a,b,1,2], Sofya Pozdniakova[b,1], Sílvia Borràs[b], Atsushi Matsuki[c], Hiroshi Tanimoto[d], Maria-Pilar Armengol[e], Irina Pey[e], Jordi Vila[f,g], Laura Muñoz[f,g], Samuel Santamaria[b], Lidia Cañas[b], Josep-Anton Morgui[b], Alejandro Fontal[b,h,3], and Roger Curcoll[i,3]

Affiliations are included on p. 11.

Edited by Nils Stenseth, Universitetet i Oslo, Oslo, Norway; received March 5, 2024; accepted July 9, 2024

The existence of viable human pathogens in bioaerosols which can cause infection or affect human health has been the subject of little research. In this study, data provided by 10 tropospheric aircraft surveys over Japan in 2014 confirm the existence of a vast diversity of microbial species up to 3,000 m height, which can be dispersed above the planetary boundary layer over distances of up to 2,000 km, thanks to strong winds from an area covered with massive cereal croplands in Northeast (NE) Asia. Microbes attached to aerosols reveal the presence of diverse bacterial and fungal taxa, including potential human pathogens, originating from sewage, pesticides, or fertilizers. Over 266 different fungal and 305 bacterial genera appeared in the 10 aircraft transects. Actinobacteria, Bacillota, Proteobacteria, and Bacteroidetes phyla dominated the bacteria composition and, for fungi, Ascomycota prevailed over Basidiomycota. Among the pathogenic species identified, human pathogens include bacteria such as *Escherichia coli, Serratia marcescens, Prevotella melaninogenica, Staphylococcus epidermidis, Staphylococcus haemolyticus, Staphylococcus saprophyticus, Cutibacterium acnes, Clostridium difficile, Clostridium botulinum, Stenotrophomonas maltophilia, Shigella sonnei, Haemophillus parainfluenzae* and *Acinetobacter baumannii* and health-relevant fungi such as *Malassezia restricta*, *Malassezia globosa*, *Candida parapsilosis* and *Candida zeylanoides, Sarocladium kiliense, Cladosporium halotolerans,* and *Cladosporium herbarum*. Diversity estimates were similar at heights and surface when entrainment of air from high altitudes occurred. Natural antimicrobial-resistant bacteria (ARB) cultured from air samples were found indicating long-distance spread of ARB and microbial viability. This would represent a novel way to disperse both viable human pathogens and resistance genes among distant geographical regions.

microbes | aerosols | long-distance transport | ARG | pathogens

Significance

Atmospheric aircraft monitoring with 10 tropospheric flights over the planetary boundary layer in Japan (between 1,000 m and 3,000 m above sea-level) demonstrate the presence of viable bacteria and fungi harmful to humans. Long-distance transport for over 2,000 km is possible in the free troposphere for air masses originating in agricultural regions enriched in fertilizers and pesticides.

It has been established that living organisms rarely reach high altitudes and survive for long in the troposphere (1–3). However, there is limited knowledge regarding the microbial species richness and diversity in this environment. Low levels of moisture and nutrients combined with high ultraviolet (UV) are known to restrict microbial proliferation in mid-to-high altitudes in the troposphere. It is noteworthy that certain extremophiles, like *Deinococcus radiodurans* can survive in the harsh conditions found at high altitudes with high radiation levels (4). The opposite is true for surface air, as demonstrated by several studies, which possesses a rich microbial biodiversity. The reported estimates for bacteria (e.g., annual mean of 2.5×10^4– 5×10^4 cells m^{-3}) and fungi (e.g., annual mean of 2.5×10^4–1×10^5 cells m^{-3}) arise from terrestrial biomes (5) or during dust events (6) or intense rains (7). Overall, 33 to 68% of these microorganisms can be transported, at times, over short distances of a few km (8) but quantitative studies are lacking (9). Whether or not these organisms stay viable and may then trigger the initiation of human pathology remains, however, an open and elusive question.

Limited data exist to date on the magnitude of microorganism propagation in the free troposphere (10), above the planetary boundary layer (PBL), whether attached to soil dust (11) or to organic aggregates (12). For a review, see ref. 13. The atmospheric PBL delimits the area of the lower troposphere most influenced by surface processes (1). The height of the PBL varies as a function of atmospheric conditions, with vertical mixing occurring below this layer. In contrast, above the PBL, materials uplifted can be subject to long-range transport due to the reduction in friction from both the surface and air density and reach the ground again through dry deposition in subsiding air masses or by being washed out by rain (14). Yet, the implications for human and plant health of this transport, as well

Author contributions: X.R. and J.-A.M. designed research; X.R., S.P., S.B., A.M., H.T., M.-P.A., I.P., J.V., J.-A.M., A.F., and R.C. performed research; X.R., S.P., S.B., A.M., H.T., M.-P.A., I.P., J.V., L.M., S.S., L.C., A.F., and R.C. contributed new reagents/analytic tools; X.R., S.P., S.B., M.-P.A., I.P., J.V., S.S., A.F., and R.C. analyzed data; and X.R., S.P., J.V., and A.F. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]X.R. and S.P. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: xavier.rodo@isglobal.org.

[3]A.F. and R.C. contributed equally to this work.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2404191121/-/DCSupplemental.

Published September 9, 2024.

as the viability of microbes in bioaerosols that can later cause infection have not been adequately explored.

Although little investigated, long-distance transport of viable airborne fungi and bacteria with soil dust has been shown to occur [e.g., from Africa to the Caribbean (15–17)] thereby causing adverse effects on the survival of corals and plants (18, 19). A consensus has been established that dust particles may pose serious risks to the environment and human health in countries in dust source regions and surrounding areas (20), though, for example, pathogens such as nontuberculosis *Mycobacterium* (8). Furthermore, tularemia outbreaks in Spain and Iran were linked, among other causes, to the inhalation of contaminated aerosols, and dust storms have been connected to regional outbreaks of meningococcal meningitis caused by *Neisseria meningitidis*, although the exact mechanism is still unknown (21). Cardiovascular, respiratory, and lung diseases can be caused by the inhalation of submicron radius particles because these can be ingested deep into the human body and are associated with respiratory morbidity and mortality (22). Cases of eye infections and diseases such as meningitis and valley fever have been recorded during and after significant dust events in some regions (23). Despite its significant relevance to human health, a mechanistic understanding of their airborne epidemiology and links to bioaerosols still remains elusive, therefore making a definitive attribution impossible.

Bacterial resistance to antimicrobial agents is becoming increasingly common and poses a serious threat to health security, as animal pathogens are among the main reservoirs of various drug resistance genes later transmitted to humans via the food chain. Antimicrobial resistance genes can occur through horizontal gene transfer (HGT) in any environment, particularly where bacterial loads are high, for example, in soil, in wastewater treatment plants (24), and in the gut microbiome of humans and animals, according to the transfer-related genes carried on plasmids (25). In addition, the natural environment and its microorganisms provide a pool of natural drug resistance genes that, when influenced by human activities, climate change, and environmental disruption, may all affect the evolution of bacteria and give rise to new genes that promote drug resistance (24). Antibiotics have also become one of the most frequently detected new pollutants in the environment, with antimicrobial resistance spreading from livestock and contaminated meat products to people.

In the present study, we sought to confirm that air masses above the PBL contain and disperse both fully viable human pathogens and bacterial species promoting drug resistance thousands of km away from their sources and that, despite their low concentrations, they may still have the potential to affect human health. Next, we briefly describe the airborne sampling protocol, the chemical analysis, DNA extraction and sequencing approaches followed. Afterward, how the cultures of samples were performed and the resistance profiles obtained. After the overall data analysis section, we end with a summarizing discussion on the conceptual progress obtained.

Methods and Analysis

Aircraft Surveys and Aerosol Sampling. Ten air surveys with a Cessna 172 aircraft were conducted over Japan in February (22nd-26th and 28th) and April (1st and 5th-8th) 2014, departing from Chofu airport (Fig. 1 and *SI Appendix*, Fig. S1 and Table S1). These days were selected for their optimal conditions in terms of flight paths because they were against the NW direction of prevailing winds over Japan (Fig. 2 *A–C* and *SI Appendix*, Figs. S1–S3). The collection of aerosol samples in the lower troposphere was automated by opening an internal manifold when the plane was above the PBL (over 1,000 m a.s.l. and ascending to around 3,000 m a.s.l.; Fig. 1 and *SI Appendix*, Fig. S1). On the same day, after each flight, aerosol samples were also collected at the surface level at the Chofu airport near Tokyo, following the same experimental workflow (Fig. 1). Air sampling was carried out

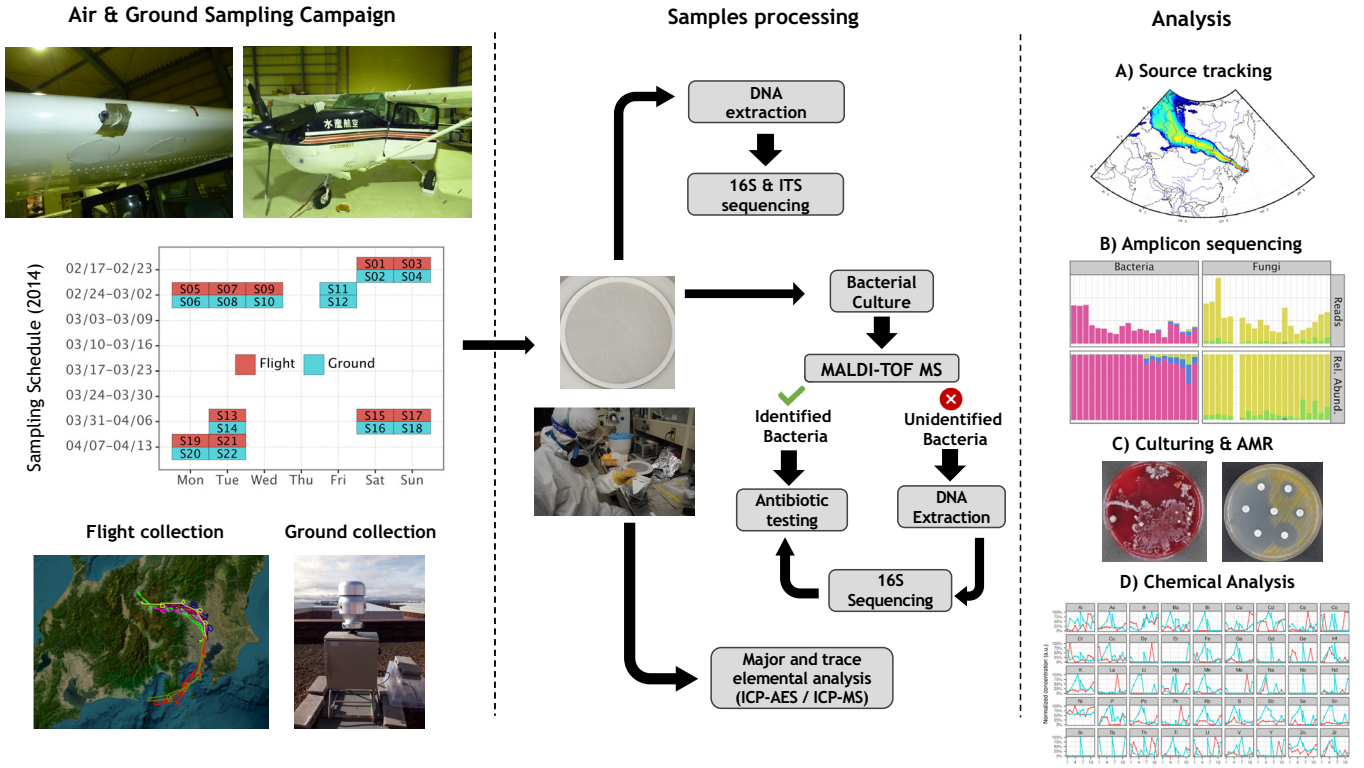


Fig. 1. Sampling and analysis pipeline for collected air samples. Flight and ground campaigns were conducted in parallel according to the established sampling schedule in parallel. The flight trajectory was documented. One-eighth of each collected filter was used for DNA extraction, bacterial culture, and major and trace elemental analysis. Comprehensive chemical and biological analysis together with source tracking of air masses was performed.

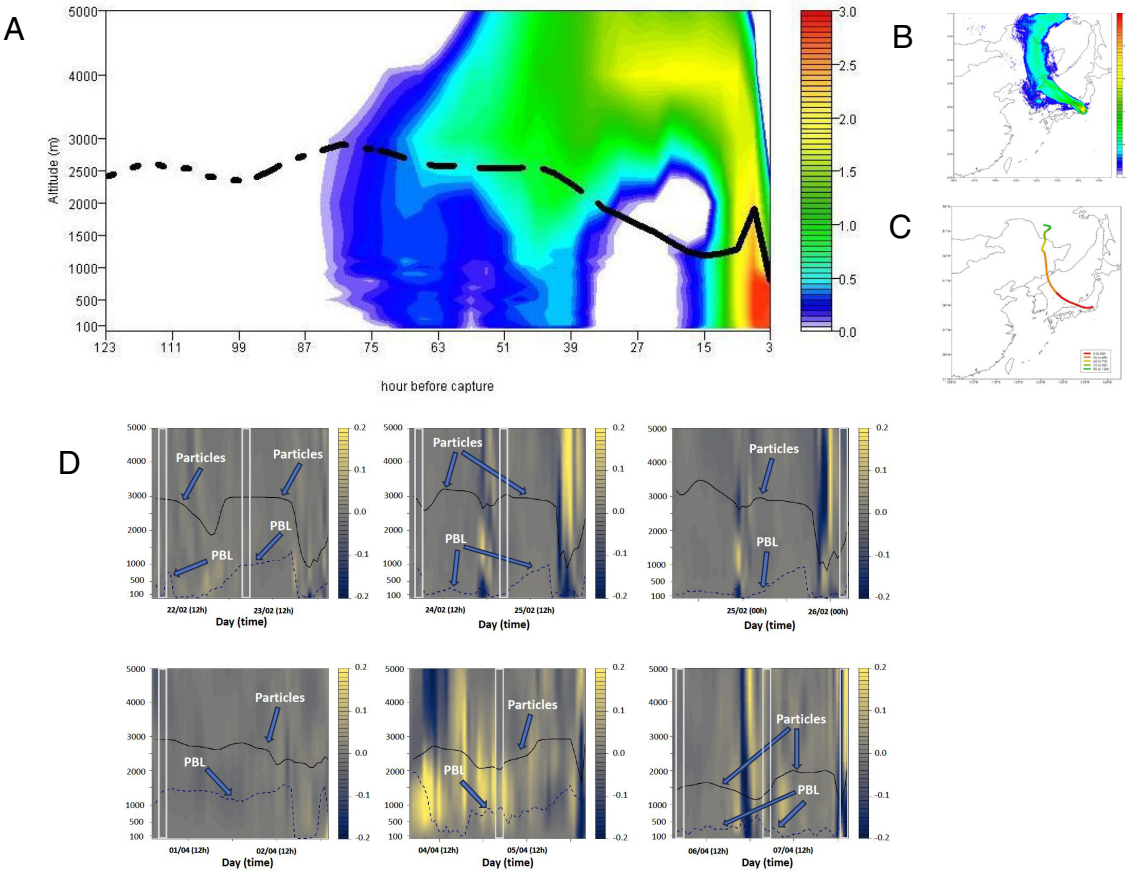


Fig. 2. Time-height distribution of particles and the PBL over the days of sampling. (*A*) Average FLEXPART time-altitude 120-h backward simulation for the composite of all 10 flights in 2014 (The *X* axis denotes hours since 10.000 particles—log10 scale—were thrown at 3 h steps from Tokyo; colors denote concentration of particles in the total column from the surface up to 5 km; see *Methods and Analysis*). (*B*) Same as (*A*) but showing the average lat-lon source dispersion plume simulated from Tokyo. (*C*) Path covered by the centroid of air mass particles at each time step. Colors denote days. (*D*) Heights of both the centroid of particles (solid line) and the PBL (dashed line) during the days of sampling. White boxes indicate the times of flights and the color intensity scale, vertical wind speed. Note the downward direction of winds and high speeds lowering the height of the PBL.

as described by Rodó et al. (14). A total of 22 aerosol samples were obtained during the two sampling periods, as detailed in *SI Appendix*, Table S1. In addition, for each sample period each day, both in flight and on the ground, a field blank filter was collected, exposed, and handled in an identical manner except that the sampler was not in operation.

Air mass trajectories were simulated 5 d back in time for each aircraft sampling date using the Lagrangian particle transport and dispersion model FLEXible PARTicle (FLEXPART v10.4) (26). Light detection and ranging (LIDAR) measurements for the same days and times were obtained from the Asian Dust and Aerosol Lidar Observation Network (https://www-lidar.nies.go.jp/AD-Net/).

Aerosol filter samples ($PM_{2.5}$) were analyzed for the determination of major ($\mu g/m^3$) and trace elements (ng/m^3) by Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) and Mass Spectrometry (ICP-MS), respectively, as described in detail elsewhere (27). One eighth of the filters (approx. 22 cm^2) was used for DNA extraction. 16S rRNA gene amplicons were obtained following the 16S rRNA gene Metagenomic Sequencing Library Preparation Illumina protocol (Cod. 15044223 Rev. A). Raw FASTQ data were filtered to facilitate high-quality reads. To account for potential contamination, the processed high-quality reads were compared to a negative control included in the sequencing run.

Considering the heterogeneity of microorganisms collected in the filters, different culturing methodologies were applied and prior to the MALDI-TOF MS analysis, individual colonies were incubated for 24 h at 37 °C. If the strain was not identified by MALDI-TOF MS, 16S rRNA sequencing was performed. The extraction of genetic material from pure bacterial cultures was carried out and DNA quantification performed. The 16S rRNA gene amplification and detection in real time was performed. Antimicrobial susceptibility was assessed with Sensititre™ Custom Antimicrobial Susceptibility Testing Plates. All the isolates were categorized as multidrug-resistant (MDR), extensively drug-resistant, or pandrug-resistant (PDR; see definitions in *SI Appendix*). All methodologies are fully detailed in the *Methods* section of *SI Appendix*.

Results

Analysis of Air Mass Trajectories. All flights were planned to track the wind currents coming from the mainland Asian continent (Fig. 2 and *SI Appendix*, Figs. S1 and S2). According to the backtrajectory analysis, over all the different aircraft flights in 2014 (*SI Appendix*, Figs. S2 and S3), the air took 2 d on average to travel from NE China to Tokyo (Fig. 2). This meant that the air traveled at a speed of around 12 m/s above the PBL (at a height between 1,500 m and 3,000 m a.s.l.) when over the Sea of Japan. This northwest (NW) atmospheric bridge (Fig. 2 *A*–*C*) did not allow the entrainment of new air from the maritime surface at any time during the transect collection (Fig. 2*A*, and PBL time-height distribution in Fig. 2*D* and *SI Appendix*, Fig. S4*B*). Afterward, the

air was normally channeled down again across the PBL over central Japan, when strong downward winds push all the air mass from up to 5,000 m a.s.l. down to the surface (Fig. 2 and *SI Appendix*, Fig. S4). This movement down the air column is facilitated by the low-pressure system that establishes itself over northern Japan in winter. Before reaching this system, the air is uplifted by the strong Siberian High, the semipermanent cell of high pressure centered poleward of 45°N in NE Siberia (28). The centroids of the mass of particles (solid lines in Fig. 2*D*) in the backward simulations are located at heights from 1,000 m to nearly 3,000 m and always well above the height of the PBL (e.g., Fig. 2*D*; dotted lines). As seen by the white boxes in the simulations (Fig. 2*D*) of the flight durations each day, the sampling took place always collecting air above the PBL, without any contamination from air entrainment from the surface below the PBL (Fig. 2*D*).

LIDAR records for the same two groups of days in February and April 2014 confirm a massive downward flux of particles from very high altitudes, as marked by the temporal evolution in the downward movement of particles (*SI Appendix*, Fig. S4*A*). Total homogenization of the air column up to 6 km may occur on days when strong air intrusions take place, with at times high cloud cover conditions also appearing (red areas in *SI Appendix*, Fig. S4*A*), therefore yielding out-of-range measurements.

Overall Abundance and Diversity of Airborne Bacteria and Fungi. Aerosols collected on quartz filters were analyzed for their chemical and biological composition. The detailed description of the experimental workflow is presented in Fig. 1. The amount of collected biomass was assessed by the quantification of extracted DNA. Although the DNA extracted from the 2 pools of blanks (one containing 10 flight blanks portions and another 11 ground blanks) was undetectable, those blanks were also included and processed in the downstream sequencing analysis, alongside the flight and ground samples. With regard to the total DNA yield, or the rates of unidentified reads per taxonomic levels, there were no significant differences between flight and ground samples (e.g., mean value of 0.1 ng m^{-3}; *SI Appendix*, Fig. S5 *A* and *B* and Tables S1 and S2).

The mean counts of raw sequencing reads were 109 K and 91 K for bacteria and 199 K and 172 K for fungi in flight and ground samples, respectively; from which 88 K and 72 K (for 16S flight and ground) and 141 K and 123 K (for ITS flight and ground) quality-filtered nonchimeric sequences were used for further analysis (Fig. 3*A* and *SI Appendix*, Table S2). The detailed information on the distribution of reads along the quality check protocol is presented in *SI Appendix*, Table S2. In-depth taxonomical analysis resulted in the identification of 98% and 94% of all reads from bacteria and fungi, up to genus level (*SI Appendix*, Fig. S5 *B* and *C*). In total, 18 different bacterial phyla and 2 fungal phyla were identified, distributed over 305 and 266 genera, respectively (*SI Appendix*, Fig. S5 *B* and *C*). The airborne microbiome associated with $PM_{2.5}$ particles was largely dominated by Proteobacteria and Bacteroidota, while the most abundant fungal populations detected were from Ascomycota (Fig. 3*A*).

The global comparison of the number of sequenced reads from flight and ground samples revealed a very high temporal similarity, confirming the massive homogenization of the air column (Figs. 3*A* and 4*A*). The number of unique genera present in samples followed a similar distribution in both bacteria and fungi (*SI Appendix*, Fig. S6), with the distribution across phyla and genera shown in *SI Appendix*, Fig. S6*B*. These results were stable after a sensitivity analysis. The similarity is clearly evident in the inspection of the most and the least similar samples (*SI Appendix*, Fig. S7), as the closest samples in time belong to the flight-to-surface

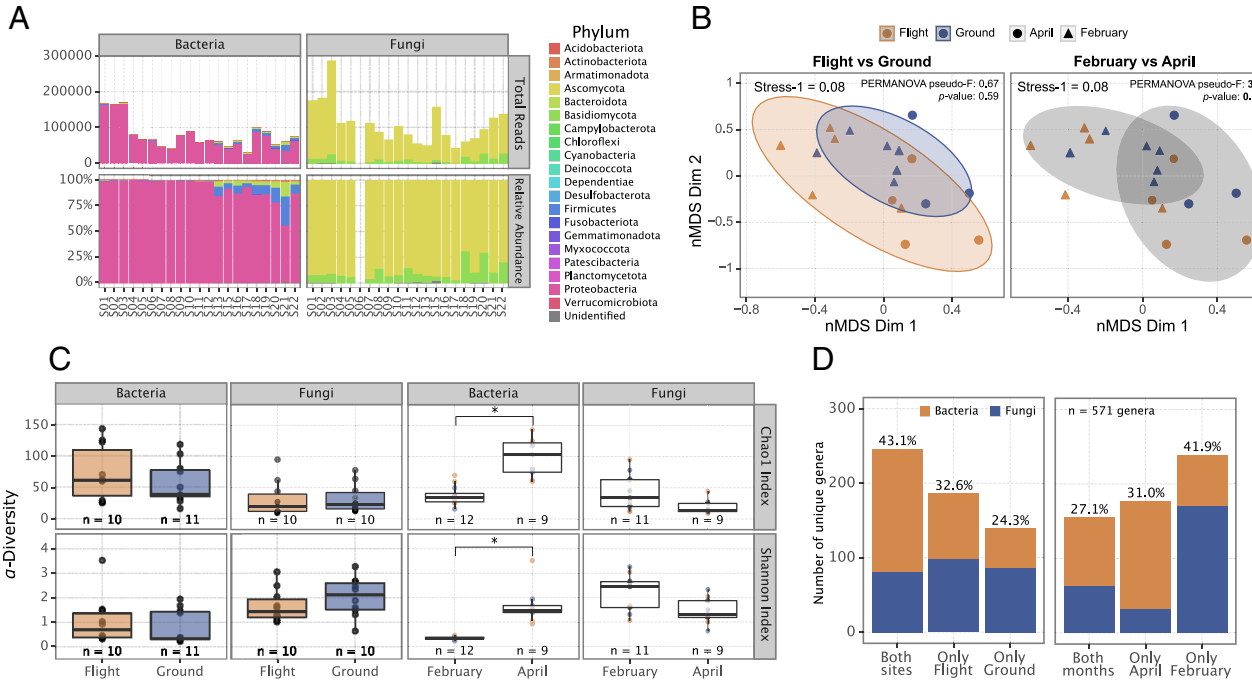


Fig. 3. Microbial diversity and richness above the PBL and in the near-surface atmosphere. (*A*) The taxonomic profile of bacterial and fungal phyla represented as the total number of reads (*Top*) and relative abundance per sample (*Bottom*). (*B*) Two-dimensional nonmetric multidimensional scaling projection depicting the distances between microbial communities at the genus level in various samples. The scatter plot differentiates sample types with color coding (blue for ground, orange for flight) and collection months with symbols (triangles for February, circles for April). Ellipses on the *Left* panel group flight and ground samples, while those on the *Right* panel differentiate between February and April collections. PERMANOVA test results for group differences are displayed atop each panel. (*C*) Variations in alpha-diversity estimators for bacteria and fungi across altitude (*Left*) and month (*Right*). Significant differences, identified using an FDR-corrected Wilcoxon Rank-Sum test, are indicated with brackets (*denotes $P < 0.05$). (*D*) Number of unique bacterial and fungal genera exclusive to flight or ground samples, or found on both (*Left*), and unique to February or April samples, or found in both (*Right*).

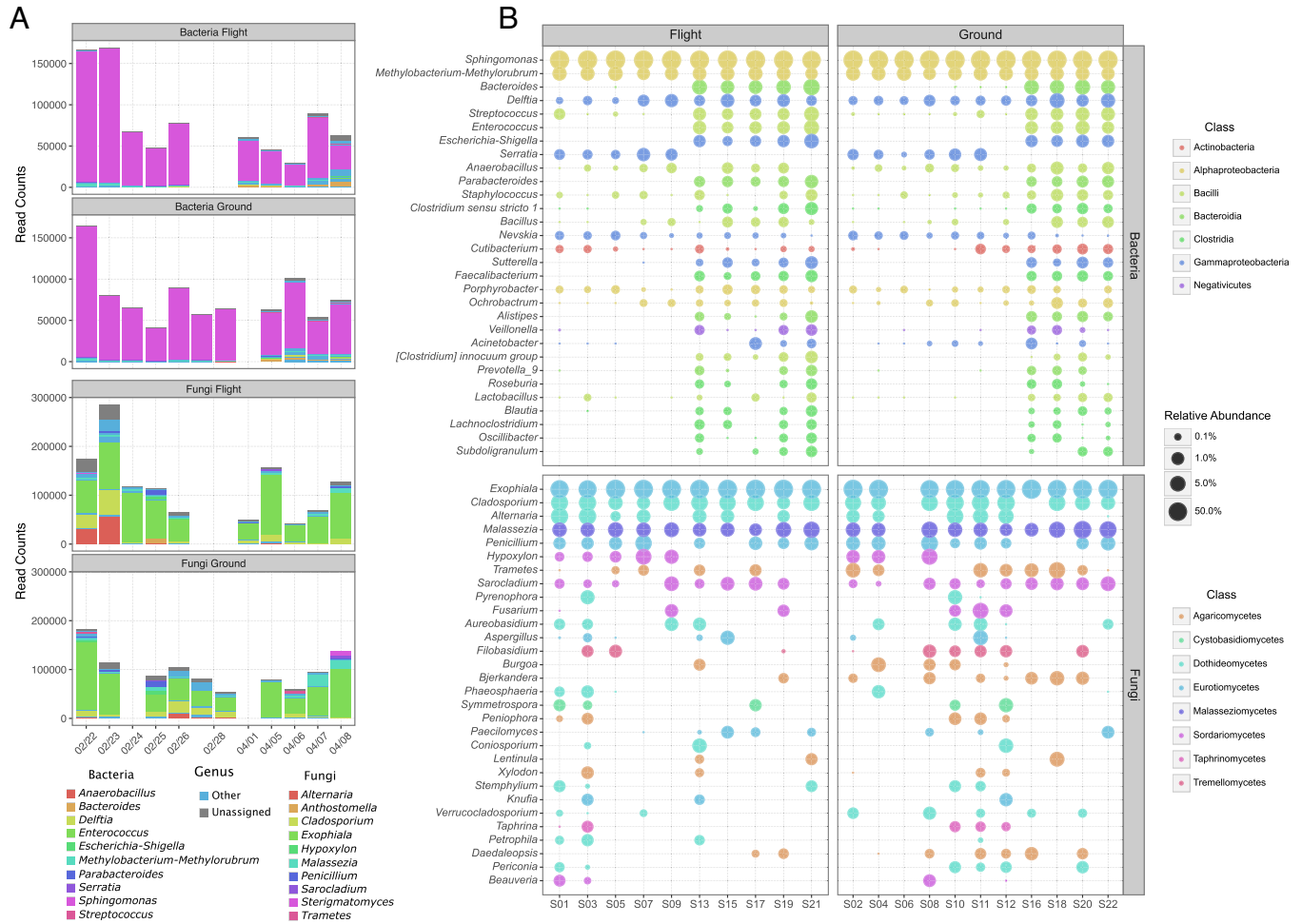


Fig. 4. Composition of air microbiome at high altitude and ground atmosphere. (*A*) Total number of read counts per sample, categorized by collection date. The top 10 most common bacterial and fungal genera are individually color-coded, with the remaining grouped in the "Other" category (blue) and unassigned reads in gray. (*B*) Relative abundances of the top 30 bacterial and fungal genera recovered from flight and ground samples. The size of each dot reflects the relative abundance of the genus in each sample with color indicating its corresponding class.

and the surface-to-surface classes and also have the larger similarity scores (*SI Appendix*, Fig. S8). Applying a permutational multivariate ANOVA (PERMANOVA), no statistically significant differences were discerned between flight and ground samples, however, samples collected in February and April were significantly different (Fig. 3*B*).

Alpha and beta diversity metrics for high-altitude air masses were similar to those of near-ground air, suggesting that high altitude air masses, at specific times, are very diverse and their species richness values are comparable to those of near ground air (Fig. 3 *B* and *C* and *SI Appendix*, Fig. S9). A substantial proportion of taxa (43.1% out of all taxa) appear in both flight and surface samples, while 32.6% out of all detected taxa were only observed in the flight samples, underscoring the distinct origin of those organisms overlapping with the different sources of air masses (Fig. 3*D* and *SI Appendix*, Figs. S2 and S3). Surprisingly, 41.9% of all taxa were only detected in February, due to the higher predominance of fungi during this temporal interval (Fig. 3*D*).

Taxonomic Assignment of Airborne Microbiome. The elevated prevalence of *Sphingomonas* (Fig. 4*A*) in $PM_{2.5}$ dominated other bacterial genera, but upon meticulous examination of the remaining, less-abundant bacterial genera, a rich bacterial diversity unfolded, marked by the dominance of *Methylobacterium*, *Bacteroides*, *Enterococcus*, *Streptococcus*, *Delftia*, *Escherichia, Nevskia,* and *Serratia* (Fig. 4 *A* and *B*). The consistent pattern in relative abundances among the principal 30 bacterial and fungal genera appears in both airborne and surface samples across the entire temporal interval (Fig. 4*B*), once again underscoring the consistent distribution of similar microbial populations in high altitude and ground air masses.

The fungal air microbiome was mainly dominated by *Exophiala,* however, a closer examination of fungal diversity revealed *Cladosporium, Alternaria, Malassezia, Penicillium*, *Hypoxylon*, *Sarocladium,* and *Pyrenophora* as the most abundant, in both altitude and ground samples (Fig. 4*B*).

Temporal shifts of airborne microbes, influenced by nearby sources or by environmental factors in the near surface atmosphere, are well-studied phenomena, but less is known about similar changes in high altitude air. Despite the small sample size, clear changes between samples from February and April were observed in both DNA yield and alpha-diversity (*SI Appendix*, Fig. S5*A* and Table S2). The bacterial diversity observed was higher in April than in February while the converse was shown for fungi (Figs. 3*C* and 4*B*). When comparing the compositions of each domain level group (bacteria and fungi) between the 2-mo in the flight samples, both bacteria and fungi exhibited substantial changes (*SI Appendix*, Fig. S10). Attribution to seasonal changes only was not possible due to the 2-mo span of the sampling. The declining proportion of Proteobacteria in April was accompanied by an increase in Firmicutes and Bacteroidota (Figs. 3*C*

and 4*B*). At the genus level, despite *Sphingomonas* standing as the most prevalent bacteria at high altitude, its relative abundance dropped in April, coinciding with a corresponding reduction in the relative abundance of *Methylobacterium* (Fig. 4*B*). Significant shifts were evident across a multitude of bacterial genera including *Bacteroides, Streptococcus, Enterococcus, Clostridium, Sutterella, Faecalibacterium, Alistipes, Prevotella, Roseburia, Blautia, Lachnoclostridium, Oscillibater,* and *Subdoligranulum,* transitioning from nearly undetectable levels in February to a markedly abundant presence in April. Conversely, *Serratia* displayed a distinct pattern, as it was detectable only in February, while its presence was largely undetectable in April (Fig. 4*B*).

No detectable shifts in fungal phylum distribution between February and April were observed, the fungal community remained predominantly composed of Ascomycota with a notably lower representation of Basidiomycota (Figs. 3*A* and 4*B*). The dominant fungal genera in both months were *Exophiala, Alternaria*, *Cladosporium*, *Malassezia,* and *Penicillium*. In February, highly diverse fungal taxa were also detected, including *Hypoxylon, Phaeosphaeria, Peniophora, Taphrina, Periconia,* and *Beauveria* (Fig. 4*B*).

Aerosol Chemical Composition and Relationships with Airborne Microbial Communities. The average composition of all $PM_{2.5}$ samples revealed S and Na (>1 μg m^{-3}) as the most prominent elements. Less prevalent were Al, K, Fe, and Ca (>100 ng m^{-3}), followed by Mg, Zn, Zr, P, Ni, Cr, Ba, and Pb (>10 ng m^{-3}). A range of other trace elements (Mn, Ti, Cu, Sn, As, Se, V, etc.) were also detected but at even lower concentrations (<10 ng m^{-3}). However, only S, K, Zn, Ni, and Sn were present in all 21 samples, with average relative abundances of 52%, 7.8%, 2.2%, 1.8%, and 0.1%, respectively (*SI Appendix*, Fig. S11).

The mean element content at the ground level (3,821 ng m^{-3}) nearly doubled that at the flight level (1,906 ng m^{-3}), which may indicate additional local intense anthropogenic activity nearby the sampling site (Fig. 5*A*). Despite the detection of elevated concentrations for the majority of elements at ground level, these values were statistically not significantly different from those in the flights. On average though, the mean element concentration in February (3,935 ng m^{-3}) was higher than in April samples (1,780 ng m^{-3}) (Fig. 5*A*). Peaks of S were accompanied by increased levels of Pb, Zn, and Zr, and a range of other potentially toxic elements (Ba, Cr, Cu, Mn, and Ni; *SI Appendix*, Fig. S11).

We used Spearman's rank-order correlation analysis to search for potential relations between the bacterial genera in flights and aerosol's chemistry (the top 40 most correlated genera present in at least two different flight samples are shown in Fig. 5*B*). With the exception of *Deinococcus* and *Delftia*, bacteria appear more strongly correlated to the chemical composition than fungi, indicating a close adherence of the former to the fine particles aerosolized. A majority of the highest correlated taxa appear strongly linked to K first and then, by groups, to Zn, Al, B, and Fe. Noticeably, *Staphylococcus*, *Dialister*, *Actinomyces*, *Campylobacter,* and *Rothia* are the ones more related to Zn than to K, denoting a different origin. *Acinetobacter* displays a strong relationship mostly to Fe, but also to Al, Ca, Cd, Cr, Mn, and Th, while *Bryocella* and *UCG-05* only to Fe. *Pseudomonas* appears strongly correlated to Hf and Zr. In the case of fungi, the correlations are not so strong or nonexistent with K, in contrast to bacteria, with only Alternaria showing strong negative associations with SB, Pb, Rb, and As.

Indeed, a strong association between bacterial communities and B was detected (*SI Appendix*, Fig. S12). Namely, *Streptococcus*, *Parabacteroides*, *Veillonella*, *Enterococcus, Faecalibacterium, Clostridium,* and *Bacteroides* had a significant positive correlation ($P < 0.01$) with B (*SI Appendix*, Fig. S12), while *Serratia* had, instead, a significant negative correlation ($P < 0.05$). *Shigella* positively ($P < 0.05$) correlated with Al, B, and Cd, *Streptococcus* showed a significant negative correlation with Hf, S, and Zr ($P < 0.01$) and positively with B ($P < 0.01$). *Parabacteroides* displayed a significant negative correlation with Hf ($P < 0.05$), *Delftia* with As and Pb ($P < 0.05$), *Anaerobacillus* with Pb and Sb ($P < 0.05$), and *Porphyrobacter* negatively correlated with a group of metals (Pb, Rb, Sb, Sn, V, and As), as well as with S.

The broad analysis of the correlation between the composite microbial community detected in our air samples and the total concentration of chemical elements revealed a group of organisms that have a correlation accounting for at least 60% of the variability. Intriguingly, whereas B may have a very strong positive correlation ($P < 0.01$), Hf and Zr, typically found to be both associated in nature, showed a strong significant negative correlation ($P < 0.05$) with the majority of the detected organisms (*SI Appendix*, Fig. S12). These correlations between specific microbial communities and chemical elements may serve as indicators of the distinct geological sources contributing to the collected $PM_{2.5}$.

Characterization of Bacteria and Fungi from a Human Health Perspective. Several genera classified as the top 10 genera hosting the majority of pathogenic species, including *Streptococcus, Prevotella, Staphylococcus, Clostridium*, and *Bacteroides* (29), were observed among the top 30 most abundant genera in our samples (Fig. 4*B*). Around 35% and 39% of bacterial and fungal species detected in $PM_{2.5}$ collected from the lower troposphere or near-surface, respectively, have the potential to pose a risk to human health since those taxa may act as opportunistic pathogens (Fig. 6*A*). Species-level bioinformatic identification was conducted and among those identified species, it is worth highlighting *Serratia marcescens*, *Delftia acidovorans, Bacteroides vulgatus, Bacteroides dorei, Bacteroides uniformis, Bacteroides thetaiotaomicron, Bacteroides fragilis, Bacteroides eggerthii,* and *Alistipes onderdonkii* (Fig. 6*A*). It is also noteworthy that the species appearing in populations larger in the lower troposphere than in the surface level belong to *Bacteroide*s genera. *B. thetaiotaomicron, B. fragilis,* and *B. eggerthii* were predominantly detected in flight samples (Fig. 6*A*), while the rest of pathogenic species were present in equal amounts at both altitudes. Another intriguing observation is that a noticeable proportion of the remaining identified species constitute normal taxa of the human intestinal or oral microbiota. The distribution of those species demonstrated a notable shift between months, and a higher abundance of the opportunistic pathogens was detected in April than in February (Fig. 6*B*). Among the fungi, several human opportunistic pathogens were detected: The most prominent were *Exophiala oligosperma*, *Malassezia globosa* and *Malassezia restricta*, *Cladosporium herbarum, Sarocladium kiliense,* and *Aureobasidium pullulans* (Fig. 6*C*). Notably, all the identified fungal pathogens were found in both February and April samples, highlighting a high degree of similarity in their compositions (Fig. 6*B*).

The reliability of these species assignments as pathogens or opportunistic pathogens in the filters is high and clearly not a result of human contamination during filter processing. This is supported by both their predominant place in the samples and their absence in the corresponding blank controls (i.e., corresponding to flight and ground monitoring; Fig. 6 *A* and *C* and *SI Appendix*, Fig. S7).

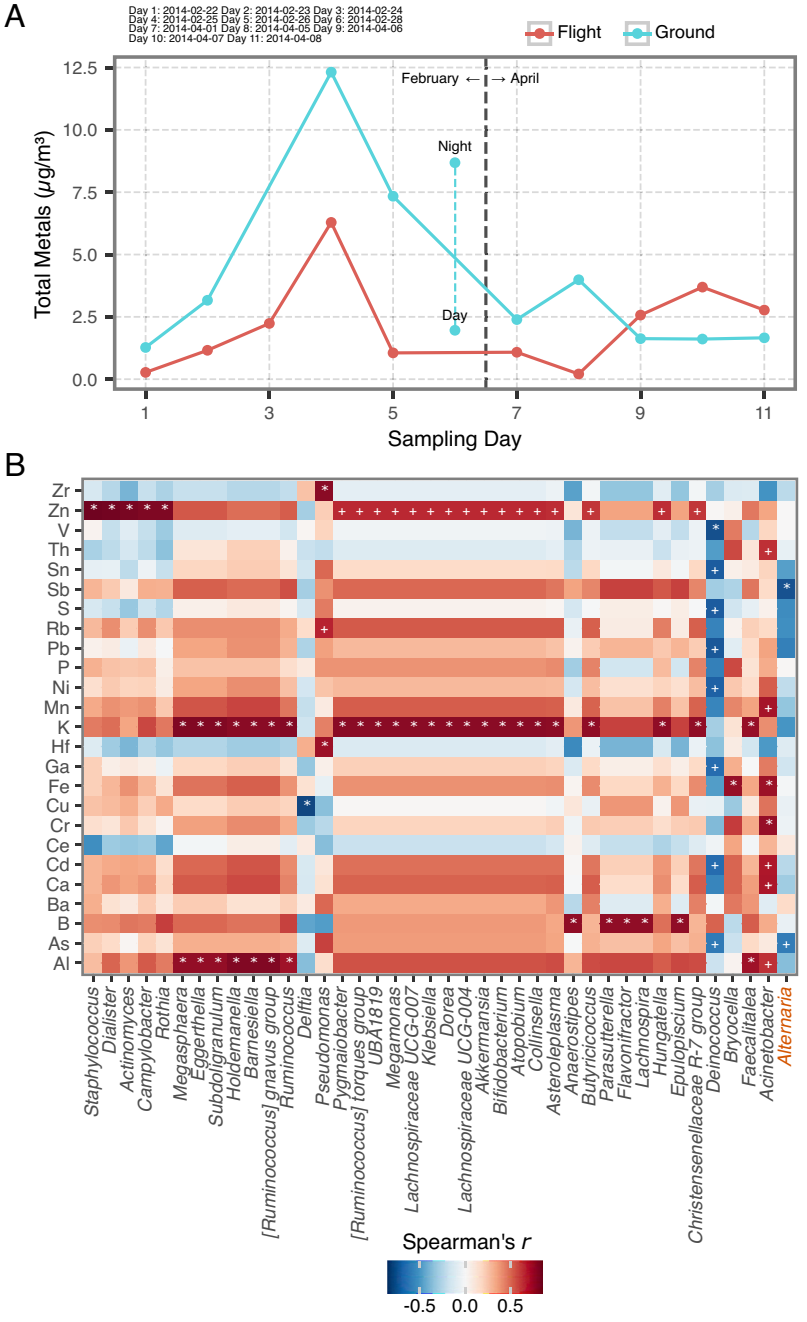


Fig. 5. Chemical composition of air masses and associations with airborne microorganisms. (*A*) Line-graph comparing daily total elemental concentrations in Flight (red) and Ground (blue) samples over the 11 d sampled. (*B*) Element-Genus Correlation Matrix: Displays Spearman's r values for the 40 bacterial and fungal genera most correlated (or anticorrelated) with chemical element/compound concentrations. Blue shades represent negative correlations, red shades indicate positive correlations. Cells with a "+" signify $P < 0.05$, and those with a "*" denote $P < 0.01$ for the Spearman rank correlation test.

Viability and Antibiotic Susceptibility Tests. To confirm whether bacteria collected from the air remained viable, we cultured a portion of the collected filters. The identification of isolated colonies was performed by MALDI-TOF MS, and when the strain could not be determined, the identification was made by sequencing the 16S rRNA gene (*Methods and Analysis*). The cultured bacteria mainly belonged to three genera: *Bacillus sp., Micrococcus sp.,* and *Lactobacillus.* (*SI Appendix*, Table S3), organisms that were rarely identified by sequencing from the DNA extracted directly from the filters.

Several *Bacillus* species were isolated from flight samples: *Bacillus pumilus* (4 strains), *Bacillus cereus* (2 strains), *Bacillus firmus* (2 strains), *Bacillus aerius* (2 strains), *Bacillus subtilis* (1 strain), *Bacillus aryabhattai* strain 2 (1 strain), *Bacillus wiedmannii* (1 strain), *Bacillus muralis* (1 strain), and *Bacillus megaterium* (1 strain). In addition, *Micrococcus luteus* (1 strain), *Micrococcus aloeverae* (1 strain) and *Lactobacillus kitasatonis* (1 strain) were cultured from samples collected in air. Furthermore, *Staphylococcus hominis* was also detected.

All isolates belonging to *B. pumilus* were susceptible to penicillin, ampicillin, amoxicillin/clavulanic acid, clindamycin, rifampicin, tetracycline, teicoplanin, imipenem, meropenem, vancomycin, ciprofloxacin, moxifloxacin, and levofloxacin, but resistant to cephalosporins. One out of four isolates were resistant to erythromycin, one strain to chloramphenicol and one to trimethoprim-sulfamethoxazole (Table 1). Both strains of *B. cereus* were resistant to

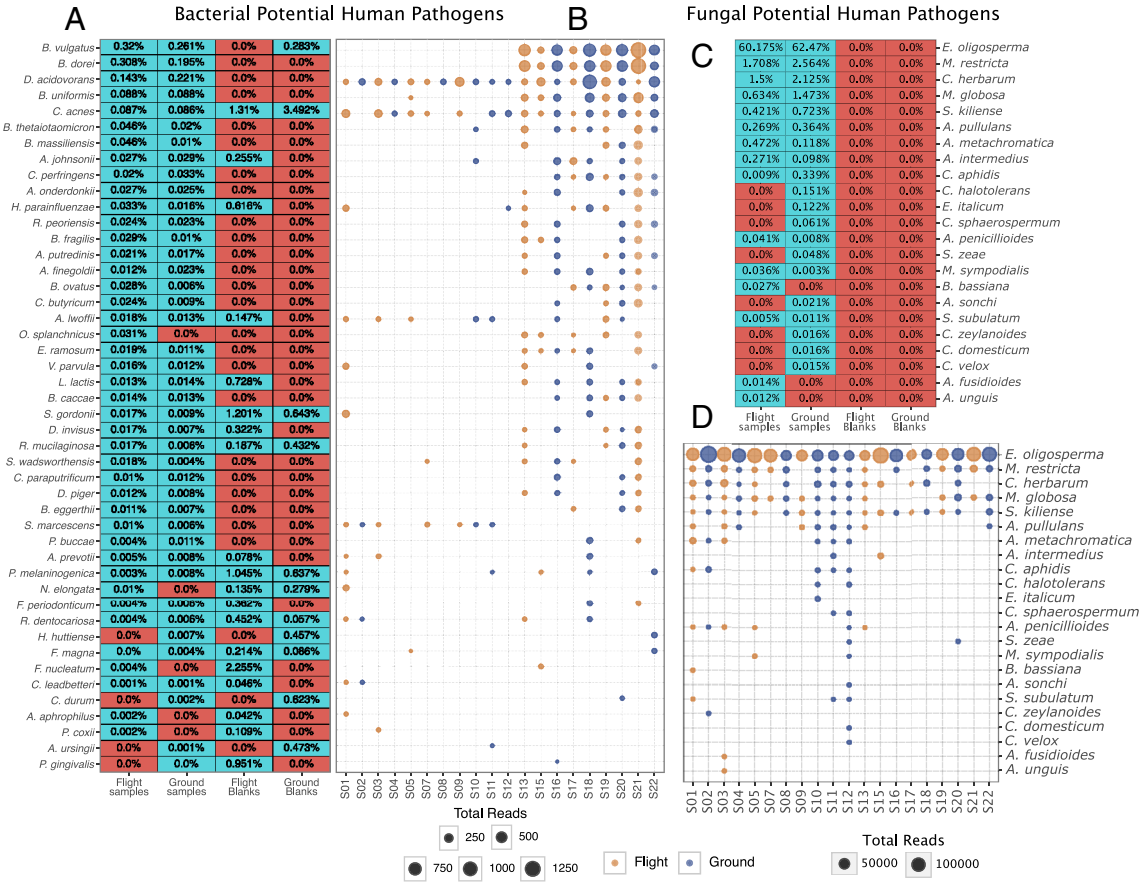


Fig. 6. Pathogenic species distribution across blanks and sample types. On (*A*) and (*B*), Average relative abundance of potentially pathogenic bacterial and fungal species identified across samples. Cells highlighted in red showcase the absence of a species in the given sample type, whereas blue cells denote a confirmed presence. On (*C*) and (*D*) Dot-plots representing the variability in total pathogen counts on a per-sample basis, distinguishing between bacterial and fungal pathogens, with blue dots for Ground samples and yellow dots for Flight samples. Only species with an average relative abundance of 0.01% or more in at least one of the sample groups are shown.

penicillin, ampicillin, cephalosporins, and clindamycin and susceptible to the remaining antibacterial agents, with the exception of trimethoprim-sulfamethoxazole, to which one of the strains was susceptible and the other one resistant. The antimicrobial susceptibility of the two isolates belonging to *B. firmus* was very similar, as both were resistant to cephalosporins, rifampicin, clindamycin, and chloramphenicol, meanwhile one of the strains was also resistant to penicillin and ampicillin but susceptible to amoxicillin-clavulanic acid and the other one resistant to erythromycin and trimethoprim-sulfamethoxazole. The other three species of *Bacillus* (*B. subtilis, B. muralis,* and *B. megaterium*), and *M. aloeverae* were susceptible to all tested antimicrobial agents. In addition, *S. hominis* was also susceptible to all tested antibiotics. The *L. kitasatonis* strain was resistant to penicillin, ampicillin, and trimethoprim-sulfamethoxazole and was intermediate to ciprofloxacin, while it could be considered susceptible to levofloxacin and resistant to moxifloxacin. The *M. luteus* strain was resistant to carbapenems, cephalosporins, glycopeptides, ciprofloxacin, and trimethoprim-sulfamethoxazole (Table 1). Thus, the antibiotic susceptibility of cultured organisms revealed that there were several MDR strains isolated from the samples.

Discussion

The identification of pathogenic organisms above the PBL indicates that large portions of the troposphere can become potential reservoirs and act as long-distance spreaders of a rich variety of microbes. These microbes are uplifted by strong air currents linked to the Siberian High and, when crossing the PBL, can travel long distances before subsiding again to the surface. Among these microbes, a vast majority of human and plant pathogens were identified, along with bacteria exhibiting antibiotic resistance profiles. In the current study, we revealed how microbial diversity values above the PBL (and up to 3,000 m a.s.l.) were comparable to those at the near-surface level, without entrainment from surface air below. This was further confirmed by both FLEXPART simulations and the daily LIDAR profiles.

Our data pointed to the air masses' mixed agricultural origin in the NE China region, given that no dust event was detected during any of the sampling days. It has been suggested that organic matrices of transparent exopolymeric particles that strongly absorb UV radiation and may contribute to the prevention of severe dehydration are responsible for the persistence and viability of both viruses and bacteria in the upper troposphere and above and may enable viable long-distance transport (30). Our study also confirms this possibility and results in Rodó et al. (14) showing a similar direction of winds from NE China and indicates that the main source of bioaerosol provenance was the vast cereal croplands in the region (31, 32).

Worth mentioning that there is, a very limited number of studies (and consequently, few datasets) on the microbial characterization at high altitude in the troposphere, as the majority of studies were performed at only a few meters above the ground level over either terrestrial or oceanic locations (10, 33) with only few exceptions (1, 34–36). Those studies implemented pumps for air

Table 1. Antibiotic resistance profile of the cultured taxa from air samples

	M. luteus (from S15)	*B. aryabhattai* (from S17)	*B. wiedmannii* (from S19)	*B. aerius* (from S21) (2 Strains)*	*B. firmus* (from S15 /S3) (2 strains)†	*B. pumilus* (from S5/S7/ S17/S19) (4 strains)†	*S. hominis* (from S1)	*L. kitasatonis* (from S19)	*B. cereus* (from S15, S17) (2 strains)†
Penicillin	≤0.03 S	>4 R	>4 R	<0.03 S	>4/2 R/S	<0.03 S		>4 R	
Ampicillin	0.25 S	>4 R	>4 R	<0.12 S	>4/1 R/ S	<0.12 S		>4 R	>4 R
Amox+clav	<1 S	<0.5 S	>4 R	<0.5 S	<0.5 S	≤1 S		≤0.5 S	>4 R
Cefepime	>2 R	<0.25 S	>2 R	>2 R	>2 R	>2 R		>2 R	>2 R
Cefotaxime	>2 R	<0.06 S	>2 R	>2 R	>2 R	>2 R		>2 R	>2 R
Imipenem	>4 R	<0.12 S	<0.12 S	<0.12 S	0.25 S	<0.25 S	<1 S	<0.12 S	≤0.5 S
Meropenem	>2 R	<0.25 S	<0.25 S	<0.25 S	<0.25 S	<0.25 S		<0.25 S	0.25 S
Vancomycin	>2 R	<0.5 S	<0.5 S	<0.5 S	<0.5 S	<0.5 S	<0.5 S	<0.5 S	<0.5 S
Teicoplanin	2	<1 S	<1 S	<1 S	<1 S	≤1 S	<0.5 S	<1 S	<1 S
Erithromycin	>4 R	4 I	<0.25 S	0.5 S	0.5 S/16 R	<0.25 S/16R	≤0.25 S	<0.25 S	≤0.25 S
Ciprofloxacin	>2 R	0.12 S	0.06 S	0.25 S	1	0.5 S		2	≤0.5 S
Moxifloxacin	2	<0.5 S	<0.5 S	<0.5 S	<0.5 S	<0.5 S	<025 S	>2	<0.5 S
Levofloxacin	>4 R	<0.5 S	<0.5 S	<0.5 S	2 S	<0.5 S		<0.5 S	<0.5 S
Co-trimoxazole	>5/76 R	<0.5 S	<0.5 S	<0.5 S	<0.5 S/>4-76 R	<0.5 S/>4-76 R		>4-76 R	<0.5 S/>4-76 R
Clindamycin	0.12 S	>0.5 R	0.5 S	>0.5 R	>0.5 R	≤0.25 S	0.25 S	0.5 S	>0.5 R
Chloramphenicol	4 S	4 S	2 S	4 S	>8 R	4 S/>8 R		8	2 S
Rifampicin	<0.25 S	<0.25 S	<0.25 S	<0.25 S	>2 R	<0.25 S		–	<0.25 S
Tetracycline	<1 S	<1 S	<1 S	<1 S	<1 S	<1 S		<1 S	<1 S
Quinupristin/ dalfopristin	2 I	2 I	2 I	<1 S	2 I	2 I	<0.5 S	<1 S	<1 S
Daptomycin							0.25 S		
Gentamicin							≤1 S		
Tobramicin							≤1 S		
Linezolid							1 S		
Tigecycline							<0.125 S		

The specified value represents the minimum inhibitory concentration (MIC), measured in μg/mL, for each antibiotic individually. "R" denotes resistance, indicating that the studied strain is resistant to the tested antibiotic at a concentration higher than the indicated MIC value. "S" signifies susceptibility, meaning that the strain is susceptible to the tested antibiotic at a concentration lower than the specified MIC value.

*Both strains showed the same antibiotic susceptibility profile.

†Differences in the MIC depending on the strain.

collection with very low flow rates (8.5 L/min, 14,15 L/min, and 17 L/min) and a short time of sampling (30 to 40 min, 5 min, and 60 min) (34–36). One of the reasons for this lack of studies is possibly the extremely low biomass content of air samples, which presents a challenge for the extraction of sufficient genetic material of high enough quality for efficient sequencing analysis (37). The main airborne bacterial genera detected both in flight and surface samples are *Sphingomonas* and *Methylobacterium*, which echoes other studies where they were reported as dominant airborne taxa in outdoor environments, particularly in the lower troposphere across different geographic regions (3, 38, 39) particularly in Japan (40). Those phyla ubiquitously reported in air samples include spore-forming bacteria with known UV and desiccation tolerance traits considered advantageous during atmospheric transport (41, 42). *Methylobacterium* can use C1-C4 carbon compounds that are ubiquitously present in the atmosphere, showcasing its adaptability to thrive in complex atmospheric conditions (43).

Other bacterial genera detected in our study are commonly distributed in natural environments (water and soil), namely: *Serratia* (44), *Bacteroides* (45), *Streptococcus* (46), *Enterococcus* (47), all of which contain species that play a role as human opportunistic pathogens.

Although the majority of *Exophiala sp.* are common outdoor and indoor environmental fungi, some of them are also involved in human infection (48). Of note, some *Exophiala* species are able to decompose toxic chemicals (49) and *Cladosporium, Alternaria,* and *Penicillium* were previously detected in air samples and considered the most prevalent fungi genera in ambient air (50, 51). *Alternaria* is a common fungus producing a variety of mycotoxins and a known plant pathogen (52). In grain facilities, *Penicillium, Alternaria, Cladosporium* were detected as predominant fungal genera, which also include plant and animal pathogenic and allergenic species (53). Other detected fungal communities include *Trametes*, a widespread wood decomposer that plays a role in urban wastewater purification by removing micropollutants (54) and *Sporisorium*, a known plant pathogen that causes sugarcane smut and is widespread in all the major sugarcane production areas in China (55).

While changes in bacterial taxa distribution in near-surface air are a well-known phenomenon (33, 56, 57), such alterations in the lower troposphere have not been widely reported. Here, we present seminal evidence demonstrating temporal changes in the bacterial and fungal diversity at high altitudes. Temperature, humidity, plant crop seasonality, and common agricultural practices significantly influence the seasonal release and diffusion of airborne microbial populations (58). Therefore, not surprisingly, the primary identified taxa at high altitudes were those that had developed an adaptation enabling them to thrive in the challenging conditions, among others

Gram-positive endospore-forming *Bacillus sp*, as well as *Exophiala, Penicillium,* and *Alternaria sp,* which were even detected in Mars 2020 spacecraft, highlighting their high radiation resistance (59).

Interestingly, among the identified genera collected well above the PBL, there were not only environmental organisms but also known potential human pathogens. Among the principal health effects are those initially associated with respiratory complications that, in turn, may affect downstream inflammatory signaling (60), also inducing cell and organ damage (61) Among those species identified from the filters, *B. fragilis* is of particular interest. Despite constituting part of a normal human gut microbiome, once it leaves the intestine due to the disruption of the intestinal wall integrity, *B. fragilis* can cause severe infections (i.e., to name a few: sepsis; peritonitis; soft tissue infections; pelvic, lung, and brain abscesses; and a toxin-associated diarrhea) (62). Less is known about the environmental presence of *B. fragilis,* despite a recent study demonstrating its presence in wastewater (63). Another organism, *S. marcescens,* is a ubiquitous bacterium that exists in multiple ecological niches, including the air (64). It has the potential to induce a wide range of infections such as pneumonia, sepsis, meningitis, peritonitis, endocarditis, arthritis, osteomyelitis, keratitis, and urinary tract and skin infections, and its involvement in various outbreaks worldwide has been confirmed (65). The most predominant fungal species in our samples was *E. oligosperma*, frequently detected in the environment but that, in immunocompromised patients, it may induce soft-tissue infection, and it has been identified as a potential trigger of ocular and periocular sarcoidosis (66). *Malassezia* species are part of the normal skin flora of humans and animals (67) and its higher abundance in the near surface area is expected. Though *M. restricta* is a common skin-resident fungus, its enhanced abundance in the gut is associated with high inflammatory cytokines and recently was reported to be involved in inflammatory bowel disease, particularly Crohn's disease, a pathology associated with changes in mycobiota (68). *M. restricta* has been implicated in various diseases of the skin, including dandruff, pityriasis versicolor, seborrheic dermatitis, folliculitis, psoriasis, and atopic dermatitis, thanks to its ability, under appropriate conditions, to invade the stratum corneum and interact with the host's immune system, both directly and through chemical mediators (69). Although the link between the inhalation of airborne particles and a suite of diseases and ailments has long been suspected, not much is known at the population level about the pathogenic effects of the microbial portion of aerosols and the chemistry of dust particles.

Due to biases generated during sampling preparation, identification at the genus level has been deemed more reliable and is frequently used (70). However, in our study, to also inform at the species level, we conducted a thorough and meticulous in-depth sequential analysis by employing the RDP Naive Bayesian Classifier together with BLASTN. Following this, we only accepted taxonomic assignments when both methods, RDP Naive Bayesian Classifier and BLASTN, yielded coinciding species identities, therefore ensuring a more robust and accurate classification process.

The analysis of elemental composition of the collected $PM_{2.5}$ particles revealed high concentrations of aluminum (Al) in both ground and flight samples. Al typically occurs as silicates in clays, feldspars, kyanite, and many other minerals as the primary terrestrial sources. This observation, along with the strong correlation with K of the majority of the microbial taxa, suggests a potential agricultural origin, in line with the vast local presence of intensively farmed croplands in NE China. The appearance of strong correlations of K and Zn with the main taxa from flight samples, together with their association to clays, Cd, and Mn further reinforces the alleged agricultural origin and the potential link to areas with high loads of fertilizers and pesticides. Indeed, it is well known that the combined application of zinc sulfate nanoparticles and potassium fertilizers effectively stimulates the growth and quality of maize (*Zea mays* L.) in soils contaminated with metals, particularly with Cd (71). These nutrients synergistically enhance maize grain quality, ensure food safety, and alleviate plant's stress by improving growth. This example illustrates the combined application of fertilizers, like zinc sulfate nanoparticles and sulfate of potash, in Cd-contaminated soils to boost maize biomass, alleviate abiotic stresses, and enhance the crop's nutritional value. In turn, plants readily absorb and accumulate Cd in their tissues and in order to combat its toxicity in soils, various zinc fertilizers, including nano-fertilizers, have proven effective in enhancing zinc efficacy in such soils (72).

Interestingly, Zr and Hf are trace elements with very limited global availability and are not mined anywhere in Japan (73), despite appearing in all our samples. Instead, China being the third-largest global producer of these elements and the backward simulation paths crossing the main producing regions like Inner Mongolia (which accounts for 70% of the national total) (74), provide additional support for the findings of our study. These also coincide well with the backtrajectory simulations included indicating this area in NE China and beyond as the likely source. Additionally, B -another element highly correlated with the microorganisms found in our study- does not appear to be mined in Japan either and, again, both the coastal and inland regions of China emerge as primary mine source regions (75). The negative correlations between these elements originate from the observation of Hf and Zr predominance in February samples, while B was predominant in April samples. Finally, given the range of nanogram concentrations found in our aerosol samples, they most likely originated from terrestrial sources in one of these regions.

Metals are considered one of the most harmful components in fine particles with reported effects on pulmonary health and allergic responses (56, 76). Additionally, bioaerosols may play a role in modifying PM toxicity by modulating the oxidative potential of toxic chemicals presented within PM, thus representing a potential hazard to human health (61). Additionally, certain microbial species detected in our sampled bioaerosols, including those known to cause human allergies and respiratory diseases, further compound this health risk (77). Although some reports highlight the antimicrobial properties of B against pathogens like *Candida, Aspergillus*, and *Staphylococcus* (78), the significant positive correlations between the *Acinetobacter* and metals such as Cd, Fe, Cr, Mn, and Th, identified in the current study, underscore the need for a reassessment of bioaerosols' health impact (79).

From a health perspective, cultured organisms indicate potential threats to human health, despite the fact that low dose concentrations obviously limit their impacts in healthy individuals (albeit possibly not for those susceptible or immunocompromised individuals). For instance, the potential risks of *Bacillus sp*. to public health have been well documented (80). *B. pumilus* is linked to sepsis in both immunocompromised and immunocompetent patients (81, 82). Some strains of this species can produce a heat-stable enterotoxin (83), classifying them as food-borne pathogens causing gastroenteritis. Moreover, *B. pumilus* can cause skin infections similar to cutaneous anthracis lesions (84). Our airborne *B. pumilus* strain seems susceptible to most tested antibiotics, and it has been reported to exhibit MDR activity in some cases (85, 86). *B. firmus*, although multidrug resistant, does not show infectivity and pathogenicity, and instead it is used as a seed

treatment resulting in increased plant and root biomass (87). *B. cereus* is recognized as potentially pathogenic in humans. Also, food poisoning caused by *B. cereus* results in acute foodborne intoxication when this microorganism produces toxins. *B. cereus*, whose spores survive in dried raw rice, is considered a relatively common cause of gastroenteritis worldwide, and, although normally only resulting in mild symptoms, it has been associated with four fatal cases. It has also been shown to cause endocarditis mainly in intravenous drug users and patients with valvular heart disease, those with pacemakers, prosthetic mitral valves, and other underlying conditions (88). In addition, for *B. cereus*, bacteremia, brain abscess, and meningoencephalitis, as well as eye infections such as pan ophthalmitis and endophthalmitis, have been reported. Neonates are also at particular risk for hospital-acquired *B. cereus* infections (89). Different studies have also shown that *B. cereus* is highly resistant to penicillin, ampicillin, and ceftriaxone, which our results largely confirm. *M. luteus* is an environmental organism found in soil, dust, water, and in human skin microbiota as well. It is an opportunistic pathogen, which can cause infections such as general bacteremia, and endocarditis. Despite that *M. luteus* is normally susceptible to most antibiotics, the strain isolated in this study showed multidrug resistance, displaying resistance to carbapenems, glycopeptides, erythromycin, fluoroquinolones, and trimethoprim-sulfamethoxazole. The mechanisms underlying this resistant phenotype have not been investigated yet. In the scientific literature, only a plasmid-borne macrolide resistance has been shown, which is the mechanism of resistance at 23S rRNA adenine methyltransferase encoded in the Erm36 gene (90). If so, it could potentially be transmitted to other microorganisms through mobile genetic elements, a matter that requires further investigation.

Our study has several limitations. First, the 10 flights do not cover all year-round variability, despite already representing a significant effort to characterize the tropospheric microbiome. Second, despite our two-stage bioinformatic analysis reinforcing the presentation of species taxonomic assignments, other sequencing approaches may offer better performances.

Conclusions

Our study uncovered a rich diversity of microbial taxa being dispersed by wind currents thousands of kilometers away from their sources. The transport takes place above the PBL, far from the surface, which enables long-distance connections among geographic sites, thereby opening the door to even much longer propagations. The number of such flights is unprecedented and should further promote others to follow exploring the presence of microbial life in the free troposphere.

The isolation of harmful species to humans (i.e., also others to animals and plants), had never been reported before for such long distances. The link of aerosol particles to remote agricultural areas, leaves it open the potential role of pesticides, fertilizers, or other components derived from anthropogenic activities ultimately capable of affecting human health, far from their sources.

Also, the fact that some of the cultured bacteria showed antimicrobial resistance capacity adds further novelty to our study as such route of propagation was never reported.

While our study does not necessarily prove causality between the presence of known human pathogens in bioaerosols and health effects, it does pave the way for further research along these lines. This future avenue could examine the exchange of microbial pathogens across very long distances and of very different sorts of pathogens (bacteria, fungi, and viruses as well, although not specifically addressed in this study). Our results introduce a novel perspective on the potential transmission routes of bioaerosol particles and their association with human, plant, and animal diseases, suggesting a need for further research to explore potential mechanistic connections in more detail.

Data, Materials, and Software Availability. Sequencing results are available in ref. 91. Data, code, and outputs of the analysis can be accessed in ref. 92 in an open repository at GitHub with the following link: https://github.com/AirLabBcn/microbial-richness-troposphere. All other data are included in the manuscript and/or *SI Appendix*.

ACKNOWLEDGMENTS. We acknowledge the financial support from the Daniel Bravo Andreu Private Foundation through the research Grant WINDBIOME and the collaboration with Fundació per al Foment de la Investigació Sanitària i Biomèdica de la Comunitat Valenciana. We also strongly thank the engineers at National Institute for Environmental Studies (Tsukuba) and pilots at Chofu airport for helping in the conditioning of the aircraft and the installation of the sampling equipment. ISGlobal researchers acknowledge support from the grant CEX2018-000806-S funded by MCIN/AEI/10.13039/501100011033, and support from the Generalitat de Catalunya through the Centres de Recerca de Catalunya program. A.F. acknowledges financial support from HELICAL as part of the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie Grant Agreement No. 81354.

Author affiliations: [a]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain; [b]Climate and Health Group, Climate, Air pollution, Nature and Urban Health, Barcelona Institute for Global Health, Barcelona 08003, Spain; [c]Division of Atmospheric Environmental Studies, Kanazawa University, Kanazawa 920-1164, Japan; [d]Earth System Division, National Institute for Environmental Studies, Tsukuba 305-8506, Japan; [e]Translational Genomics Facility, Fundació Institut de Recerca en Ciències de la Salut Germans Trias i Pujol, Badalona 08916, Spain; [f]Department of Clinical Microbiology, Biomedical Diagnostic Center, Hospital Clinic School of Medicine, University of Barcelona, Barcelona 08036, Spain; [g]Centro de Investigación Biomédica en Red de Enfermedades Infecciosas, Instituto de Salud Carlos III, Madrid 28029, Spain; [h]Department of Microbiology, Genetics and Statistics, Faculty of Biology, University of Barcelona, Barcelona 08028, Spain; and [i]Ionising Radiation, Health and Environment, Institute of Energy Technologies, Universitat Politècnica de Catalunya, Barcelona 08028, Spain

1. D. I. Drautz-Moses *et al.*, Vertical stratification of the air microbiome in the lower troposphere. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117293119 (2022).
2. H. Behzad, T. Gojobori, K. Mineta, Challenges and opportunities of airborne metagenomics. *Genome Biol. Evol.* **7**, 1216–1226 (2015).
3. A. M. Klein, B. J. M. Bohannan, D. A. Jaffe, D. A. Levin, J. L. Green, Molecular evidence for metabolically active bacteria in the atmosphere. *Front. Microbiol.* **7**, 772 (2016).
4. M. M. Cox, J. R. Battista, Deinococcus radiodurans–The consummate survivor. *Nat. Rev. Microbiol.* **3**, 882–892 (2005).
5. D. V. Spracklen, C. L. Heald, The contribution of fungal spores and bacteria to regional and global aerosol number and ice nucleation immersion freezing rates. *Atmos. Chem. Phys.* **14**, 9051–9059 (2014).
6. K. Tang *et al.*, Characterization of atmospheric bioaerosols along the transport pathway of Asian dust during the Dust-Bioaerosol 2016 Campaign. *Atmos. Chem. Phys.* **18**, 7131–7148 (2018).
7. Y. S. Joung, Z. Ge, C. R. Buie, Bioaerosol generation by raindrops on soil. *Nat. Commun.* **8**, 14668 (2017).
8. T. Maki *et al.*, Long-range transport of airborne bacteria over East Asia: Asian dust events carry potentially nontuberculous Mycobacterium populations. *Environ. Int.* **168**, 107471 (2022).
9. P. Amato *et al.*, The aeromicrobiome: The selective and dynamic outer-layer of the Earth's microbiome. *Front. Microbiol.* **14**, 1186847 (2023).
10. E. Mayol *et al.*, Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nat. Commun.* **8**, 201 (2017).
11. N. Yamaguchi, T. Ichijo, A. Sakotani, T. Baba, M. Nasu, Global dispersion of bacterial cells on Asian dust. *Sci. Rep.* **2**, 525 (2012).
12. I. Reche, G. D'Orta, N. Mladenov, D. M. Winget, C. A. Suttle, Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J.* **12**, 1154–1162 (2018).
13. T. Šantl-Temkiv, P. Amato, E. O. Casamayor, P. K. H. Lee, S. B. Pointing, Microbial ecology of the atmosphere. *FEMS Microbiol. Rev.* **46**, fuac009 (2022).
14. X. Rodó *et al.*, Tropospheric winds from northeastern China carry the etiologic agent of Kawasaki disease from its source to Japan. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7952–7957 (2014).
15. D. W. Griffin, V. H. Garrison, J. R. Herman, E. A. Shinn, African desert dust in the Caribbean atmosphere: Microbiology and public health. *Aerobiologia* **17**, 203–213 (2001).
16. J. M. Prospero, "Interhemispheric transport of viable fungi and bacteria from Africa to the Caribbean with soil dust" in *Biological Resources and Migration*, D. Werner, Ed. (Springer Berlin Heidelberg, 2004), pp. 127–133.
17. V. R. Després *et al.*, Primary biological aerosol particles in the atmosphere: A review. *Tellus B Chem. Phys. Meteorol.* **64**, 15598 (2012).
18. J. K. M. Brown, M. S. Hovmøller, Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537–541 (2002).

19. S. Otake, S. Dee, C. Corzo, S. Oliveira, J. Deen, Long-distance airborne transport of infectious PRRSV and Mycoplasma hyopneumoniae from a swine population infected with multiple viral variants. *Vet. Microbiol.* **145**, 198–208 (2010).
20. WMO Secretariat, "Sand and dust storm warning advisory and assessment system (SDS-WAS). Science and implementation plan 2015–2020" (GAW Report No. 279, Research Department, Atmospheric Research and Environment Branch, 2015).
21. A. Zargar, M. Maurin, E. Mostafavi, Tularemia, a re-emerging infectious disease in Iran and neighboring countrie. *Epidemiol. Health* **37**, e2015011 (2015).
22. A. M. Rule, K. A. Koehler, Particle constituents and oxidative potential: Insights into differential fine particulate matter toxicity. *Am. J. Respir. Crit. Care Med.* **206**, 1310–1312 (2022).
23. Z. Ren-Jian, H. Kin-Fai, S. Zhen-Xing, The role of aerosol in climate change, the environment, and human health. *Atmos. Ocean. Sci. Lett.* **5**, 156–161 (2012).
24. Z. Zhou *et al.*, Association between particulate matter $(PM)_{2.5}$ air pollution and clinical antibiotic resistance: A global analysis. *Lancet Planet. Health* **7**, e649–e659 (2023).
25. A. Font, J. A. Morguí, X. Rodó, Atmospheric CO_2 in situ measurements: Two examples of Crown Design flights in NE Spain. *J. Geophys. Res.* **113**, D12308 (2008).
26. I. Pisso *et al.*, The Lagrangian particle dispersion model FLEXPART version 10.4. *Geosci. Model Dev.* **12**, 4955–4997 (2019).
27. T. Moreno *et al.*, Daily and hourly chemical impact of springtime transboundary aerosols on Japanese air quality. *Atmos. Chem. Phys.* **13**, 1411–1424 (2013).
28. X. Rodó *et al.*, Association of Kawasaki disease with tropospheric wind patterns. *Sci. Rep.* **1**, 152 (2011).
29. A. Bartlett, D. Padfield, L. Lear, R. Bendall, M. Vos, A comprehensive list of bacterial pathogens infecting humans. *Microbiology*, 10.1099/mic.0.001269 (2022).
30. J. Y. Aller *et al.*, Size-resolved characterization of the polysaccharidic and proteinaceous components of sea spray aerosol. *Atmos. Environ.* **154**, 331–347 (2017).
31. X. Rodó *et al.*, Revisiting the role of environmental and climate factors on the epidemiology of Kawasaki disease. *Ann. N. Y. Acad. Sci.* **1382**, 84–98 (2016).
32. X. Rodó *et al.*, Sub-weekly signatures relate ultrafine aerosols enriched in metals from intensive farming and urban pollution to Kawasaki disease. *Environ. Res. Lett.* **18**, 074011 (2023).
33. J. Uetake *et al.*, Seasonal changes of airborne bacterial communities over Tokyo and influence of local meteorology. *Front. Microbiol.* **10**, 1572 (2019).
34. C. Jaing *et al.*, Sierra Nevada sweep: Metagenomic measurements of bioaerosols vertically distributed across the troposphere. *Sci. Rep.* **10**, 12399 (2020).
35. F. Kobayashi *et al.*, Atmospheric bioaerosol, Bacillus sp., at an altitude of 3,500 m over the Noto Peninsula: Direct sampling via aircraft. *Asian J. Atmos. Environ.* **5**, 164–171 (2011).
36. L. N. Kobziar *et al.*, Accessing the life in smoke: A new application of unmanned aircraft systems (UAS) to sample wildland fire bioaerosol emissions and their environment. *Fire* **2**, 56 (2019).
37. I. Luhung *et al.*, Experimental parameters defining ultra-low biomass bioaerosol analysis. *NPJ Biofilms Microbiomes* **7**, 37 (2021).
38. T. Maki *et al.*, Variations in airborne bacterial communities at high altitudes over the Noto Peninsula (Japan) in response to Asian dust events. *Atmos. Chem. Phys.* **17**, 11877–11897 (2017).
39. J. Cáliz, X. Triadó-Margarit, L. Camarero, E. O. Casamayor, A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12229–12234 (2018).
40. D. Tanaka *et al.*, Size resolved characteristics of urban and suburban bacterial bioaerosols in Japan as assessed by 16S rRNA amplicon sequencing. *Sci. Rep.* **10**, 12406 (2020).
41. S. D. J. Archer *et al.*, Airborne microbial transport limitation to isolated Antarctic soil habitats. *Nat. Microbiol.* **4**, 925–932 (2019).
42. A. C. Tastassa, Y. Sharaby, N. Lang-Yona, Aeromicrobiology: A global review of the cycling and relationships of bioaerosols with the atmosphere. *Sci. Total Environ.* **912**, 168478 (2024).
43. N. DeLeon-Rodriguez *et al.*, Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2575–2580 (2013).
44. S. D. Mahlen, Serratia infections: From military experiments to current practice. *Clin. Microbiol. Rev.* **24**, 755–791 (2011).
45. H. M. Wexler, Bacteroides: The good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**, 593–621 (2007).
46. C. Dupont *et al.*, Streptococcus pseudopneumoniae, an opportunistic pathogen in patients with cystic fibrosis. *J. Cyst. Fibros.* **19**, e28–e31 (2020).
47. S. Ramos, V. Silva, M. Dapkevicius, G. Igrejas, P. Poeta, Enterococci, from harmless bacteria to a pathogen. *Microorganisms* **8**, 1118 (2020).
48. Y. Song *et al.*, Comparative ecology of capsular exophiala species causing disseminated infection in humans. *Front. Microbiol.* **8**, 2514 (2017).
49. E. Estévez, M. C. Veiga, C. Kennes, Biodegradation of toluene by the new fungal isolates Paecilomyces variotii and Exophiala oligosperma. *J. Ind. Microbiol. Biotechnol.* **32**, 33–37 (2005).
50. M. Shams-Ghahfarokhi, S. Aghaei-Gharehbolagh, N. Aslani, M. Razzaghi-Abyaneh, Investigation on distribution of airborne fungi in outdoor environment in Tehran, Iran. *J. Environ. Health Sci. Eng.* **12**, 54 (2014).
51. C. Woo, C. An, S. Xu, S.-M. Yi, N. Yamamoto, Taxonomic diversity of fungi deposited from the atmosphere. *ISME J.* **12**, 2051–2060 (2018).
52. M. DeMers, Alternaria alternata as endophyte and pathogen. *Microbiol. Read. Engl.* **168**, 001153 (2022).
53. A. Straumfors, S. Mundra, O. A. H. Foss, S. K. Mollerup, H. Kauserud, The airborne mycobiome and associations with mycotoxins and inflammatory markers in the Norwegian grain industry. *Sci. Rep.* **11**, 9357 (2021).
54. A. Cruz Del Álamo, M. I. Pariente, F. Martínez, R. Molina, Trametes versicolor immobilized on rotating biological contactors as alternative biological treatment for the removal of emerging concern micropollutants. *Water Res.* **170**, 115313 (2020).
55. P. Singh *et al.*, Whole genome analysis of sugarcane root-associated endophyte Pseudomonas aeruginosa B18-A plant growth-promoting bacterium with antagonistic potential against Sporisorium scitamineum. *Front. Microbiol.* **12**, 628376 (2021).
56. S. Jiang *et al.*, Airborne microbial community structure and potential pathogen identification across the PM size fractions and seasons in the urban atmosphere. *Sci. Total Environ.* **831**, 154665 (2022).
57. A. Núñez, A. M. García, D. A. Moreno, R. Guantes, Seasonal changes dominate long-term variability of the urban air microbiome across space and time. *Environ. Int.* **150**, 106423 (2021).
58. D. Lagomarsino Oneto, J. Golan, A. Mazzino, A. Pringle, A. Seminara, Timing of fungal spore release dictates survival during atmospheric transport. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5134–5143 (2020).
59. A. M. Chander *et al.*, Draft genome sequences of fungi isolated from Mars 2020 spacecraft assembly facilities. *Microbiol. Resour. Announc.* **11**, e0046422 (2022).
60. T. Ruiz-Gil *et al.*, Airborne bacterial communities of outdoor environments and their associated influencing factors. *Environ. Int.* **145**, 106156 (2020).
61. A. Samake *et al.*, The unexpected role of bioaerosols in the oxidative potential of PM. *Sci. Rep.* **7**, 10978 (2017).
62. M. Yekani *et al.*, To resist and persist: Important factors in the pathogenesis of Bacteroides fragilis. *Microb. Pathog.* **149**, 104506 (2020).
63. M. A. Tariq *et al.*, Genome characterization of a novel wastewater Bacteroides fragilis bacteriophage (vB_BfrS_23) and its host GB124. *Front. Microbiol.* **11**, 583378 (2020).
64. N. E. Gaultier *et al.*, Complete genome sequence of the bacterium Serratia marcescens SGAir0764, isolated from Singapore air. *Genome Announc.* **6**, e00637-18 (2018).
65. F. Tavares-Carreon, K. De Anda-Mora, I. C. Rojas-Barrera, A. Andrade, *Serratia marcescens* antibiotic resistance mechanisms of an opportunistic pathogen: A literature review. *PeerJ* **11**, e14399 (2023).
66. A. S. Shifera *et al.*, Identification of microbial agents in tissue specimens of ocular and periocular sarcoidosis using a metagenomics approach. *F1000Research* **10**, 820 (2021).
67. C. W. Saunders, A. Scheynius, J. Heitman, Malassezia fungi are specialized to live on skin and associated with dandruff, eczema, and other skin diseases. *PLoS Pathog.* **8**, e1002701 (2012).
68. J. J. Limon *et al.*, Malassezia is associated with Crohn's disease and exacerbates colitis in mouse models. *Cell Host Microbe* **25**, 377–388.e6 (2019).
69. A. Gupta, R. Bluhm, Seborrheic dermatitis. *J. Eur. Acad. Dermatol. Venereol.* **18**, 13–26 (2004).
70. X. C. Morgan, C. Huttenhower, Chapter 12: Human microbiome analysis. *PLoS Comput. Biol.* **8**, e1002808 (2012).
71. M. Umair *et al.*, Unraveling the effects of zinc sulfate nanoparticles and potassium fertilizers on quality of maize and associated health risks in Cd contaminated soils under different moisture regimes. *Sci. Total Environ.* **896**, 165147 (2023).
72. R. Ahmed *et al.*, Differential response of nano zinc sulphate with other conventional sources of Zn in mitigating salinity stress in rice grown on saline-sodic soil. *Chemosphere* **327**, 138479 (2023).
73. J. Sanz, O. Tomasa, A. Jimenez-Franco, N. Sidki-Rius, "Zirconium (Zr) [Z = 40]" in *Elements and Mineral Resources. Springer Textbooks in Earth Sciences, Geography and Environment* (Springer, Cham, 2022), pp. 255–258.
74. X. Zhu *et al.*, Investigating zirconium flows and stocks in China: A dynamic material flow analysis. *Resour. Policy* **80**, 103139 (2023).
75. C. Helvacı, M. R. Palmer, Origin and distribution of evaporite borates: The primary economic sources of boron. *Elements* **13**, 249–254 (2017).
76. N. V. Dos Santos *et al.*, Accumulation of trace element content in the lungs of Sao Paulo city residents and its correlation to lifetime exposure to air pollution. *Sci. Rep.* **12**, 11083 (2022).
77. C. Cao *et al.*, Inhalable microorganisms in Beijing's PM2.5 and PM10 pollutants during a severe smog event. *Environ. Sci. Technol.* **48**, 1499–1507 (2014).
78. Z. Iyigundogdu, I. Saribas, The effect of various boron compounds on the antimicrobial activity of hardened mortars. *Constr. Build. Mater.* **351**, 128958 (2022).
79. M. Sauget, B. Valot, X. Bertrand, D. Hocquet, Can MALDI-TOF mass spectrometry reasonably type bacteria? *Trends Microbiol.* **25**, 447–455 (2017).
80. C. J. Carlson *et al.*, The global distribution of Bacillus anthracis and associated anthrax risk to humans, livestock and wildlife. *Nat. Microbiol.* **4**, 1337–1343 (2019).
81. V. M. Shivamurthy *et al.*, Bacillus pumilus septic arthritis in a healthy child. *Can. J. Infect. Dis. Med. Microbiol.* **2016**, 3265037 (2016).
82. D. C. Garcia-Ramon *et al.*, The parasporal crystals of Bacillus pumilus strain 15.1: A potential virulence factor? *Microb. Biotechnol.* **11**, 302–316 (2018).
83. M. M. Shah, G. Miringu, A. Wada, S. Kaneko, Y. Ichinose, Case report: Bacillus pumilus-caused bacteremia in a patient with food poisoning. *Am. J. Trop. Med. Hyg.* **100**, 688–690 (2019).
84. D. Tena *et al.*, Cutaneous infection due to Bacillus pumilus: Report of 3 cases. *Clin. Infect. Dis.* **44**, e40–e42 (2007).
85. F. Celandroni *et al.*, Identification and pathogenic potential of clinical bacillus and Paenibacillus isolates. *PLoS One* **11**, e0152831 (2016).
86. J. Freitas-Silva, T. Silva-Oliveira, G. Muricy, M. S. Laport, Bacillus strains associated to homoscleromorpha sponges are highly active against multidrug resistant bacteria. *Curr. Microbiol.* **77**, 807–815 (2020).
87. M. Seymen, *Sustainable Horticulture Microbial Inoculants and Stress Interaction* (Academic Press, 2022).
88. M. K. Steen *et al.*, Bacillus cereus endocarditis: Report of a case and review. *Clin. Infect. Dis.* **14**, 945–946 (1992).
89. M. Ehling-Schulz, D. Lereclus, T. M. Koehler, The *Bacillus* cereus group: *Bacillus* species with pathogenic potential. *Microbiol. Spectr.*, 10.1128/microbiolspec.gpp3-0032-2018 (2019).
90. W. Liebl, W. E. Kloos, W. Ludwig, Plasmid-borne macrolide resistance in Micrococcus luteus. *Microbiology* **148**, 2479–2487 (2002).
91. X. Rodó, S. Pozdniakova, S. Borràs, A. Fontal, WINDBIOME dataset: Microbial species richness in tropospheric samples above the PBL confirms thousands of km long-distance transport of potential human pathogens. European Nucleotide Database. https://www.ebi.ac.uk/ena/browser/view/PRJEB64345. Deposited 14 July 2023.
92. X. Rodó *et al.*, microbial-richness-troposphere. Github. https://github.com/AirLabBcn/microbial-richness-troposphere. Deposited 8 August 2024.

PAPER V

Benchmarking Laser-Induced Fluorescence and Machine Learning for real-time identification of bacteria in bioaerosols

Alejandro Fontal[1, 2], Sílvia Borràs[1], Lídia Cañas[1, 2], Sofya Pozdniakova[1], and Xavier Rodó[1,3*]

[1]Climate and Health Group, Climate, Air pollution, Nature and Urban Health, Barcelona Institute for Global Health, Barcelona 08003, Spain;

[2]University of Barcelona (UB), Department of Microbiology, Faculty of Biology, 08028 Barcelona, Spain

[3]ICREA, Barcelona, Spain

Article in preparation

Benchmarking Laser-Induced Fluorescence and Machine Learning for real-time identification of bacteria in bioaerosols.

Alejandro Fontal[1, 2], Sílvia Borràs[1], Lídia Cañas[1, 2], Sofya Pozdniakova[1], and Xavier Rodó[1,3*]
[1]AIRLAB, Climate and Health (CLIMA) Group, ISGlobal, Barcelona, Spain
[2]University of Barcelona (UB), Department of Microbiology, Faculty of Biology, 08028 Barcelona, Spain
[3]ICREA, Barcelona, Spain

ABSTRACT

Microorganisms are ubiquitous in the environment, playing key roles in all ecosystems, including the atmosphere, with airborne dissemination via particulate matter being essential for many microorganisms' life cycles. However, the atmosphere as a microbial ecosystem has been severely understudied, mostly due to the challenging technical difficulties in sampling and characterizing it. So far, most recent studies use metagenomic sequencing to assess aerobiome diversity, which can be biased and hurdled due to the inherent ultra-low DNA yield of air samples.

Previous research has already demonstrated the possibility of using Laser-Induced Fluorescence (LIF) and machine learning (ML) to characterize the vegetal fraction of bioaerosols, by classifying pollen particles using the Rapid-E bioaerosol detector (Plair SA) and neural network classifiers.

In this study, we present a new methodology for near real-time automatic recognition of microbial particles in the air: by replacing Rapid-E's UV laser (337 nm) with another optimized (266 nm) to excite fluorophores in bacterial and fungal cell membranes. We tested this new setup with artificially generated aerosols enriched with five distinct bacterial species. Employing Random Forest classifiers, we were able to: (a) detect bacterial particles (96.74% class-balanced accuracy), and (b) discriminate between the different species (69.24% class-balanced accuracy across the different species in the validation set). This innovative approach sets a new range of possibilities for the rapid and precise monitoring of airborne microbial communities, offering a valuable tool for both ecological studies and public health surveillance.

1. Introduction

Microbial communities are essential components of all ecosystems, contributing significantly to ecological processes and human health. While extensive research has been conducted on the microbiome present in soil, water and different surfaces in both urban and rural environments, the atmospheric microbial ecosystem (aerobiome), remained relatively unexplored until more recently. The atmosphere is a dynamic environment where microorganisms can be transported over long distances, influencing both local and global ecological patterns. These airborne microorganisms play crucial roles in nutrient cycling, weather patterns, and even disease transmission (1–4). Despite their importance, the study of aerobiomes has been hampered by significant technical challenges (5).

Historically, methods for characterizing the airborne microbiome relied on culture-based approaches that missed a majority of microbial species, as most are non-culturable. The advent of metagenomic next-generation sequencing (mNGS) revolutionized microbial detection in both clinical and environmental samples, offering culture-independent analysis with higher throughput, faster turnaround times, and greater resolution. However, the ultralow biomass nature of air samples results in extremely low DNA yields. This presents significant challenges for both qualitative and especially quantitative characterization of microbial diversity. To accumulate enough material for analysis, the temporal resolution of samples often needs to be compromised (6,7).

Alternatively, a distinct set of fluorescence-based techniques has been developed for the detection, monitoring, and identification of bioaerosols. Initially driven by military efforts to detect biological agents, these technologies have expanded into various civil and scientific fields where timely bioaerosol surveillance is essential. In recent years, several instruments capable of characterizing bioaerosols (primarily focused on pollen) have been commercialized. Notable examples include the Wideband Integrated Bioaerosol Spectrometer (WIBS) series from Droplet Measurement Technologies (USA) (8), the KH-3000 from Yamatronics (Japan) (9), the BAA500 from Helmut Hund GmbH (Germany) (10), and the Rapid-E series from Plair SA (Switzerland) (11,12). Maya-Manzano *et al*. conducted a comprehensive comparison and validation of these instruments, among others, for detecting and classifying pollen, using Hirst-type traps as a ground truth reference (13).

While the automation of pollen particle counting and identification is valuable, expanding these capabilities to the microbial fraction of the bioaerosols is not trivial. The relatively large size of pollen makes it more suitable for analysis using spectroscopy, especially outside of controlled laboratory conditions. Studies have shown that UV-LIF is precise enough to discriminate even the smallest viral particles (picornaviruses) when used in combination with machine learning-trained classifiers (14), but translating this to a real-time field-deployable device is challenging. This is due to the need for high accuracy and the complexity of dealing with the heterogeneous mixtures in which microbial particles are typically found in real world scenarios.

In this study, we use the Rapid-E Real-Time Airborne Particle Analyzer, developed by Plair SA (Switzerland), to attempt, to the best of our knowledge, the first methodology to automate the detection and classification of bacterial particles in bioaerosols using a portable device suitable for real-time surveillance in both indoor and outdoor environments.

2. Methods

2.1 Particle Analyzer: Rapid-E

Commercial Device. Rapid-E is the successor to Plair's first particle analyzer, the PA-300, which uses morphological analysis based on scattered light and chemical analysis through UV-LIF (15, 16). The instrument is continuously aspirating ambient air and its aerosol particles. Once these are captured, they enter the nozzle which creates a thin laminar flow in the measurement chamber. Here, two laser beams hit each particle subsequently. First, the particle is hit by a red-light laser beam, whose light gets scattered and detected by 24 detectors positioned at different angles (from 45° to 130°) at a high acquisition rate and at the same wavelength as the red laser beam. The detection of the aerosol particles by these detectors generates a trigger signal for the UV laser (337 nm), which in turn hits the particle

exciting its fluorophores. The emission fluorescence spectra (if any) is then acquired by 32 detectors on the spectral range from 350 to 800 nm, and the spectral lifetime is continuously acquired for 4 different spectral bands for a total of 64 ns (at a 1 ns acquisition frequency). See Figure 1 for a schematic view of the working principle.

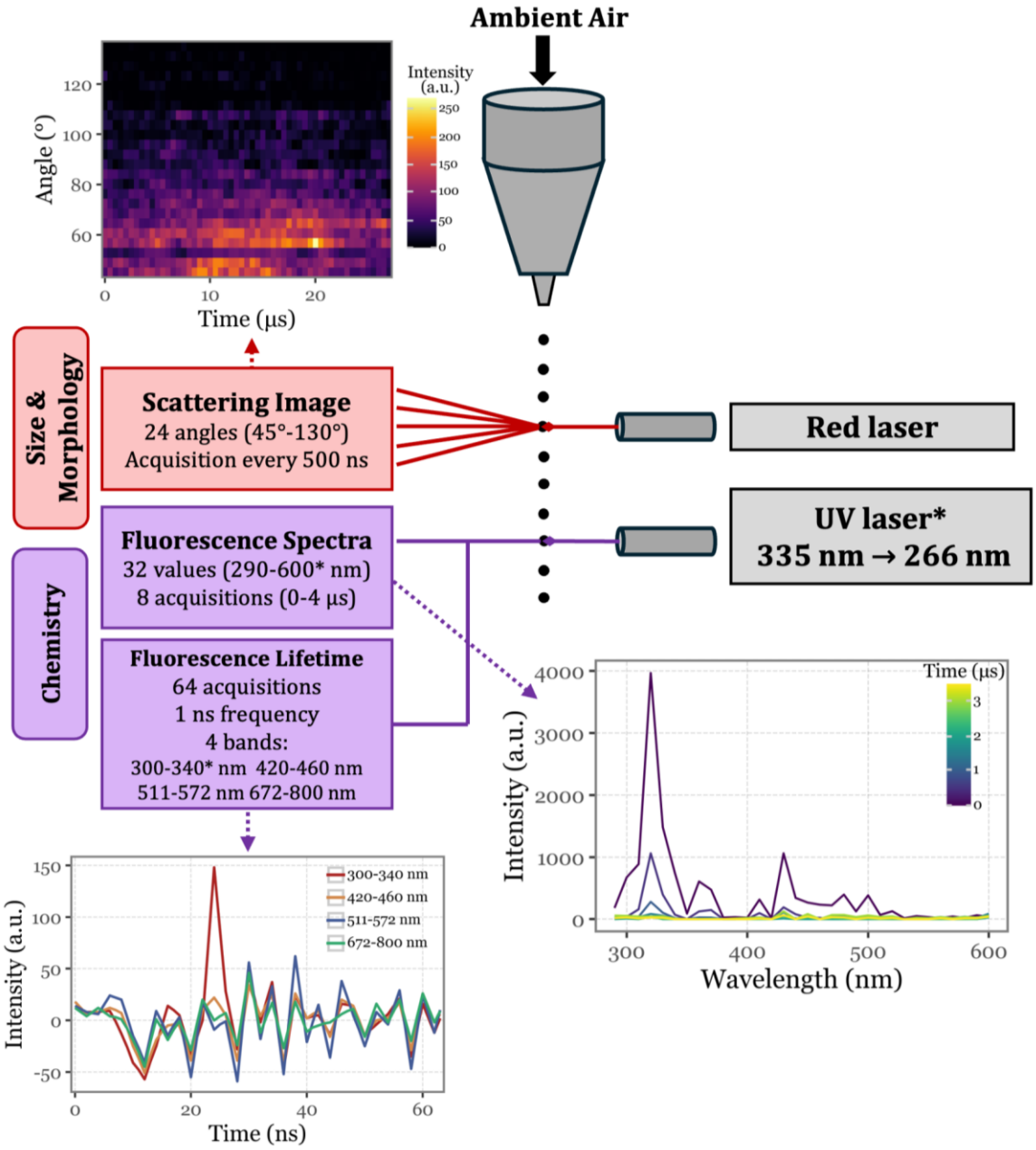


Figure 1. Schematic representation of Rapid-E's operational mechanism, illustrating the data points generated for each particle as it passes through the two integrated laser systems. The asterisks denote modifications made with respect to the commercial version of Rapid-E: (1) the UV laser's excitation wavelength changed from 337 nm to 266 nm, (2) the fluorescence spectra acquisition range was changed from 350-800 nm to 290-600 nm, and (3) the first lifetime channel was shifted from 350-400 nm to 300-340 nm.

Customized Device. For our objective of repurposing the device to focus on the microbial fraction of aerosols, several modifications to the commercially available Rapid-E device configuration were implemented in this study. These adjustments were specifically aimed at adapting to the smaller particle

size range of microorganisms compared to pollen, and accounting for the different fluorophore distribution present in microbial cell walls.

The original UV laser integrated into the Rapid-E device operates at a wavelength of 337 nm, selected as a compromise to effectively excite key vegetal fluorophores such as chlorophyll A and B, carotenoids, coenzymes, and various flavins, even though this wavelength does not correspond to the peak absorption of any specific fluorophore. However, to better discriminate the microbial fraction in bioaerosols, it is advantageous to minimize excitation of these plant fluorophores and instead focus on those that are more characteristic of microbial cell walls. These include amino acids (tryptophan, tyrosine, phenylalanine), flavins (riboflavin, FAD, FADH), and coenzymes (NADH, NADPH). These fluorophores typically have excitation peaks at sub-300 nm wavelengths or possess broad excitation spectra that include these shorter wavelengths, generating distinct emission spectra when excited below 300 nm (17). Consequently, we opted to replace the standard laser with a different one with a sub-300 nm primary wavelength.

The UV laser selected as a replacement was the Onda ns Q-Switched Diode-Pumped Solid State (DPSS) laser (Bright Solutions Srl, Italy), which operates at a primary wavelength of 266 nm, with a maximum peak power of 40 kW, a maximum pulse energy of 80 μJ, and a pulse width ranging from 2 to 6 ns. This laser was integrated directly into the Rapid-E's measurement chamber using mirrors and tubing, along with a new bottom module added to the standard Rapid-E configuration, increasing its height by 15 cm (see Figures 1 and 2 for the updated configuration). To accommodate the lower excitation wavelength and capture a narrower range of emission spectra, modifications were made to the detectors in both the fluorescence spectra and lifetime modules: the fluorescence spectra acquisition range was adjusted to 290-600 nm (previously 350-800 nm), and the first channel of the lifetime module now covers the 300-340 nm range (instead of 350-400 nm).

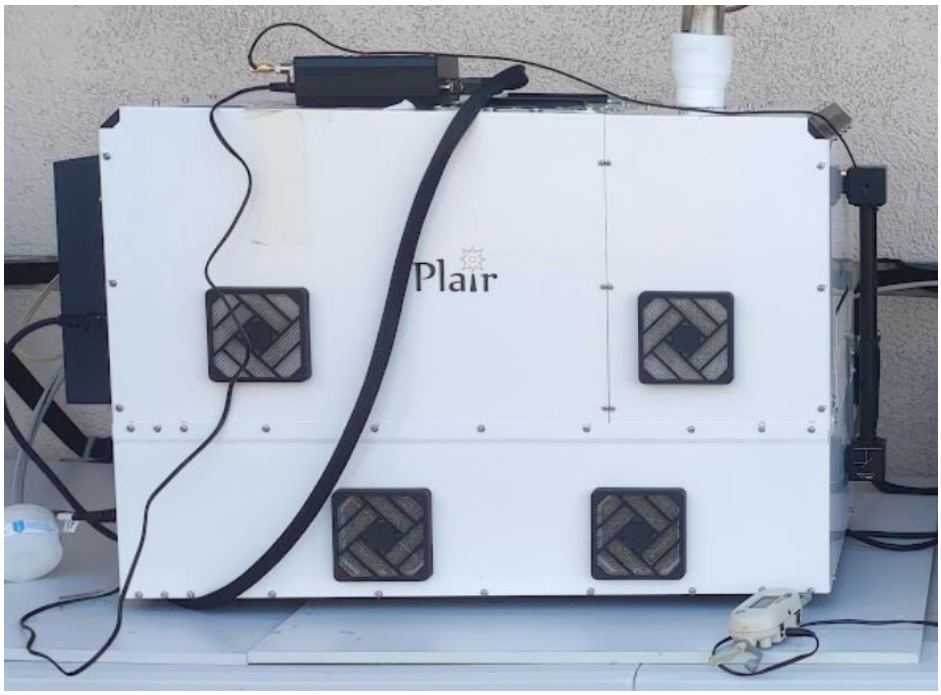


Figure 2. Appearance of the customized Rapid-E after integrating the new UV laser and its power supply.

2.2 Experimental set-up

After successfully implementing the modifications into the device, we designed a series of experiments to evaluate:

(1) the feasibility of generating aerosols from solutions and directly introducing them to Rapid-E's inlet;
(2) the effectiveness of the newly integrated laser in detecting and discriminating target fluorophores when aerosolized and isolated;
(3) the device's detection capability when aerosolizing and analyzing different isolated and cultured bacterial species.

Aerosolization. We use an AGK 2000 aerosol generator (Palas GmbH, Germany) to generate solid particles from suspensions. In all experiments depicted in the study, we maintained an air flow rate of approximately 3 L/min and a pressure between 0.7 and 1.1 bar, with direct connection to the Rapid-E inlet via a sealed metallic arm. Solutions to be aerosolized were introduced into the generator's 500 mL reservoir and stirred continuously at low RPM using a magnetic stirrer (Thermo Fisher Scientific, USA) to ensure homogeneity throughout the process.

Fluorophores. We obtained powdered forms of the following molecules, known to be components of bacterial cell walls and exhibiting fluorescent emission when excited at sub-300 nm wavelengths: NADH, riboflavin, tryptophan, and tyrosine. Solutions were prepared by dissolving 5 mg of each fluorophore in 150 mL of deionized water. For the riboflavin solution, 0.1M acetic acid was added to a second solution to enhance fluorescence. To ensure the system was free of obstructions, contaminants, or residual particles from previous samples, we aerosolized Milli-Q water for 30 minutes between each solution.

Bacteria. We analyzed five bacterial species commonly found in urban bioaerosols, which were obtained from air samples collected on quartz fiber filters using a high volume sampler (MCV, Spain) on the rooftop of our laboratory (Barcelona, Spain). Filters were cultured on agar plates and isolated colonies were identified by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (Bruker LT microflex MALDI-TOF MS, Bruker Daltonics, Germany). The complete taxonomic classification of the bacterial species used in the experiments is presented in Table 1.

Table 1. Taxonomic classification of the bacterial species aerosolized in the study.

Phylum	Class	Order	Family	Genus	Species
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Micrococcaceae*	*Micrococcus*	*luteus*
Firmicutes	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*	*endophyticus*
Firmicutes	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*	*cereus*
Actinobacteria	*Actinobacteria*	*Micrococcales*	*Micrococcaceae*	*Kocuria*	*salsicia*
Firmicutes	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*	*hominis*

For all the bacteria, a Ringer solution 1:4 was used as the medium, with the negative control consisting of Ringer solution 1:4 alone in the generator's reservoir at the time of aerosolization. Each bacterial isolate was aerosolized for at least 15 minutes, and to prevent cross-contamination, Milli-Q water was aerosolized for 30 minutes between each sample. Throughout the experiment, the external controller of the Onda ns QS-DPSS laser was set to 80% (36 A) and Rapid-E was configured in fine dust mode (0.5-100 µm).

2.3 Data analysis

Data Wrangling. After detection, 1-minute data files are stored on Rapid-E's internal hard drive. Based on the recorded timeline for each sample, we extracted the last 10 minutes of aerosolization data for each fluorophore and bacterial sample, ensuring a buffer period between samples to prevent any potential interference. Given the anticipated irregularity in aerosol particle formation, the uncertainty regarding the proportion of particles containing the targeted biological content, and the need for clear signals to distinguish subtle differences in chemical or morphological composition, we filtered the data to include only particles with recorded peak fluorescence intensities exceeding 2000 a.u. at any wavelength and time point. This threshold is higher than the 1500 a.u. used by Šaulienė *et al.* for pollen recognition models using the Rapid-E (11), as our laser and spectral conditions differ, and we observed significant background noise in peaks below that threshold.

No transformation is needed for incorporating fluorescence spectra and lifetime data into the models. However, light scattering images present a challenge due to their irregular shapes, as the total number of acquisitions depends on the duration of the detected scattered light signal. While most particles exhibit signals lasting just 25-30 µs, a significant number of particles require over 100 µs of scattering signal. Expanding all samples to match the dimensionality of these outliers would drastically increase dimensionality, severely hindering model performance. To address this, we observed that the scattering signal peaks consistently around 10 to 13 µs for all samples, and that the 80th to 90th percentiles of the total scattering signal duration fall within the 25 to 35 µs range. Based on these findings, we decided to crop the scattering images at 30 µs (equivalent to 60 acquisitions), zero-pad shorter signals, and normalize the values to the [0-1] range. Figure 3 showcases the statistics discussed, with an example of the final transformed scattered images shown in Suppl. Figure 1.

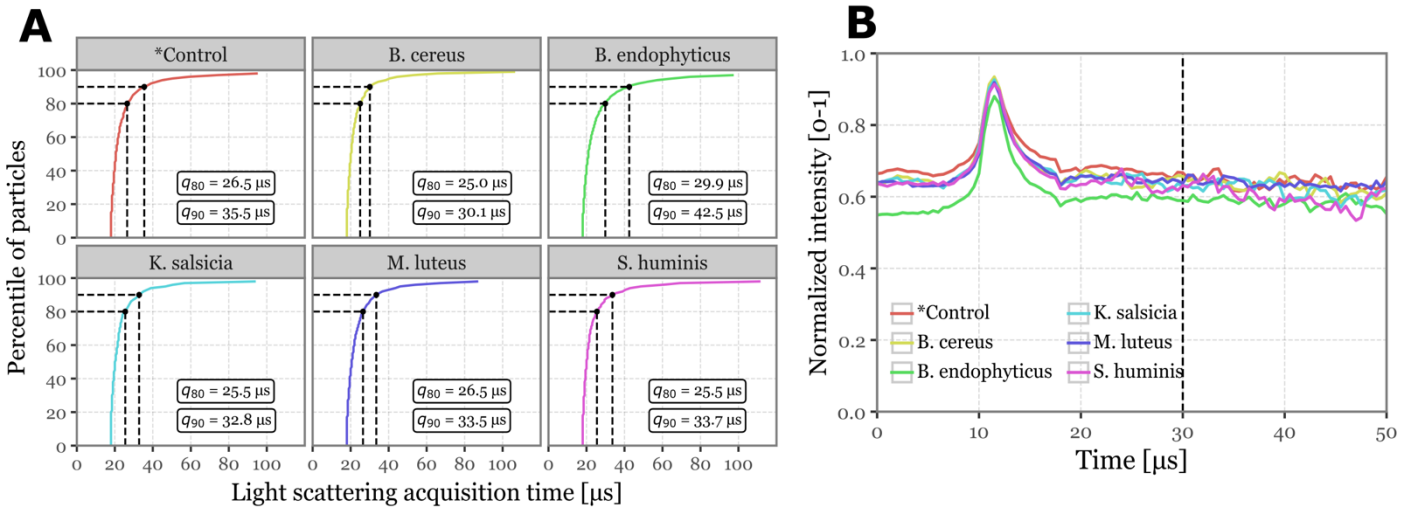


Figure 3. Light scattering intensity acquisition statistics. Panel A (left) displays the cumulative distribution functions for the total time of light scattering acquisition per sample for each of the sample groups. Dashed lines and the text boxes indicate the values for the 80th and 90th percentiles. Panel B (right) shows the median normalized intensity across all scattering angles as a function of acquisition time for each of the sample groups. The vertical dashed line marks the selected threshold of 30 μs.

Random Forest Classifier. For bacterial identification and classification, we trained tree-based machine learning models, specifically Random Forests (18), to perform two tasks: (a) binary classification to distinguish between control and bacterial samples, and (b) multiclass classification to identify specific bacterial species. Hyperparameter tuning involved adjusting the number of trees (estimators), the maximum depth of the trees, and the predictors to include in the model (all combinations of fluorescence spectra, lifetimes and scattering images), optimizing for test set accuracy before finalizing the model. We split the particles of each group into training, test, and validation sets, representing 60%, 20%, and 20% of the total particles, respectively. The training set is used for model training, i.e. the only portion of the data that the models 'see'. The test set is used to evaluate model performance during hyperparameter tuning, and the validation set is employed for the final model evaluation. A complete diagram of the data flow, model training, and evaluation process is provided in Figure 4.

Given the class imbalance present in both the binary classification and the bacterial species classification tasks, model evaluation was conducted using class-balanced accuracy, calculated as the average of the recall metric (True Positives / True Positives + False Negatives) for each class. To address class imbalance during training, we applied class weight adjustments inversely proportional to the frequency of each class, minimizing potential biases. The implementation was carried out using the RandomForestClassifier class from the scikit-learn Python library (19).

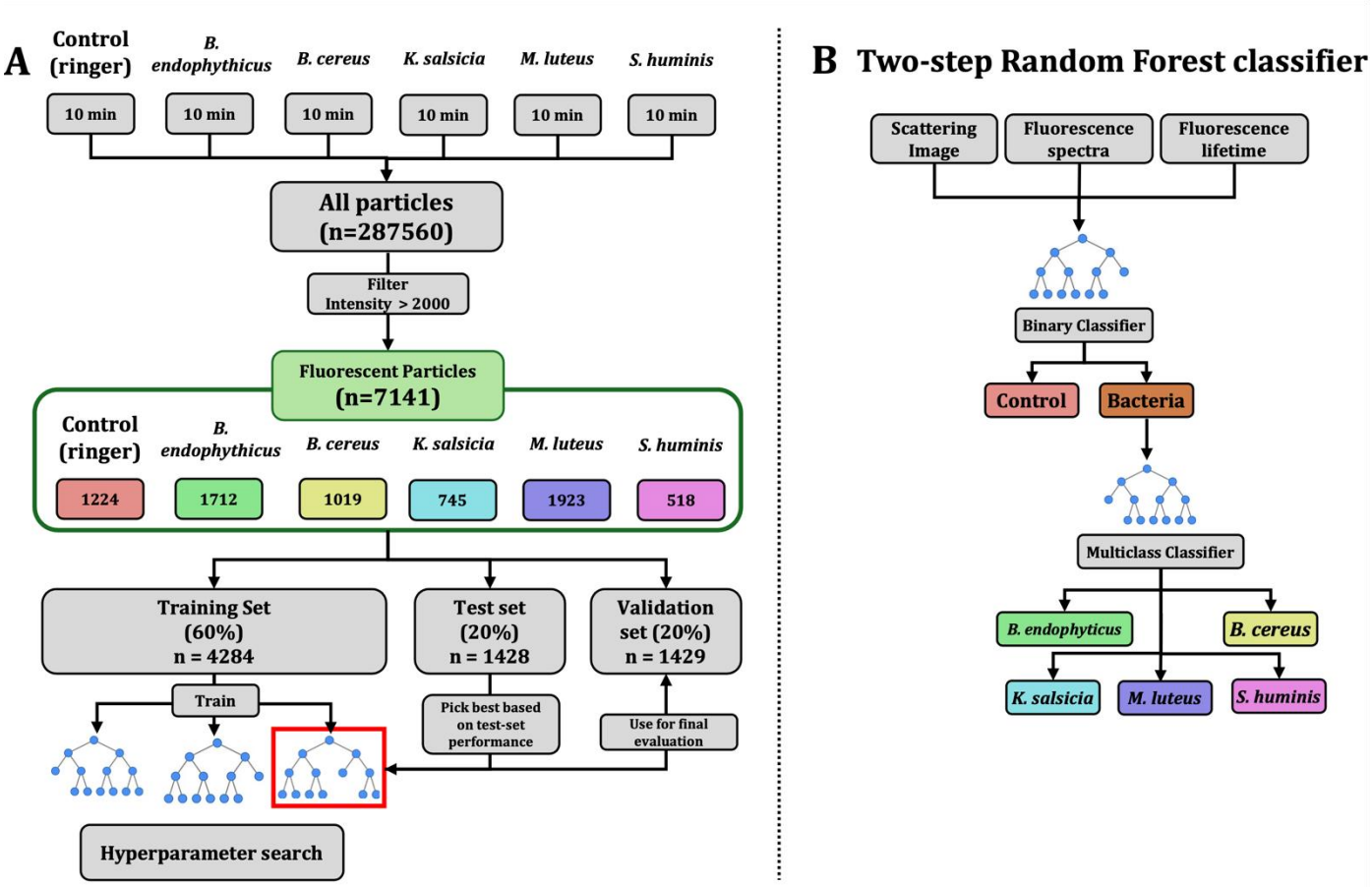


Figure 4. Data flow scheme and model hierarchy. Panel A (left) illustrates how the experimental data is handled, filtered and used to train, evaluate and ultimately validate the models. Panel B (right) visually describes the two-step approach employed for classifying aerosol particles using random forests: the first step involves a binary classifier to determine whether a particle is bacterial, and the second step employs a multiclass classifier to discriminate between bacterial species if the particle is identified as bacterial by the previous model.

3. Results

3.1 Fluorophores

The aerosolization of fluorophores commonly found in biological particles produced a range of aerosol particles detectable by the Rapid-E device, with varying success rates. The device detected the following particle counts per minute of aerosolization: 4064 for tyrosine, 3132 for riboflavin in acetic acid, 2162 for tryptophan, 1885 for NADH, and 853 for riboflavin in deionized water. Regarding fluorescence intensity, most particles did not exhibit sufficient intensity to be classified as fluorescent (2000 a.u. threshold), with median peak intensities below 800 a.u. for all fluorophores. See Figure 5 for the median, 95th and 99th percentile values of fluorescence intensity for each class.

Using the fluorescence threshold of 2000 a.u., the percentage of fluorescent particles detected was 10.88% for riboflavin in acetic acid, 4.90% for tryptophan, 2.46% for riboflavin in deionized water, 1.59% for NADH, and 1.03% for tyrosine (Figure 5). These results suggest that metrics focusing on the central part of the particle distribution are likely to be ineffective in distinguishing between groups. Most

particles generated in the aerosolization chamber either do not adequately contain the fluorophores or are not effectively excited by the laser, likely due to their small size.

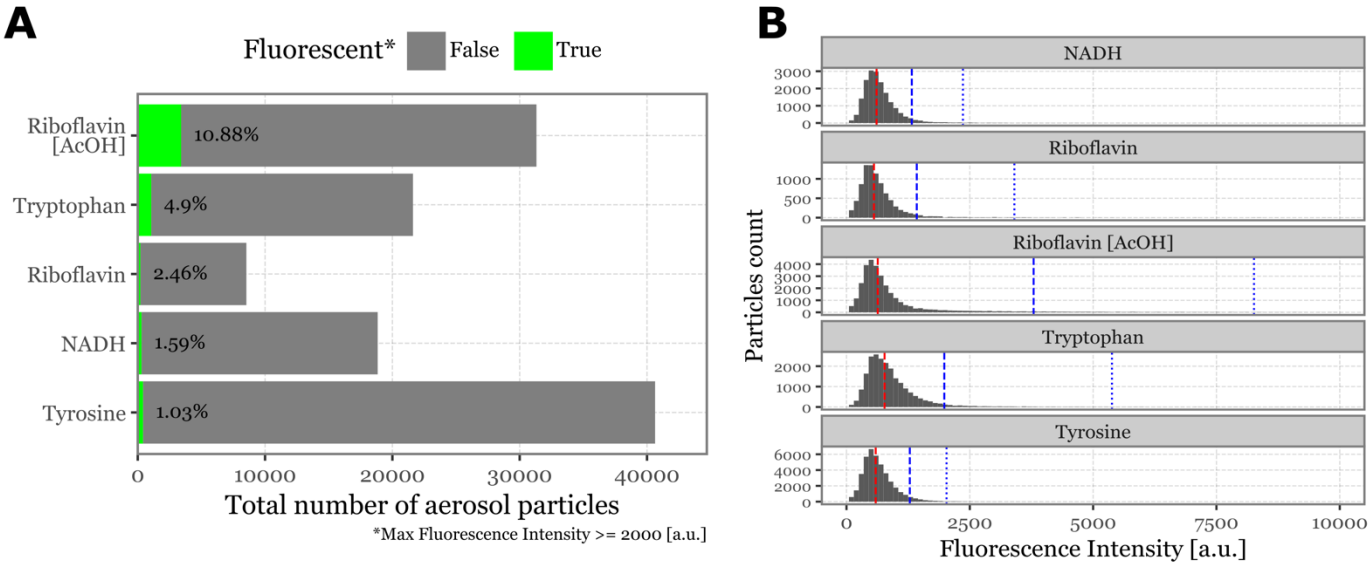


Figure 5. Particle count and fluorescence intensity distribution for biological fluorophores. Panel A (left) displays a bar plot showing the total number of aerosol particles generated and detected by Rapid-E over a 10-minute period. The green bars (and text) represent the fraction of particles detected as fluorescent, based on a threshold of 2000 a.u. for peak intensity at any given time or wavelength. Panel B (right) presents histograms illustrating the intensity distribution of all detected particles during the aerosolization of each fluorophore. The red line indicates the median intensity for each group, while the blue dashed line and blue dotted line represent the 95th and 99th percentiles, respectively.

Fluorescence spectra for the different fluorophores are presented in Figure 6. To ensure a clean signal, we focused on the top 100 particles with the highest peak fluorescence intensity for each fluorophore. The median spectra and interquartile ranges are shown in Figure 6A, while the spectra for each of the top 100 particles are displayed in the heatmap in Figure 6C. Figure 6B provides a reference for the expected fluorescence spectra, adapted from (17), where a single-shot 266 nm laser was used with a higher wavelength resolution to excite particles of the same fluorophores at a controlled diameter of 5 µm. The relative order of the fluorescence peaks we observed — tyrosine, tryptophan, NADH, and riboflavin — aligns with the reference data, validating our results. However, we observed a shift towards lower wavelengths: tyrosine peaks between 300-320 nm (close to reference values), tryptophan peaks at 330 nm (versus 350 nm in the reference), NADH at 390 nm (versus 450 nm), and riboflavin at 460 nm (versus 550-560 nm). This suggests a potential misalignment in the wavelength calibration of the Rapid-E's fluorescence acquisition system, possibly extending beyond the specified 600 nm range.

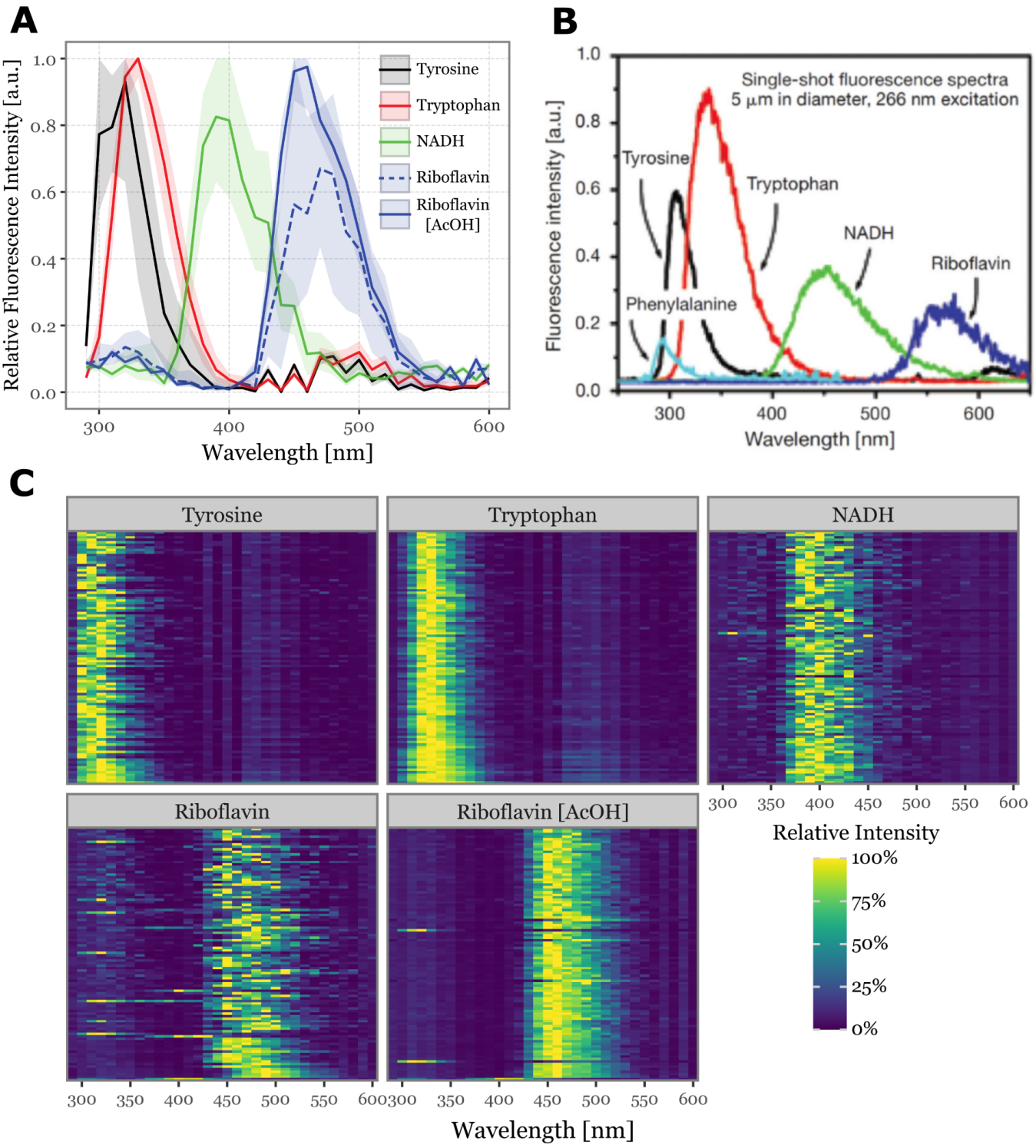


Figure 6. Fluorescence spectra detected for aerosolized biological fluorophores. Panel A (top-left) shows the median fluorescence spectra, aggregated across all acquisitions, for the 100 most fluorescent particles during each fluorophore's sampling period. The shaded areas represent the interquartile range (25th to 75th percentiles). Panel B (top-right) displays single-particle fluorescence spectra excited at 266 nm for common fluorophores found in biological particles, adapted from (17). The colors in Panels A and B are matched to facilitate comparison. Panel C (bottom) presents a heatmap of the fluorescence spectra for the 100 most fluorescent particles from each fluorophore, ordered from top to bottom.

The particles with the highest fluorescence intensities also produced the cleanest spectral signals. Riboflavin in acetic acid and tryptophan exhibited low variability among particles, in contrast to tyrosine, NADH, and riboflavin in water for which the particle-to-particle variability was detectably higher (Figure 6C).

The fluorescence lifetime values also produced clearly distinguishable signals for all fluorophores, showing fluorescence intensity peaks at around 20 ns after excitation in a single wavelength band each: tyrosine and tryptophan peak in the 300-340 nm band, NADH in the 420-460 nm band and riboflavin in the 672-800 nm band (Suppl. Figure 4) Again, these values appear to be shifted and do not match the observed fluorescence peaks for the spectra, especially for riboflavin.

These results prove that, while the absolute spectral values may be shifted, the integration of the 266 nm laser into the Rapid-E successfully generates distinct and recognizable fluorescence spectra and lifetime. Furthermore, the aerosolization protocol produces solid particles consistent enough to be effectively detected by the device, which validates both the laser's performance and the reliability of the particle aerosolization process.

3.2 Bacteria

Aerosolization and Descriptive Statistics. The aerosolization of the bacterial cultures was successful, generating a substantial number of detectable particles. However, only a small fraction of these particles met the fluorescence threshold of 2000 a.u. set for the study. The particle counter of Rapid-E approached near-saturation levels (5000 particles/minute) during the aerosolization of all bacterial samples, including the negative control with Ringer solution 1:4, resulting in nearly 50,000 particles per group over the 10-minute sampling period (see Table 2 for the specific numbers).

Table 2. Summary statistics for the particles of each of the aerosolized bacterial groups. Fluorescent here implies at least a single reading at any wavelength and acquisition time of at least 2000 a.u, and the particle diameters are estimated using the Mie equations based on the scattering images.

	Total (n)	Fluorescent (%)	Fluorescent (n)	Diameter (median, µm)	F.I. (median)	F.I. (Q95)	F.I. (Q99)
Control (ringer)	47922	2.55%	1224	4.88	985	1806	2284
B. cereus	47935	2.13%	1019	4.97	1022	1761	2192
B. endophyticus	47916	3.57%	1712	4.14	1117	1904	2354
K. salsicia	47936	1.55%	745	4.95	970	1681	2112
M. luteus	47923	4.01%	1923	4.72	1136	1938	2392
S. huminis	47928	1.08%	518	4.81	917	1601	2021

Despite this, only 1-4% of the total detected particles exhibited fluorescence above the threshold, leading to a total of 7141 particles included in the final models, with a non-homogeneous distribution: 1923 particles during *M. luteus* aerosolization, 1712 for *B. endophyticus*, 1224 for the negative control, 1019 for *B. cereus*, 745 for *K. salsicia*, and only 518 for *S. huminis*. The particle size distributions were consistent across groups, with most particles falling within the 4-5 µm diameter range. Notably, *B. endophyticus* particles appeared to have a slightly smaller size, ranging from approximately 3.5 to 4.75 µm in diameter. See Suppl. Figure 2 for the full distribution.

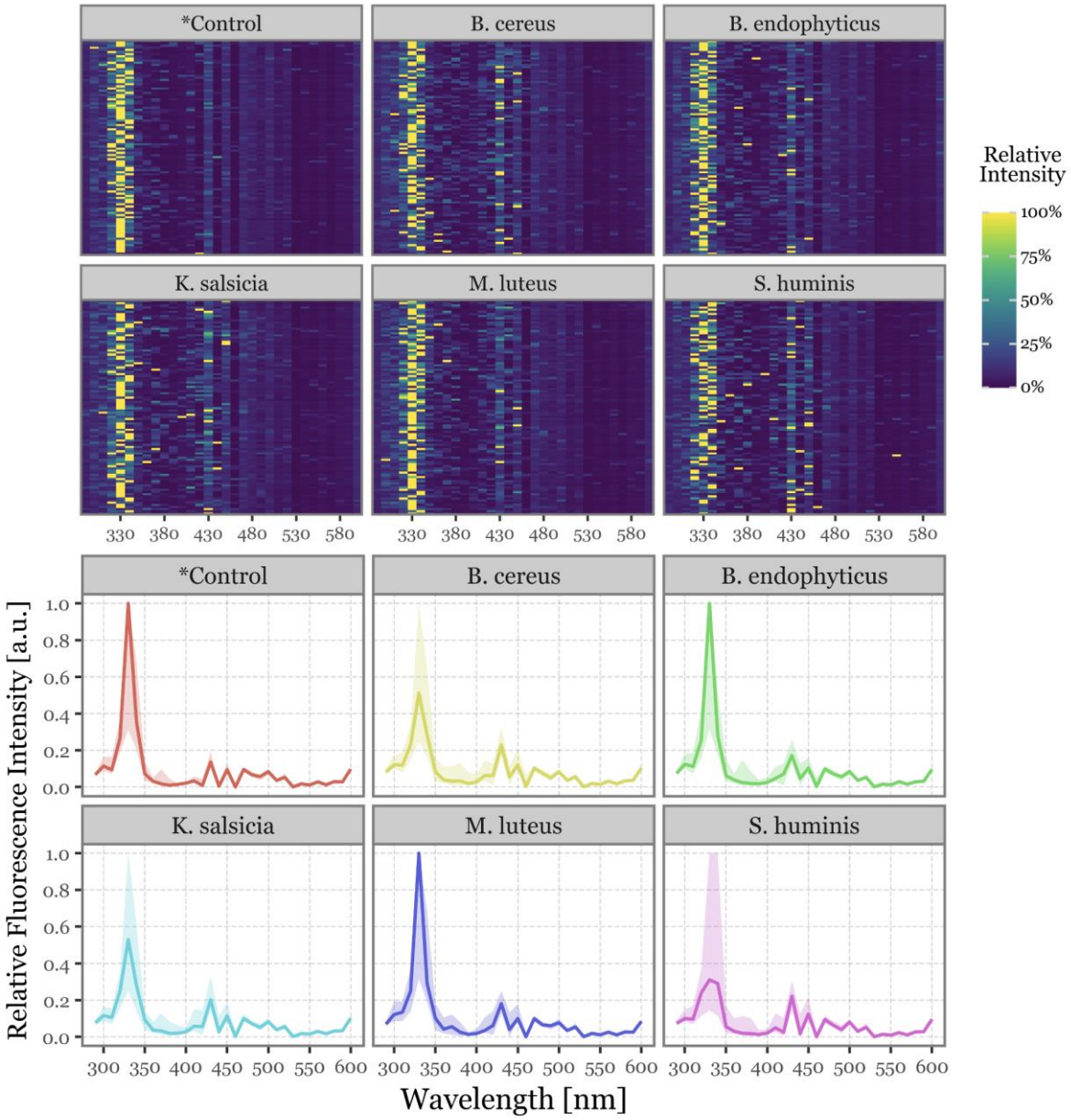


Figure 7. Fluorescence spectra for the aerosolized bacterial samples. In the top panel, heatmaps displaying the relative fluorescence intensity detected across all wavelengths within the detector range for the top 100 most intensely fluorescent particles. The bottom panel shows the same data represented as line plots, with the solid line indicating the median value across all samples and the shaded area representing the interquartile range (25^{th} to 75^{th} percentiles).

The signals for the fluorescence spectra, lifetimes, and scattering images were not as distinctly separable across groups compared to the aerosolized fluorophore particles discussed in the previous section, as expected due to the complex biological composition of the bacterial cells and media.

The fluorescence spectra for each group, as shown in Figure 7, are presented both for the top 100 most fluorescent particles of each group and as an aggregated distribution of all particles within each group. All groups exhibited a prominent peak around 330 nm, accompanied by a secondary, less pronounced peak in the 400 to 470 nm region. No group-specific spectral features were identified.

The fluorescence lifetimes measured over the 64 ns post-excitation were notably less intense and exhibited higher noise levels compared to the fluorophores (see Suppl. Figure 6 for the bacteria and Suppl. Figure 4 for the fluorophores). Nevertheless, the median fluorescence lifetime profiles indicate some variability across groups, particularly in the 300-340 nm and 511-572 nm bands. The extent to which this variability can facilitate group discrimination remains unclear, as the median values may not fully capture the subtle differences that machine learning models could potentially exploit.

The scattering images (Suppl. Figure 5) are largely similar, but showed some shifts in distribution, intensity and total length of signal as we observed earlier in Figure 3. Specifically, *B. endophyticus* particles demonstrated lower intensity and longer scattering signals.

To better visualize the differences among the groups, a principal component analysis (PCA) was conducted independently for each of the three data sources (fluorescence lifetime, fluorescence spectra and light scattering), with scatterplots projecting the first two principal components (Figure 8). The scatterplots show how the bioaerosol particles cluster across groups when projected in 2D space: while all three variables exhibit some degree of overlap among the bacterial groups, the scattering images and especially the fluorescence lifetimes display distinct patterns in the distribution of particles across the first two components, suggesting potential exploitable differences for classification.

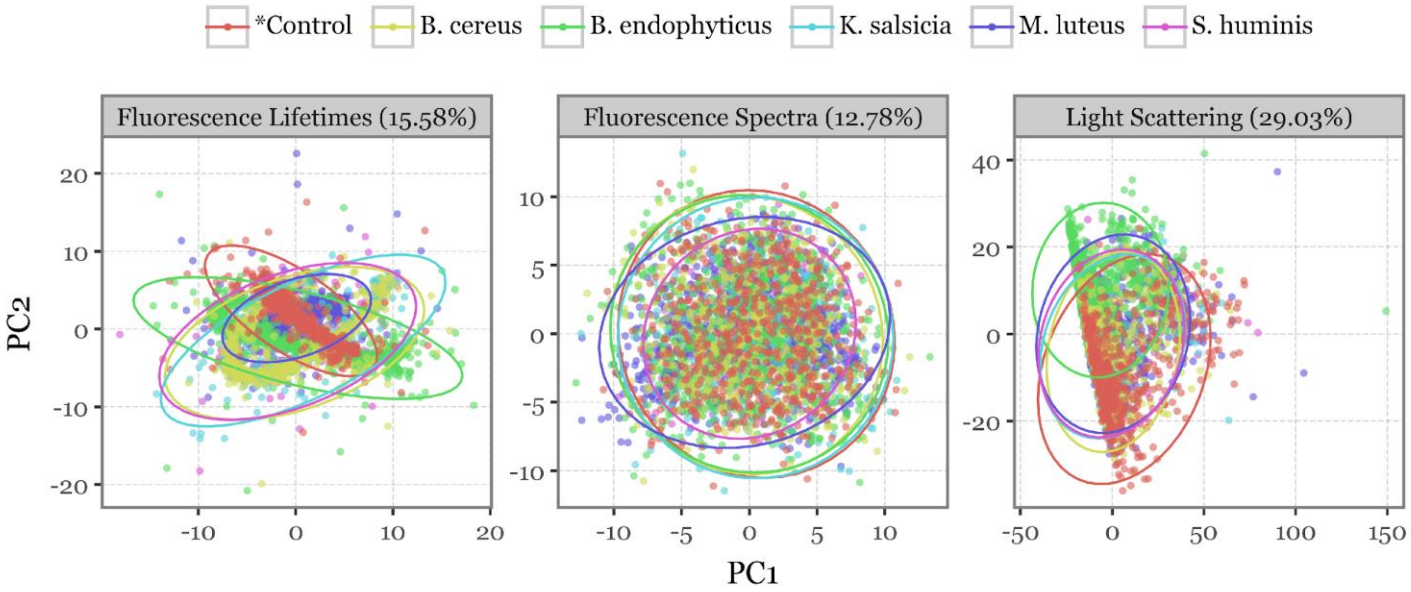


Figure 8. PCA projection of all particles. Scatterplots depicting the first two principal components after a principal components analysis (PCA) decomposition of all variables describing each particle (fluorescence spectra, lifetime, and scattering images) used as input into the models. Colours indicate the bacterial group of each particle,

with ellipses representing the 95% confidence interval of a multivariate T-distribution for the particles in each group. The percentages above indicate the total variance explained by the first two components.

Model Training and Performance. The performance of the different models during hyperparameter exploration is shown in Figure 9, Figure 10 and Suppl. Figure 7. All models, across all combinations of predictor variables and parameter choices, showed a marked improvement over the naïve baseline model in both the binary and multiclass classification tasks. The choice of predictor variables emerged as a key factor, not only contributing to the variance in model performance but also highlighting the critical features necessary for distinguishing between bacterial groups. When evaluating the best-performing models for each combination of predictor variables (Figure 9), those that included fluorescence lifetimes (LT) achieved the highest balanced accuracy, with 95-97% on the test set for the binary classifier and 65-69% for the multiclass classifier. Conversely, model performance dropped significantly when fluorescence lifetimes were excluded: models using only fluorescence spectra (FS), and scattering images (SI) achieved a maximum of 73-80% accuracy in the binary classifier and 44-51% in the multiclass classifier.

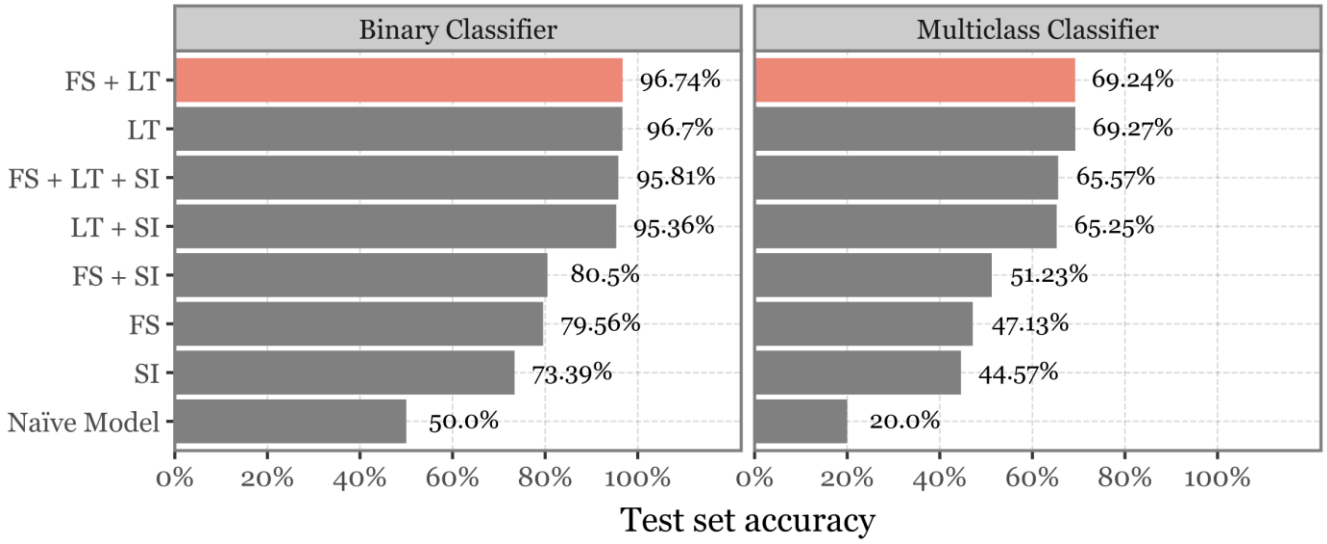


Figure 9. Evaluation of the predictor variables and model skill. Bar plot depicting the class-balanced accuracy of the best-performing model for each combination of model predictors (FS = Fluorescence Spectra, LT = Lifetimes, SI = Scattering Images) for both binary (left) and multiclass (right) classifiers. The bar corresponding to the overall best-performing model, which uses both Fluorescence Spectra and Lifetimes, is highlighted in red. The baseline performance of a naïve classifier, equivalent to a completely uninformed prediction is also included for reference.

Models that included both FS and SI, in addition to LT, showed only marginal improvements, with the best-performing model achieving 96.74% in binary and 69.24% in multiclass classification, suggesting that LT alone captures most of the relevant information. FS alone provided a reasonable accuracy of 79.56% in binary classification but showed a noticeable drop to 47.13% for multiclass classification. SI were the least effective when used in isolation, achieving 73.39% in binary classification and 44.57% in multiclass classification. These observations suggest that while combining multiple predictors can improve accuracy, LT remains the most informative feature for this classification task. The best overall model, which was selected for evaluation on the validation set, was the FS + LT model. However, this model only marginally improved upon the LT-only model in binary classification (96.74% vs. 96.7%) and showed slightly lower performance in multiclass classification (69.24% vs. 69.27%).

Figure 10 illustrates the class-balanced accuracies for both the train and test sets across various combinations of the number of estimator trees and the depth of decision trees for the FS + LT model. The results indicate that, while the training set consistently achieved near-perfect accuracy across all parameter combinations, the test set accuracies exhibited a more nuanced pattern, with a modest decline in performance for the binary classifier and a more pronounced drop (close to 25 percentage points) in the multiclass classification. The highest accuracy for the binary classifier (96.70%) was achieved with 125 estimator trees and a tree depth of 5, whereas the optimal performance for the multiclass classifier (69.24%) was observed with 275 estimator trees and a tree depth of 11. The observed disparity in performance between the train and test sets in the multiclass scenario highlights the challenges posed by the high dimensionality of the predictor space, where the model may capture subtle noise patterns rather than genuine signals. This likely explains why a higher number of estimator trees appears to benefit the test set performance by reducing the model's variance, thereby mitigating overfitting. Although some degree of overfitting is evident, the model still demonstrates strong generalization capability.

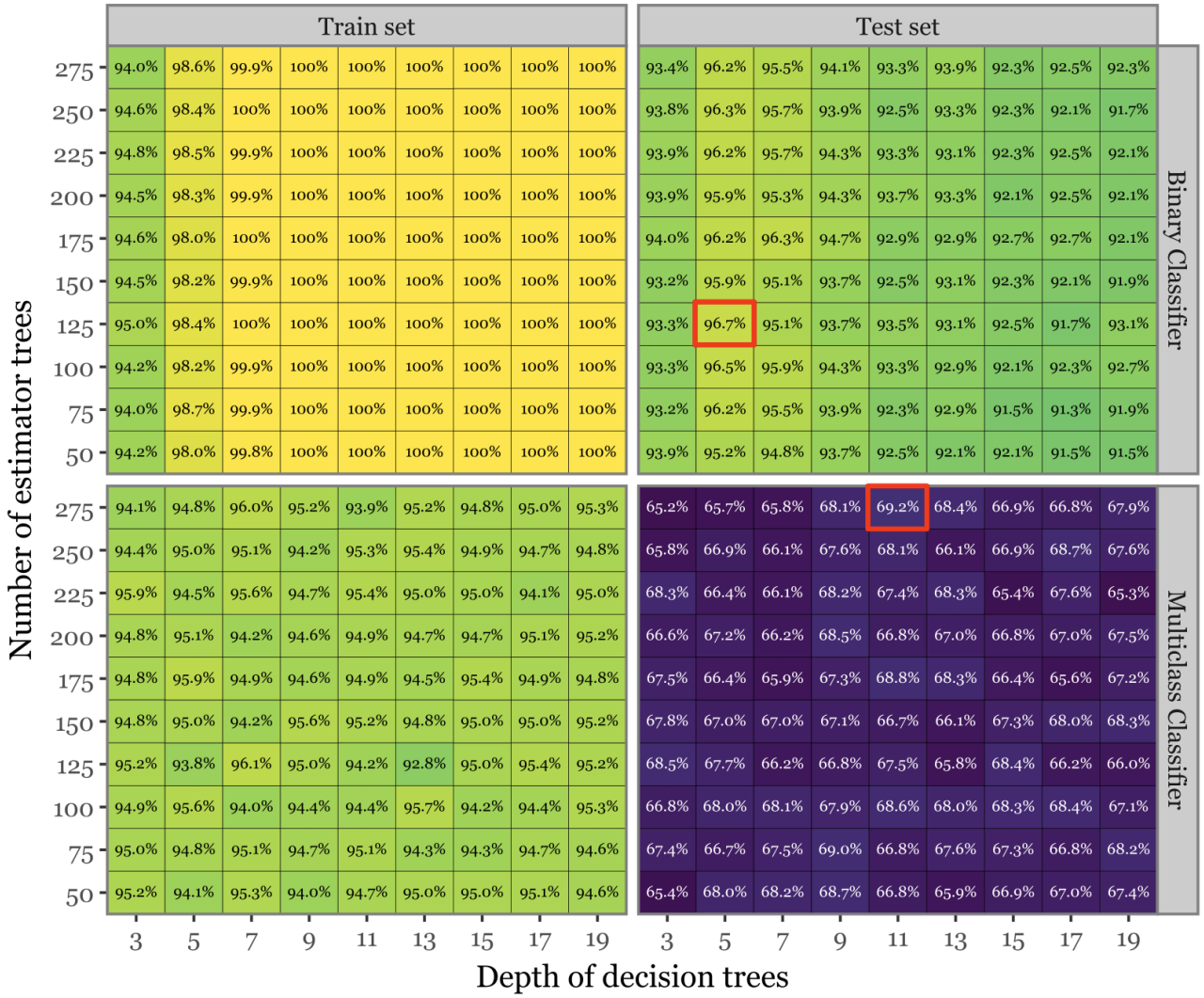


Figure 10. Hyperparameter search and model evaluation summary. Heatmap showing the class-balanced accuracy for both the train and test sets across both model types, focusing solely on the FS + LT models. The squares marked in red indicate the combination of parameters that yielded the best results.

In the final evaluation on the unseen validation set, the binary classifier exhibited strong performance, with a precision of 99.2% and a recall of 92.6% for the control class, and a precision of 98.3% and a recall of 99.8% for the bacterial class (Figure 11A). The multiclass classifier, while demonstrating high precision (90.9% to 99.2%) and recall (92.6% to 98.9%) for certain classes like *B. endophyticus* and *M. luteus*, showed significant variability across other classes. Specifically, *S. huminis* and *K. salsicia* were less reliably identified, with precision dropping to around 51-53% and recall as low as 15.2% for *S. huminis* (Figure 11B). These worst-performing classes also had lower overall abundance, and despite oversampling them during the training process to compensate, their performance remained suboptimal. This could be attributed to the fact that their lower abundance is due to fewer particles in those groups being fluorescent, indicating that their signals might be less distinct or produce a lower signal-to-noise ratio. Notably, the good performance of both models on the validation set is unbiased and reflects the model's true generalization capability, as the validation set was not used during training or hyperparameter tuning.

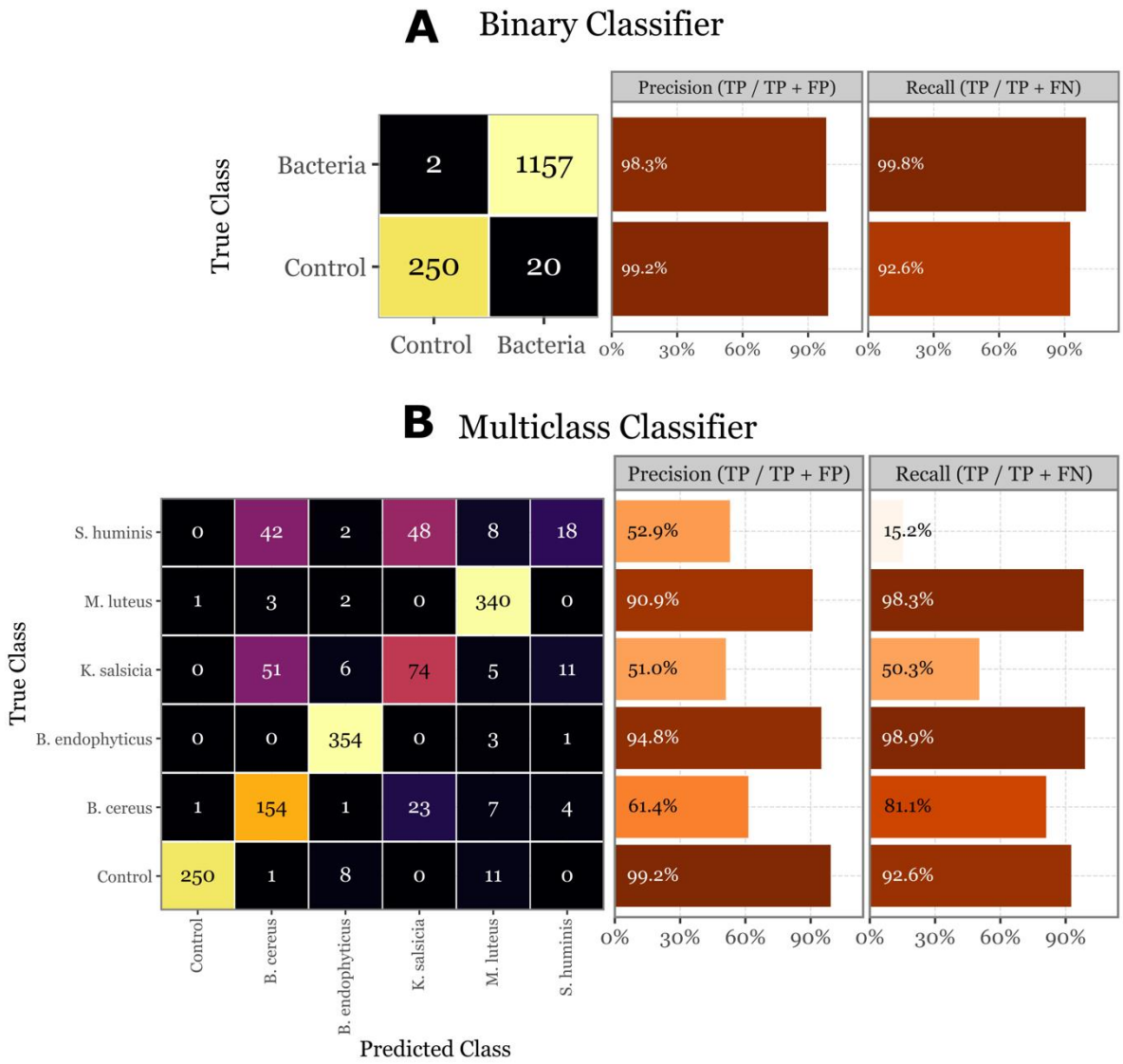


Figure 11. Evaluation of the skill of the Random Forest classifier on the validation set. On A (top panel) the confusion matrix of the validation set for the binary classifier, with the class-specific precision and recall metrics on the right side bar plots. On B (bottom panel) the same but for the multi-class classifier. Control class

added for comparison and to show the performance on the 20 control samples that the binary classifier predicted as Bacteria.

4. Discussion

The experimental results presented in this study highlight a novel application of a Rapid-E bioaerosol analyzer modified with a 266 nm laser for the detection and classification of bacterial particles in aerosols. The experimental findings demonstrate that the system is effectively integrated, with a clear ability to distinguish between fluorophores commonly found in bacterial cells when isolated. Moreover, the system successfully differentiates between aerosols enriched with bacteria and those without and even performs much better than baseline models in discriminating between different bacterial species. This positions the system as a promising tool for real-time bioaerosol monitoring. Despite this, the transition from controlled laboratory settings to real-world environmental sampling introduces several challenges that would need to be addressed before this technology can be widely implemented.

On a first note, it seems that the scattering images contribute minimally to the predictive performance of the models, with the integration of this data into the model failing to improve accuracy and, in some cases, even diminishing it. This could indicate that the morphological information contained in the scattering patterns of bioaerosols, despite the system's high-resolution capabilities, does not provide additional discriminatory power for distinguishing bacterial species or differentiating bacterial from non-bacterial particles when compared to fluorescence spectra and lifetime data. Alternatively, this may reflect limitations in our current feature engineering approach or the suitability of random forest classifiers for extracting meaningful features from scattering data. Random forests and tree-based models generally outperform more complex machine learning models such as Neural Networks when dealing with tabular data (20), but may fall short when dealing with complex spatial patterns inherent in scattering images, which bear resemblance to image data. More sophisticated machine learning techniques, such as Convolutional Neural Networks (21) may be better suited for processing this type of data due to their ability to automatically extract spatial hierarchies and features. Previous studies have successfully employed convolutional blocks for feature extraction from scattering images, achieving high accuracy in classifying pollen particles (11,12). However, it is important to note that pollen particles are significantly larger and more morphologically distinct than bacterial cells, potentially making morphological differences more detectable and relevant for classification in those contexts.

While the two-step model demonstrated a substantial degree of generalization, this capability remains inherently tied to the specific conditions of the training data - namely, the artificially generated bioaerosols. Even though the bacterial samples aerosolized in Ringer solution 1:4 represent a more complex mixture than standard lab-controlled samples, they still fall short of capturing the full complexity and heterogeneity of real environmental aerosols. Natural aerosols comprise a diverse array of biological and non-biological particles, such as pollen, fungal spores, plant debris, dust, and various pollutants, all of which are capable of scattering light and emitting fluorescence to varying extents (22–24). In such environments, the spectral and morphological signatures of bacterial particles may be obscured or mimicked by other particles, leading to potential misclassification or reduced detection sensitivity. To address this challenge, future work should focus on developing methods to generate training datasets that more accurately reflect the variability and complexity of real-world bioaerosols. However, this is not a trivial task, given the inherent variability and the dynamic nature of environmental aerosols.

5. Conclusion and Outlook

In conclusion, while challenges remain, the results of this study highlight the potential of the Rapid-E particle analyzer, with a UV laser replacement (337 nm to 266 nm), as a valuable tool for bioaerosol monitoring beyond pollen. With continued research and development, there is a clear pathway toward overcoming the current limitations and realizing the full potential of this technology in diverse real-world applications. The promising results obtained in this study suggest that with further refinement, this approach could become an integral part of environmental monitoring and public health strategies, offering a real-time solution for the detection and identification of airborne bacterial threats.

6. References

1. Morris CE, Sands DC, Bardin M, Jaenicke R, Vogel B, Leyronas C, et al. Microbiology and atmospheric processes: research challenges concerning the impact of airborne micro-organisms on the atmosphere and climate. Biogeosciences. 2011 Jan 3;8(1):17–25.

2. Griffin DW. Atmospheric Movement of Microorganisms in Clouds of Desert Dust and Implications for Human Health. Clin Microbiol Rev. 2007 Jul;20(3):459–77.

3. Tellier R, Li Y, Cowling BJ, Tang JW. Recognition of aerosol transmission of infectious agents: a commentary. BMC Infect Dis. 2019 Jan 31;19(1):101.

4. Fröhlich-Nowoisky J, Kampf CJ, Weber B, Huffman JA, Pöhlker C, Andreae MO, et al. Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. Atmospheric Res. 2016 Dec 15;182:346–76.

5. Behzad H, Gojobori T, Mineta K. Challenges and Opportunities of Airborne Metagenomics. Genome Biol Evol. 2015 May 1;7(5):1216–26.

6. Brodie EL, DeSantis TZ, Parker JPM, Zubietta IX, Piceno YM, Andersen GL. Urban aerosols harbor diverse and dynamic bacterial populations. Proc Natl Acad Sci. 2007 Jan 2;104(1):299–304.

7. Gusareva ES, Acerbi E, Lau KJX, Luhung I, Premkrishnan BNV, Kolundžija S, et al. Microbial communities in the tropical air ecosystem follow a precise diel cycle. Proc Natl Acad Sci. 2019 Nov 12;116(46):23299–308.

8. Healy DA, Huffman JA, O'Connor DJ, Pöhlker C, Pöschl U, Sodeau JR. Ambient measurements of biological aerosol particles near Killarney, Ireland: a comparison between real-time fluorescence and microscopy techniques. Atmospheric Chem Phys. 2014 Aug 13;14(15):8055–69.

9. Miki K, Fujita T, Sahashi N. Development and application of a method to classify airborne pollen taxa concentration using light scattering data. Sci Rep. 2021 Nov 16;11(1):22371.

10. Oteros J, Weber A, Kutzora S, Rojo J, Heinze S, Herr C, et al. An operational robotic pollen monitoring network based on automatic image recognition. Environ Res. 2020 Dec 1;191:110031.

11. Šaulienė I, Šukienė L, Daunys G, Valiulis G, Vaitkevičius L, Matavulj P, et al. Automatic pollen recognition with the Rapid-E particle counter: the first-level procedure, experience and next steps. Atmospheric Meas Tech. 2019 Jun 28;12(6):3435–52.

12. Matavulj P, Panić M, Šikoparija B, Tešendić D, Radovanović M, Brdar S. Advanced CNN Architectures for Pollen Classification: Design and Comprehensive Evaluation. Appl Artif Intell. 2023 Dec 31;37(1):2157593.

13. Maya-Manzano JM, Tummon F, Abt R, Allan N, Bunderson L, Clot B, et al. Towards European automatic bioaerosol monitoring: Comparison of 9 automatic pollen observational instruments with classic Hirst-type traps. Sci Total Environ. 2023 Mar 25;866:161220.

14. Gabbarini V, Rossi R, Ciparisse JF, Malizia A, Divizia A, De Filippis P, et al. Laser-induced fluorescence (LIF) as a smart method for fast environmental virological analyses: validation on Picornaviruses. Sci Rep. 2019 Aug 29;9(1):12598.

15. Kiselev D, Bonacina L, Wolf JP. Individual bioaerosol particle discrimination by multi-photon excited fluorescence. Opt Express. 2011 Nov 21;19(24):24516–21.

16. Kiselev D, Bonacina L, Wolf JP. A flash-lamp based device for fluorescence detection and identification of individual pollen grains. Rev Sci Instrum. 2013 Mar 5;84(3):033302.

17. Pan YL. Detection and characterization of biological and other organic-carbon aerosol particles in atmosphere using fluorescence. J Quant Spectrosc Radiat Transf. 2015 Jan 1;150:12–35.

18. Breiman L. Random Forests. Mach Learn. 2001 Oct 1;45(1):5–32.

19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12(85):2825–30.

20. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? Adv Neural Inf Process Syst. 2022 Dec 6;35:507–20.

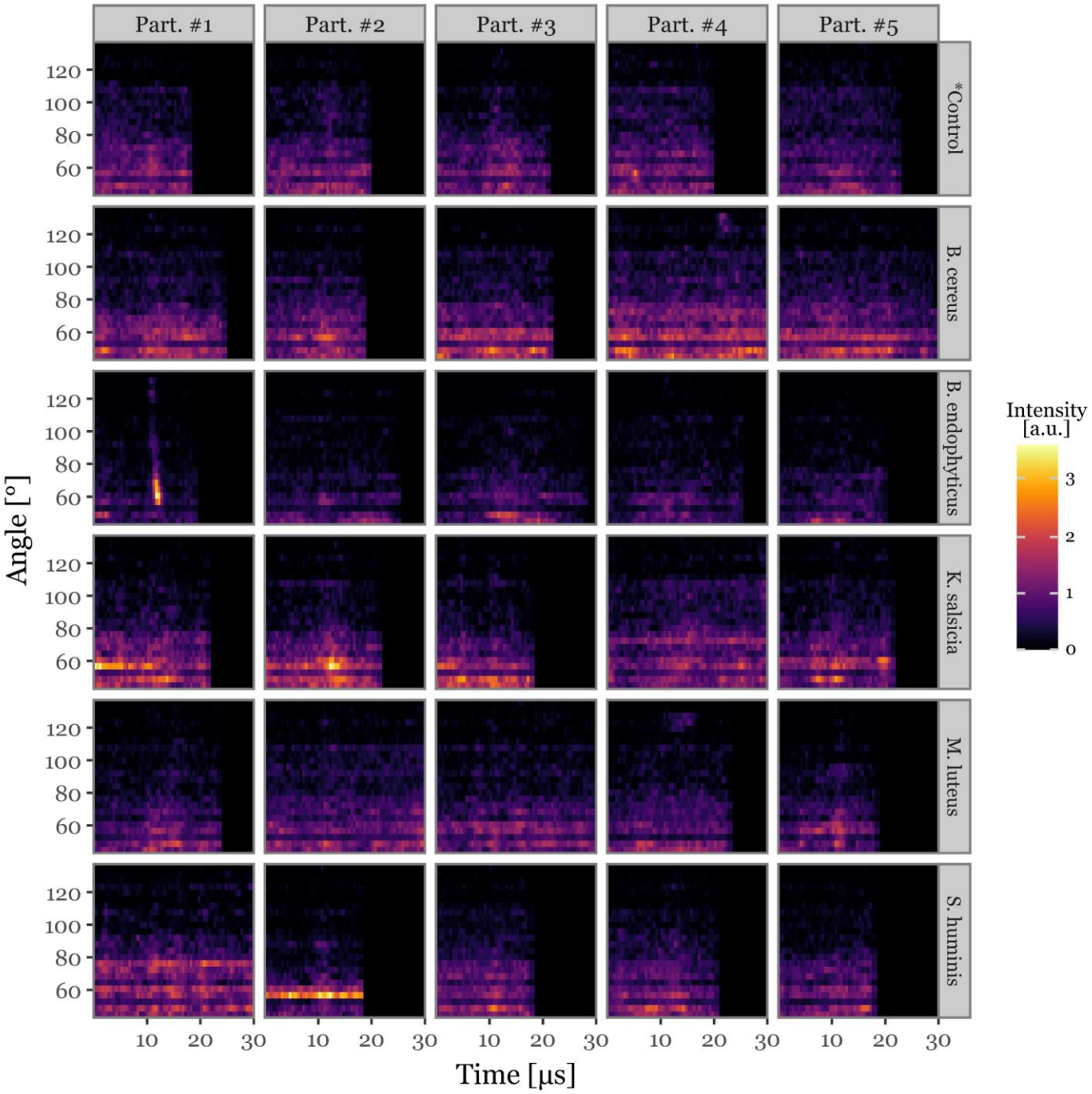
21. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. Handb Brain Theory Neural Netw. 1995;3361(10):1995.

22. Pöhlker C, Huffman JA, Pöschl U. Autofluorescence of atmospheric bioaerosols – fluorescent biomolecules and potential interferences. Atmospheric Meas Tech. 2012 Jan 9;5(1):37–71.

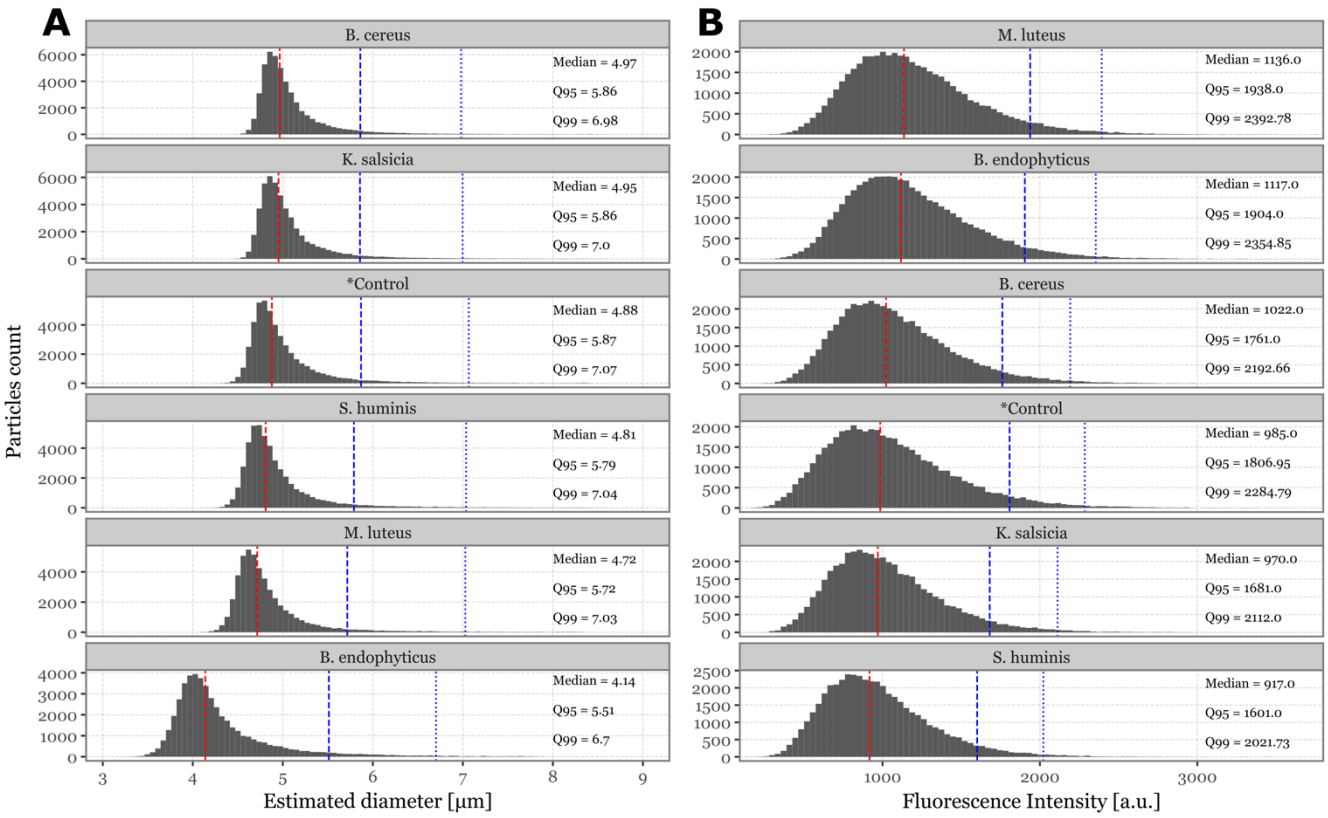
23. Després VR, Huffman JA, Burrows SM, Hoose C, Safatov AS, Buryak G, et al. Primary biological aerosol particles in the atmosphere: a review. Tellus B Chem Phys Meteorol [Internet]. 2012 Jan 1 [cited 2024 Aug 19];64(1). Available from: https://b.tellusjournals.se/articles/10.3402/tellusb.v64i0.15598

24. Hill SC, Pinnick RG, Niles S, Fell NF, Pan YL, Bottiger J, et al. Fluorescence from airborne microparticles: dependence on size, concentration of fluorophores, and illumination intensity. Appl Opt. 2001 Jun 20;40(18):3005–13.

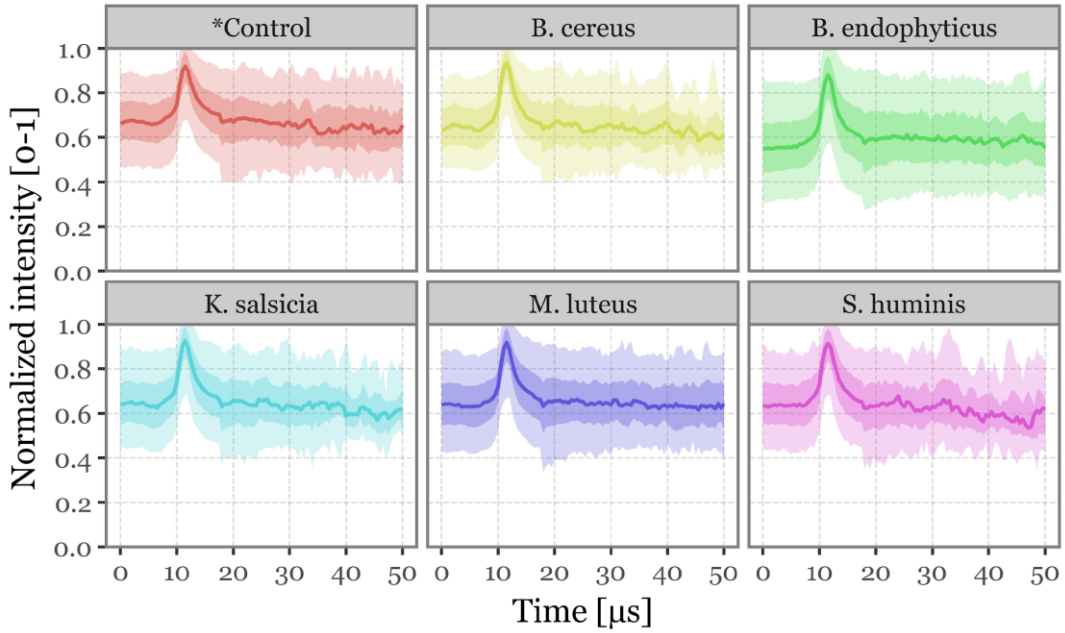
7. Supplementary Tables and Figures



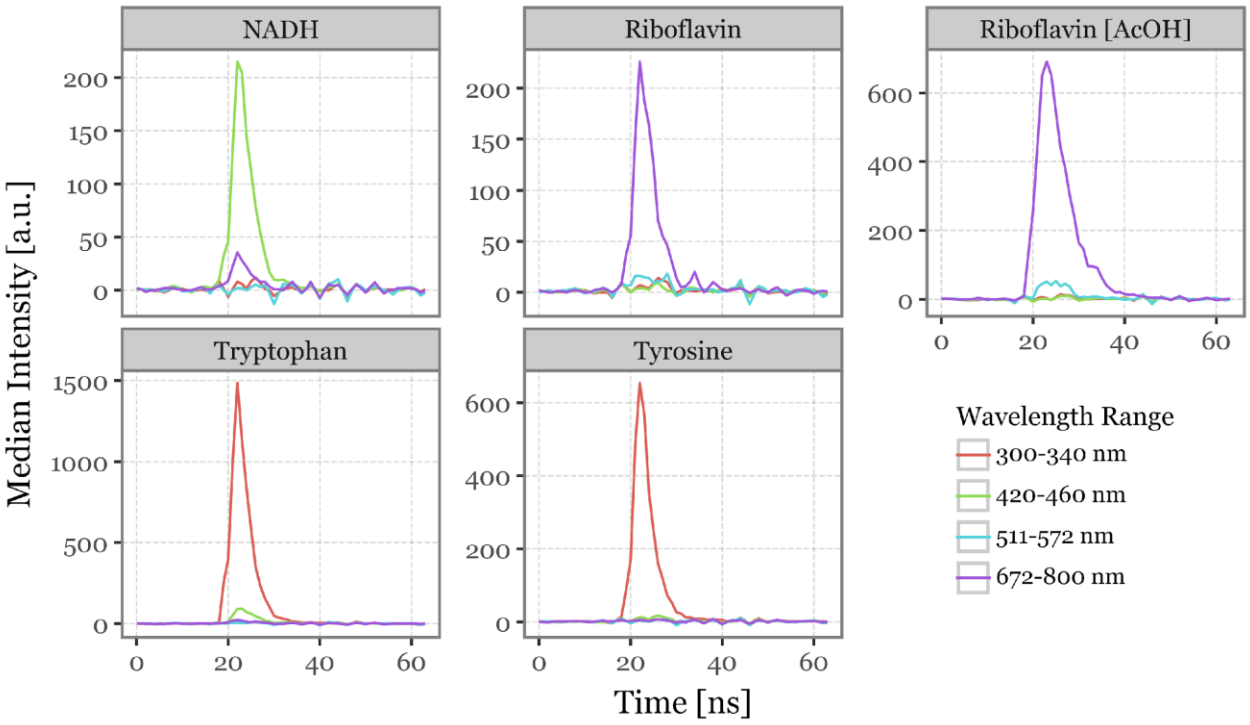
Suppl. Figure 1. Example of the scattering images of 5 particles per sample group as they are either cropped or zero padded to 60 acquisitions (30 µs time) to be included in the models.



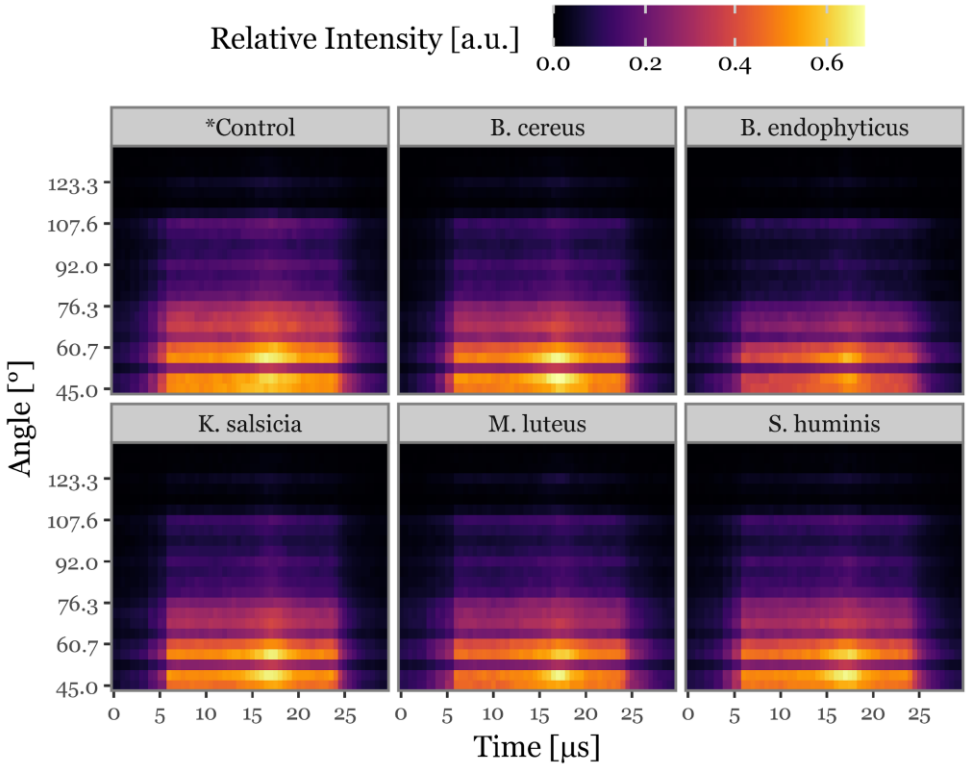
Suppl. Figure 2. Size and fluorescence intensity distribution for bacterial samples. Panel A (left) depicts the distributions of estimated diameters per sample group, and panel B (right) depicts the distribution of maximum recorded fluorescence intensities per sample group. Red vertical line indicates the median, with the blue dashed line indicating the 95th percentile and the blue dotted line indicating the 99th percentile.



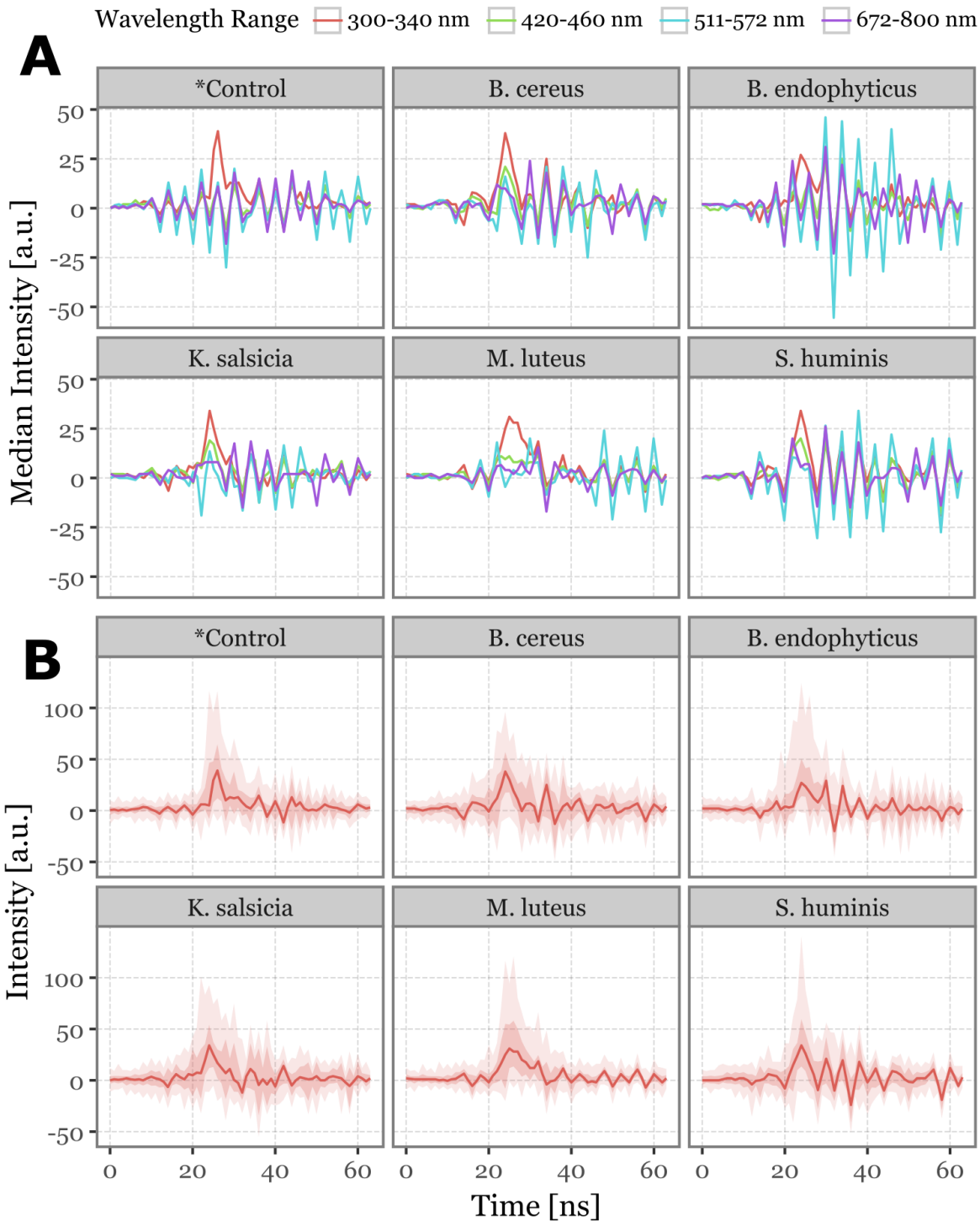
Suppl. Figure 3. Distribution of scattering intensity across all angles as a function of acquisition time. The solid line represents the median value for all particles within each group. The darker shaded area indicates the interquartile range (25th to 75th), while the lighter shaded area represents the 5th to 95th percentile range.



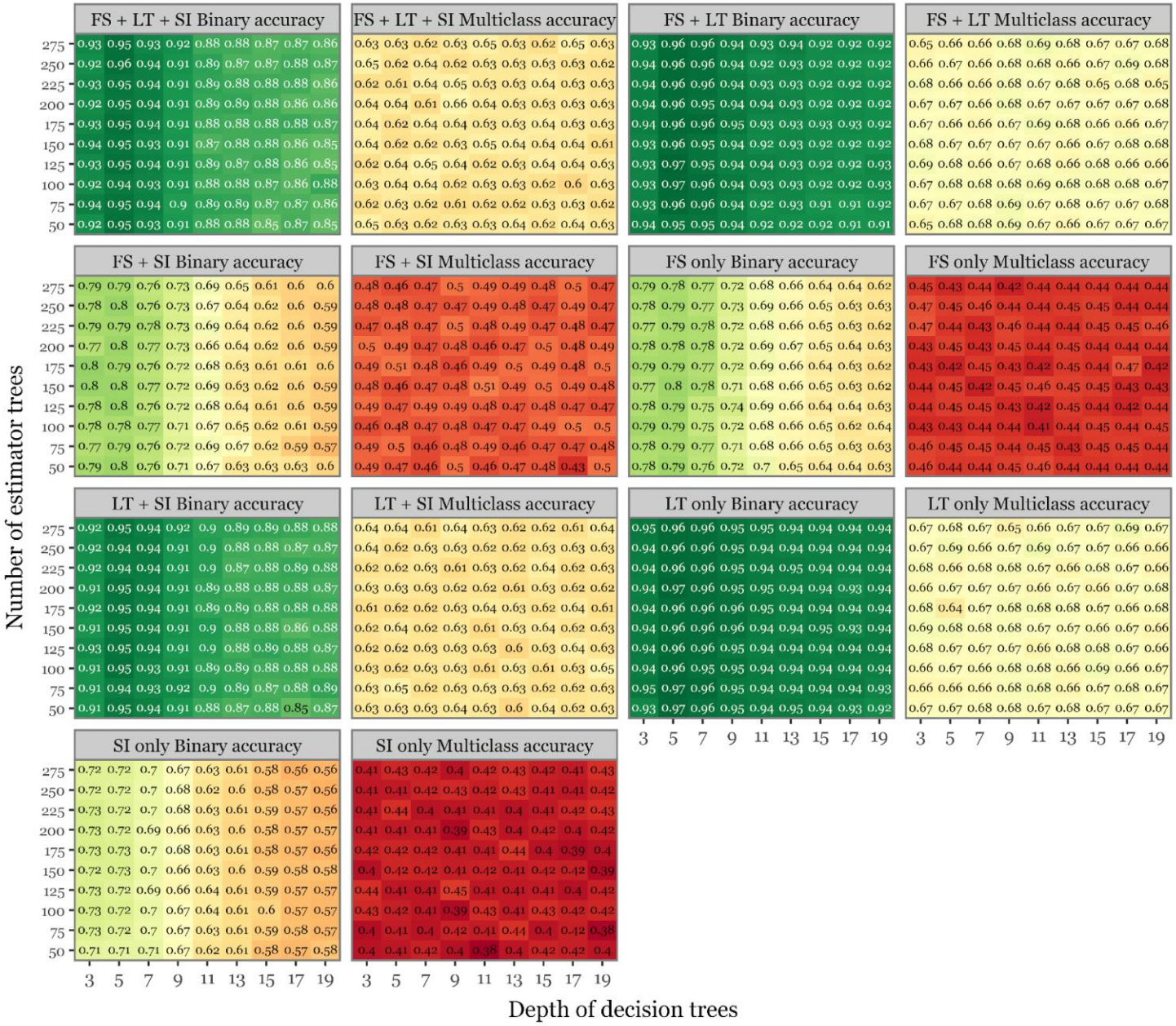
Suppl. Figure 4. Fluorescence lifetimes for aerosolized fluorophores. Median fluorescence intensity as a function of time (in ns) for the 4 specified wavelength ranges across the 100 most fluorescent particles for each of the aerosolized fluorophores.



Suppl. Figure 5. Average scattering images for bacterial aerosols. Heatmaps depicting the average relative light intensity for each combination of timely acquisition (x-axis) and angle (y-axis) and group of aerosolized particles.



Suppl. Figure 6. Fluorescence lifetimes for bacterial aerosols. On panel A (top), the median fluorescence intensity as a function of time (in ns) for the 4 specified wavelength ranges across the 100 most fluorescent particles for each of the bacterial groups. On panel B (bottom), the same but focusing only on the 300-340 nm wavelength range, with the dark shaded area representing the interquartile range (25^{th} to 75^{th} percentiles) and the lighter shaded area representing the 5^{th} to 95^{th} percentiles.



Suppl. Figure 7. Summary of the hyperparameter optimization process. Each box shows the test set balanced accuracy performance as a function of the depth of decision trees and the number of estimators used on the trained random forest models, both for the binary and multiclass problem (indicated on top of each box). 7 different combinations of predictors are used, also indicated on top (FS=Fluorescence Spectra, LT=Lifetime, SI=Scattering Images).

4. Discussion

4.1 Kawasaki Disease

One of the main aims of this thesis is to investigate the complex relationships between environmental and atmospheric factors and their impact on public health, with a particular focus on KD. Through two comprehensive studies, presented as Paper I and Paper II, we have explored how air quality, specifically airborne particulate matter enriched in metals, and seasonal patterns influence the incidence and progression of KD in Japan. This global discussion section synthesizes the key findings from these studies, integrates them with current literature, and examines their implications for understanding KD's etiology.

4.1.1 Paper I: sub-weekly signatures and metal and metalloid-enriched PMs

In Paper I, we employed signal-detection approaches to uncover linkages between air pollution and the incidence of KD in Japan.

Main findings

Correlation Between Metal and Metalloid-Enriched Particulate Matter and KD Incidence: Over a 37-day period in the Spring of 2011, an intensive survey was conducted in Kumamoto, collecting air samples that measured concentrations of nearly 60 major and trace elements, predominantly metals and metalloids (MM) [117]. We identified a cluster of 38 MM whose concentrations exhibited a striking coherence with the number of KD cases in the Kumamoto prefecture during the same period, with a response lag of up to three days (consistent with established KD response times). Notably, fluctuations in these MM concentrations explained over 40% of the total variability in KD cases. Days with high MM concentrations (exceeding 227 ng/m^3) consistently corresponded with significantly higher numbers of KD cases.

Identification of a Sub-Weekly Cycle in KD Incidence: Through frequency decomposition using Singular Spectrum Analysis (SSA)[118] on long-term KD records for the whole of Japan, we discovered a significant sub-weekly cycle (SWC1) with a frequency of 3.5 days, distinct from the typical 7-day cycle. This cycle has been consistently present since the early 1970s and became more pronounced in recent years, coinciding with increased seasonality in KD incidence. The SWC1 was also strongly evident during the nationwide KD epidemics of 1979, 1982, and 1986.

Association Between Ultrafine Particulate Matter and KD Incidence via LIDAR Observations: Using LIDAR optical measurements from 2010 to 2016, we estimated the presence of the ultrafine (<1 μm) fraction of particulate matter (PM1) in the atmosphere using the absc532 coefficient. We observed that the presence of the SWC1 in the LIDAR time series intensified as we approached the vertical coordinates associated with synoptic-scale air intrusion events. The amount of ultrafine aerosols in PM clearly increased during events associated with KD maxima, exhibiting a striking positive coevolution between the SWC1-PM1 cycle and KD incidence, with the SWC1 in the LIDAR data slightly preceding KD cases. Furthermore, synchronous annual cycles appeared in both variables, with coincident maxima in winter and minima in summer, reinforcing the central role of aerosols in KD epidemiology.

Significance of the findings

Previous studies on KD have primarily focused on seasonal and long-term variability, often overlooking day-to-day fluctuations in disease incidence and aerosol changes. The discovery of a consistent sub-weekly cycle (SWC1) with a 3.5-day frequency in all KD epidemiological records marks a significant advancement in understanding the disease's dynamics. This cycle is present continuously since the early 1970s, including during the major epidemics of 1979, 1982, and 1986, suggesting a persistent environmental or atmospheric influence on KD incidence.

The 3.5-day cycle does not align with the conventional 7-day weekly cycle associated with human activities, such as reduced emissions during weekends [119,120]. Our findings indicate that the SWC1 is not a mere artifact of reporting biases or healthcare access patterns but reflects a genuine environmental periodicity influencing KD cases. Such co-variability between KD incidence and aerosol MM concentrations is unlikely to occur by chance, highlighting a strong and consistent association.

The intensive 37-day air quality survey in Kumamoto revealed that fluctuations in the concentrations of 38 MM correlate strongly with KD incidence. These metals and metalloids, including zinc (Zn), lead (Pb), manganese (Mn), barium (Ba), selenium (Se), and copper (Cu), are known to originate from anthropogenic sources, particularly industrial emissions and agricultural activities [121]. The composition denotes a mixture of dust and fine aerosols, suggesting that air masses carrying these particles are a complex blend of emissions from different parts of East Asia.

High KD incidence days coincided with elevated continental pollution episodes characterized by fine sulfatic transboundary aerosol intrusions and higher concentrations of metallic trace elements. In contrast, days with minimum KD cases were associated with the advection of cleansing marine air containing lower levels of toxic metals. This pattern indicates that specific aerosol compositions, especially those enriched with certain metals, may contribute to KD onset.

The use of LIDAR data allowed us to examine the vertical distribution of ultrafine particulate matter and its association with KD incidence. We observed that the SWC1 in the LIDAR time series intensified near the vertical coordinates corresponding to synoptic-scale air intrusion events. These intrusions bring ultrafine particles (<1 μm) from above the planetary boundary layer (PBL) into the lower atmosphere, influencing ground-level air quality. The synchronization of the SWC1 in both KD incidence and PM1 concentrations suggests that atmospheric dynamics play a crucial role in modulating KD cases. The fact that these patterns are evident up to 6 km in the lower troposphere emphasizes the significance of long-range transport of pollutants in affecting local health outcomes.

Exposure to fine and ultrafine particulate matter, especially those enriched with metals, is known to induce pulmonary oxidative stress and inflammation [122,123]. Inhalation of these particles can generate reactive oxygen species (ROS), leading to both innate and adaptive immune responses [124,125]. Such immune modulation may be critical in triggering KD in genetically predisposed children.

The metals identified in our study, such as Zn, Cu, and Pb, have shown high correlations with pulmonary inflammation in toxicological studies using animal models [126]. The presence of soluble transition metals enhances inflammatory responses via increased oxidative stress, potentially contributing to the processes observed in KD.

Alignment with previous research

Our findings contribute to an evolving understanding of the environmental factors influencing KD. While earlier studies on urban air pollution and KD yielded inconclusive or negative results [127], recent research has started to identify associations between specific pollutants and KD incidence. Buteau and colleagues found that prenatal exposure to ambient and industrial air pollution was associated with KD in childhood [128]. Similarly, Jung and colleagues reported

positive associations between ozone (O_3) levels and KD admissions in a case-crossover study in Taiwan [129].

By focusing on the chemical nature and physical size of aerosols, particularly metal-enriched ultrafine particles, our study harmonizes these findings and provides a plausible mechanism for the observed associations. The unprecedented analysis of a high-resolution 46-year daily record of KD admissions strengthens the validity of our conclusions.

Implications for Public Health and future Research:

The findings of the study clearly point to several public health policies and research directions that might reduce disease burden and provide answers to solve the many unknowns of KD:

Inclusion of LIDAR data for Early-Warning Systems: Our study is the first to demonstrate how close scrutiny of LIDAR profiles can be used as effective sentinels for alerting high-risk KD events. The extensive AD-Net LIDAR network in Japan provides real-time data that could be leveraged to develop important air quality services for health, potentially reducing KD incidence through timely interventions.

Policy and environmental controls: Recognizing that KD exacerbation is associated with external air intrusions of anthropogenic origin emphasizes the need for regional cooperation in controlling emissions. Policies aimed at reducing agricultural and industrial emissions of metal-enriched aerosols could have significant public health benefits.

Research Directions: While our findings point toward an unequivocal link between ultrafine metal-enriched aerosols and KD, causality cannot be directly inferred. Further research is necessary to investigate the specific biological mechanisms involved. Exploring whether other factors, such as microbes or organic by-products embedded in aerosols, contribute to KD pathology is also essential. Understanding the dose-response relationship to MM and their role in immune activation could provide deeper insights into KD and similar vasculitides.

4.1.2 Paper II: Spatiotemporal analysis of age-stratified KD incidence

In Paper II, we conducted an extensive age-stratified spatiotemporal analysis of KD incidence across Japan, using daily prefectural records from 2000 to 2019, yielding several key findings that expand the current knowledge of the disease epidemiology.

Main findings

Stable Sex ratio and temporal dynamics: KD incidence consistently showed a higher prevalence in males compared to females, with a male-to-female ratio of approximately 1.4:1. This sex difference remained stable over time, across regions, and among different age groups. Both sexes exhibited identical temporal dynamics in terms of incidence trends and seasonal patterns.

Incidence trends: The overall incidence of KD has been increasing over the past two decades, particularly among toddlers (6–24 months) and young children (24–60 months). For these groups, the incidence rates nearly doubled from 2000 to 2019. In contrast, the incidence among infants under 6 months plateaued after 2010 and has remained stable ever since.

Shift in Seasonal patterns for toddlers post-2016: While most age groups maintained a stable seasonal pattern characterized by a winter peak, the toddler group exhibited a striking shift after 2016. This group developed a new seasonal peak from August to November, effectively replacing the previously dominant winter peak. This shift was consistent across most regions in Japan and was specific to the toddler age group, which accounts for over 40% of total KD cases.

Stable spatial distribution across age groups: Despite the increasing incidence, KD's spatial distribution across Japan remained stable, with no significant clustering or hotspots over time at the prefectural level. The prefectural incidence patterns were consistent across all age groups, indicating that factors influencing higher or lower overall KD incidences in specific prefectures affect all age groups similarly.

No significant changes in environmental factors explaining the seasonal shift: Our comprehensive analysis of various environmental factors—including wind patterns, air source regions, temperature, precipitation, vegetation dynamics (EVI), and air pollution—did not reveal any significant changes between the pre-2016 and post-2016 periods that could account for the observed shift in KD seasonality among toddlers.

Significance of the findings

The abrupt and age-specific shift in KD seasonality after 2016 is unprecedented in the epidemiology of the disease. The emergence of a new peak in KD cases among toddlers during the fall months (August to November) represents a dramatic change from the traditional winter peak that has been consistently observed across all age groups. This shift indicates that toddlers are being influenced by factors distinct from those affecting other age groups or, at the very least, that one of the triggers responsible for disease onset has undergone a dynamic change specific to this age group. The reason behind this unique impact on toddlers remains a mystery: it could stem from developmental immunological differences, unique exposure patterns linked to age, or behavioural factors specific to toddlers. The failure to associate this shift with changes in traditional environmental variables suggests that more complex factors may be involved or, alternatively, it may indicate a need for finer temporal analyses that can capture the highly nonlinear dynamics potentially driving the connectivity between the environment and KD.

The stable spatial distribution of KD incidence across prefectures, despite temporal changes and increasing incidence rates, indicates that the local factors influencing the baseline KD incidence rates at each prefecture are consistently affecting all age groups similarly. This suggests that while the triggers of KD may vary temporally and differ by age group, the baseline risk factors characteristic of each prefecture have not changed with time.

Previous studies have documented the seasonal variation in KD incidence, commonly observing winter peaks across various regions and age groups [25,29,45]. The higher incidence in males compared to females is also well-established in the literature, and is consistent with the heightened activity of both the innate (inflammatory) and adaptive immune responses during early life stages in males, in contrast with post-puberty stages [130,131]. Our findings align with these established patterns but introduce new complexities by identifying an age-specific seasonal shift. While some studies have hinted at age-related differences in KD presentation and incidence [45,132] none have reported a sudden and sustained seasonal shift confined to a specific age group. This divergence adds a new dimension to the understanding of KD epidemiology and will have to be thoroughly studied in further research.

Implications for Public Health and future research directions

The results from this study might also inform some public health policies and provide guidance on future research directions, such as:

- **Targeted (age-stratified) surveillance**: Public health authorities could consider implementing age-specific surveillance for KD, particularly focusing on toddlers during the newly identified high-risk months (August to November). This could facilitate earlier detection and treatment, potentially reducing the risk of complications, but also provide better data for research. The findings demonstrate that if we are to look for etiological clues of the disease, age deserved to be considered as a key factor in the way that the pediatric population both interacts with the environment and reacts to it.
- **Investigate non-traditional environmental factors**: Future research should explore non-traditional environmental factors that may contribute to KD onset, such as airborne allergens, microbial content of air masses, or chemical compounds not typically monitored in standard air quality assessments.
- **Behavioural and societal factors**: Studies should examine changes in childcare practices, vaccination schedules, dietary factors, or other societal changes that may have occurred around 2016 and disproportionately affect toddlers.
- **Immunological studies**: Research into the immunological differences between toddlers and other age groups could provide insights into why this group is uniquely affected by the seasonal shift. Understanding the maturation of the immune system and its interaction with environmental exposures is crucial.
- **Global Comparisons**: Investigating whether similar age-specific seasonal shifts have occurred in KD incidence in other countries could help determine if the phenomenon is unique to Japan or reflects a broader trend.

Summary

In summary, the age-specific shift in KD seasonality among toddlers in Japan represents a significant development in the understanding of KD epidemiology. While the triggers of KD remain elusive, our study highlights the necessity of considering age-specific factors and the potential limitations of focusing solely on traditional environmental variables. By broadening the scope of research and incorporating multidisciplinary perspectives, there is potential to uncover the complex interactions that contribute to KD onset and unveil, at last, the etiology of this mysterious ailment after decades of research.

4.2 Infectious Diseases and climatic drivers

4.2.1 Paper III: Climatic signatures during the COVID-19 pandemic

The investigation presented in Paper III sought to elucidate the role of climate factors—specifically temperature (T) and absolute humidity (AH)—in the transmission dynamics of COVID-19 across various spatial and temporal scales. By employing advanced statistical methods designed to detect transient associations, such as Scale-Dependent Correlation Analysis (SDC) [133,134] and its multiscale extension (MSDC) developed specifically for the study, we uncovered consistent and robust negative relationships between these climate variables and COVID-19 case numbers globally, regionally, and locally.

Main findings

At the global scale, analysing the initial data from 162 countries, we observed significant negative associations between the basic reproductive number (R_0) of COVID-19 and both temperature and absolute humidity. Lower temperatures and humidity levels were linked to higher transmission rates of SARS-CoV-2, accounting for 10–25% of the variance in R_0.

This negative relationship was evident not only globally during the initial stage of the pandemic, but also when we zoomed into regional and local scales during the subsequent waves. For instance, in Europe—specifically in countries like France, the United Kingdom, Italy, Spain, and Germany—we found that the negative correlation persisted during both the first and second pandemic waves. These associations were particularly strong during the rising and declining phases of the waves, with a noticeable break during the summer months.

By analysing the frequency of significant correlations across different ranges of temperature and humidity, we identified critical thresholds where climate effects on COVID-19 transmission were most pronounced. Specifically, temperatures between 12–18 °C and absolute humidity levels between 4–12 g/m³ were associated with stronger negative correlations. This suggests that SARS-CoV-2 transmission is facilitated under cooler and drier conditions, aligning with patterns observed in other respiratory viruses like influenza.

To further substantiate the role of climate factors, we incorporated temperature into a process-based epidemiological model as a driver of the transmission rate. This model consistently outperformed those without climate variables or with only sinusoidal seasonality across different

regions (Catalonia, Lombardy) and spatial scales (regional and city levels). The enhanced model more accurately captured observed cases and deaths, particularly during the first wave of the pandemic. These results highlight the significant influence of temperature on the transmission dynamics of COVID-19.

Our findings classify COVID-19 as a seasonal low-temperature infection, similar to influenza and other respiratory viruses. The strong association between lower temperatures and humidity levels and increased COVID-19 cases suggests an important contribution of the airborne pathway in SARS-CoV-2 transmission. Colder and drier conditions favour the survival and transmission of respiratory aerosols, supporting the hypothesis that airborne spread plays a significant role in the pandemic's dynamics.

Significance of findings

The confirmation of climate influence on COVID-19 transmission has substantial implications for public health strategies and pandemic preparedness. Understanding that lower temperatures and humidity levels facilitate the spread of SARS-CoV-2 allows for more accurate forecasting of seasonal surges and informs the timing of interventions such as vaccination campaigns, public messaging, and resource allocation. Emphasizing measures that address airborne transmission—like improving indoor ventilation, promoting mask-wearing in indoor settings, and considering air filtration systems—becomes crucial, especially during colder months.

Comparison within current context

Since the publication of our study in October 2021, substantial new data and research have emerged regarding the transmission dynamics of COVID-19, with the susceptibility of the population largely affected by natural immunity after SARS-CoV-2 infection and the widespread vaccination.

Studies using later data and published after the beginning of 2022 generally support the notion that COVID-19 exhibits seasonal patterns influenced by climatic factors, with temperature and humidity being negatively correlated with COVID-19 transmission and mortality [135–138].

However, recent research also highlights that the relationship between climate factors and COVID-19 transmission is more complex than initially understood. Factors such as human behaviour, public health interventions, and the emergence of new variants significantly modulate

transmission dynamics. The highly transmissible Delta and Omicron variants have shown high transmission rates regardless of climatic conditions, potentially diminishing the relative impact of temperature and humidity [139–141]. Moreover, widespread vaccination and natural immunity have altered susceptibility patterns, affecting transmission in ways that may confound simple climatic associations[142].

Despite these complexities, the importance of airborne transmission under varying climatic conditions continues to be supported. Studies confirm that SARS-CoV-2 maintains infectivity longer under cooler and less humid conditions, reinforcing the airborne transmission hypothesis [143]. Research emphasises the need for improving indoor air quality, especially during colder months when people spend more time indoors with limited air circulation [144].

In summary, the new information from the last few years generally aligns with our study's conclusions, with some additional nuances. While COVID-19 does exhibit seasonality influenced by temperature and absolute humidity, the impact of climatic factors is modulated by human behaviour, emergence of new variants, and immunity levels. Our study's conclusions remain relevant but should be considered within the broader context of these additional factors. Ongoing research accounting for new variants and population status is essential to refine our understanding and adapt public health strategies accordingly, and climate should be considered as a key variable, especially in the context of climate change, as the current trends make it likely to incentivise the emergence of new pathogens both via affecting transmission chains but also facilitating zoonotic spillovers [145].

Summary

Our study provides compelling evidence that climate factors, particularly temperature and absolute humidity, significantly influence the transmission dynamics of SARS-CoV-2. The consistent negative associations observed across different regions and scales support the classification of COVID-19 as a seasonal low-temperature infection. These findings have critical implications for public health strategies and stress the need to incorporate climate considerations into pandemic planning and to focus on measures that mitigate airborne transmission. By enhancing epidemiological models with climate data and prioritizing interventions that address the airborne spread of SARS-CoV-2, we can improve our response to the ongoing pandemic and better prepare for future respiratory virus outbreaks.

4.3 The Aerobiome

4.3.1 Paper IV: Microbiome of the troposphere and long-distance transport of pathogens

In paper IV, we present a study involving the collection of air samples from the free troposphere using aircraft surveys over Japan. Using advanced DNA extraction, metagenomic sequencing, and microbial culturing techniques, the goal was to identify and characterize microorganisms present at these altitudes. The main findings of the study were rather striking:

Main Findings

Our study provides compelling evidence that viable microorganisms, including potential human pathogens and antibiotic-resistant bacteria (ARB), are present in the troposphere above the planetary boundary layer (PBL) and can be transported over long distances—up to 2,000 kilometers—by atmospheric currents. Through ten aircraft surveys over Japan, we collected aerosol samples at altitudes up to 3,000 meters and discovered a rich microbial diversity comparable to that found at ground level when there is entrainment of air from high altitudes.

The backtrajectory analyses using the FLEXPART model revealed that air masses originating from Northeast (NE) China traveled to Japan within two days, moving above the PBL without contamination from ground-level sources en route. Chemical analyses of the aerosols showed elevated levels of elements associated with agricultural activities, such as potassium (K), zinc (Zn), and boron (B), suggesting that these air masses carried particulate matter from agricultural regions rich in fertilizers and pesticides.

The microbial communities detected were diverse, comprising over 266 different fungal genera and 305 bacterial genera. Dominant bacterial phyla included Actinobacteria, Bacillota, Proteobacteria, and Bacteroidetes, while Ascomycota was the predominant fungal phylum. Notably, several identified taxa are known human pathogens, including *Escherichia coli*, *Serratia marcescens*, *Staphylococcus* spp., *Clostridium difficile*, and *Candida* spp. Culturing of airborne microbes confirmed their viability, and some isolates exhibited multidrug resistance, highlighting the potential for ARB to be disseminated through atmospheric transport.

Significance of the Findings

These findings challenge the traditional understanding that harsh conditions at high altitudes—such as low moisture, nutrient scarcity, and high ultraviolet (UV) radiation—limit microbial survival and proliferation. While previous studies have documented the presence of microorganisms in the atmosphere, few have focused on the upper troposphere or addressed the viability and pathogenic potential of these organisms [80,99]. Our study extends this knowledge by confirming not only the presence but also the viability of a wide range of microbes, including potential human pathogens and ARB, at significant altitudes.

The detection of viable pathogens and ARB over such distances and altitudes represents a novel contribution to atmospheric microbiology and infectious disease epidemiology. It suggests that the upper troposphere can serve as a reservoir and conduit for microorganisms, facilitating their dispersal across geographical regions. This has implications for understanding the global spread of infectious diseases and antibiotic resistance, particularly in the context of increasing global connectivity and climate change. The findings also have implications for understanding the transmission of diseases such as KD. As mentioned in previous sections and Papers I and II, previous research has suggested that tropospheric winds from Northeast China carry the etiologic agent of KD to Japan [38,39,46]. The detection of viable human pathogens in air masses traveling over 2,000 kilometers supports the hypothesis that airborne agents transported via these routes could contribute to the onset of KD. Given the multifactorial nature of KD, the presence of a diverse microbial community, including potential pathogens, in the upper troposphere provides a plausible mechanism for the observed epidemiological patterns. Our study adds weight to the possibility that the agent(s) causing KD might be dispersed through long-distance atmospheric transport.

Our findings also highlight the interconnectedness of environmental processes and human activities. The association of specific microbial taxa with chemical elements linked to agricultural practices highlights the impact of intensive farming on atmospheric microbial composition. The presence of elements like zirconium (Zr) and hafnium (Hf), not mined in Japan but abundant in NE China, further supports the conclusion that these microbes originated from distant agricultural regions.

Implications for Public Health and Future Research

The discovery of viable human pathogens and ARB in the upper troposphere has substantial implications for public health. While the concentrations of these microbes in the atmosphere are low, their potential impact cannot be entirely dismissed. The inhalation of pathogenic microbes and ARB could contribute to respiratory infections and other health issues, particularly in susceptible or immunocompromised individuals. However, it remains uncertain whether the doses encountered in the atmosphere are sufficient to pose a real threat to human health.

This uncertainty emphasises the need for further research to assess the viability, infectivity, and pathogenicity of these airborne microbes upon deposition. Experiments simulating atmospheric conditions could provide insights into the survival mechanisms of these organisms and their potential health impacts. Additionally, epidemiological studies could explore correlations between atmospheric microbial content and health outcomes in populations downwind of agricultural regions. The findings also raise questions about the role of atmospheric transport in the dissemination of antibiotic resistance genes. The identification of multidrug-resistant strains among the cultured isolates suggests that ARB can be spread over long distances through the atmosphere. This could contribute to the global challenge of antimicrobial resistance, stressing the importance of monitoring environmental pathways of ARB transmission [146].

Some limitations of the study include the specific temporal and spatial scope of sampling, as flights were conducted over two periods in February and April 2014. Expanding the sampling to cover different seasons and regions would provide a more comprehensive understanding of atmospheric microbial dynamics. Additionally, while we identified potential pathogens and ARB, we did not assess their infectivity or the likelihood of causing disease in humans, so studies developing on that might provide valuable answers.

Summary

In conclusion, our study reveals that the upper troposphere harbours a diverse and viable microbial community, including potential human pathogens and antibiotic-resistant bacteria. The long-distance transport of these organisms attached to aerosols originating from agricultural regions suggests that the atmosphere plays a role in the global dispersal of microorganisms. While the concentrations are ultra-low, and the actual health risk remains uncertain, the findings highlight the need for further research into the atmospheric microbiome and its implications for public health.

Understanding these processes is essential for developing strategies to mitigate the spread of infectious diseases and antimicrobial resistance on a global scale. Our findings also provide support for the hypothesis that airborne agents transported via tropospheric winds could contribute to diseases like KD, highlighting the importance of considering atmospheric pathways in disease ecology. Future research should focus on assessing the viability and pathogenicity of airborne microbes, exploring their survival mechanisms in the atmosphere, and evaluating their potential impact on human health. Such efforts will contribute to a more comprehensive understanding of the atmosphere's role in microbial ecology and the interconnectedness of environmental factors and human health.

4.3.2 Paper V: Real-time characterization of the aerobiome with LIF

Main Findings

In this study, we successfully modified the Rapid-E Real-Time Airborne Particle Analyzer by replacing its standard 337 nm ultraviolet (UV) laser with a 266 nm laser. This modification was aimed at optimizing the excitation of fluorophores characteristic of microbial cell walls, such as amino acids (tryptophan, tyrosine, phenylalanine), flavins (riboflavin, FAD, FADH), and coenzymes (NADH, NADPH). Using this customized device, we demonstrated the capability to detect and classify bacterial particles in aerosols in near real-time. By generating artificially aerosolized samples enriched with five different bacterial species—*Micrococcus luteus*, *Bacillus endophyticus*, *Bacillus cereus*, *Kocuria salsicia*, and *Staphylococcus hominis*—we achieved a class-balanced accuracy of 96.74% in distinguishing bacterial particles from controls and 69.24% accuracy in discriminating among the different bacterial species using Random Forest classifiers.

Significance of the findings

The ability to detect and classify bacterial particles in bioaerosols in real-time represents a significant advancement in atmospheric microbiology. Traditional methods, such as metagenomic next-generation sequencing, face challenges due to the ultra-low biomass of air samples, resulting in low DNA yields and compromised temporal resolution[102]. Our methodology addresses these limitations by providing a rapid, culture-independent approach that enhances both detection speed and specificity.

This work is, to our knowledge, the first to automate the detection and classification of bacterial particles in bioaerosols using a portable device suitable for real-time surveillance. By focusing on

microbial fluorophores and employing machine learning techniques, we have expanded the capabilities of existing bioaerosol detectors, which have predominantly been used for pollen detection and classification [111,112]. The high accuracy achieved in both detection and classification tasks emphasises the potential of this approach for applications in environmental monitoring, public health, and biosecurity.

Additionally, this methodology provides a foundation for real-time monitoring of airborne pathogens, which is crucial for early detection and response to biohazards. The integration of laser-induced fluorescence (LIF) with machine learning models offers a non-invasive, rapid, and accurate tool that can significantly enhance our ability to monitor and respond to microbial presence in the atmosphere.

Areas of Impact and Future Research

While the results are promising, translating this methodology from controlled laboratory settings to real-world environmental sampling presents challenges. Natural aerosols are complex and heterogeneous, comprising a diverse mix of biological and non-biological particles such as pollen, fungal spores, dust, and pollutants [77]. These components can interfere with the detection of bacterial particles by obscuring or mimicking their spectral and morphological signatures.

Future advancements should focus on developing methods to generate training datasets that more accurately reflect the variability and complexity of real-world bioaerosols. This includes collecting environmental aerosol samples and annotating them to train machine learning models that are robust against the noise and variability inherent in natural settings. Additionally, exploring more sophisticated machine learning algorithms, such as convolutional neural [147], may improve the extraction of meaningful features from scattering images, potentially enhancing classification performance.

Improvements in the device hardware could also be considered. Fine-tuning the wavelength calibration of the fluorescence acquisition system may address the observed shifts in spectral peaks, leading to more accurate detection of fluorophores. Enhancing the sensitivity of the detectors and optimizing the aerosolization process to ensure a higher proportion of particles containing the target biological content could further improve detection capabilities.

Summary

In summary, with this study we demonstrate the feasibility of using a modified Rapid-E bioaerosol analyzer equipped with a 266 nm laser and machine learning techniques to detect and classify bacterial particles in aerosols in near real-time. The methodology offers a rapid, accurate, and portable solution for monitoring the microbial fraction of bioaerosols, with significant implications for environmental monitoring and public health. While challenges remain in adapting this approach to complex real-world conditions, the findings provide a solid foundation for future developments aimed at comprehensive, real-time bioaerosol surveillance.

5. Conclusions and Outlook

The work presented in this thesis advances our understanding of the complex interactions between atmospheric processes, environmental pollutants, microbial ecology and their impacts on human health. Through a series of five studies, we have explored the multifaceted relationships between climate, air quality, airborne microorganisms and diseases such as Kawasaki Disease and COVID-19. The research has integrated expertise from many different disciplines, including epidemiology, atmospheric science, microbiology, data science, bioinformatics and several more to provide new insights and knowledge.

In **Paper I**, we identified a significant association between ultrafine particulate matter enriched with metals and metalloids (MM) and the incidence of KD in Japan. By uncovering a consistent sub-weekly cycle (SWC1) of 3.5 days in KD cases, we demonstrated that fluctuations in specific MM concentrations, likely originating from anthropogenic sources in East Asia, correlate strongly with KD incidence. The synchronization of SWC1 in both KD cases and ultrafine PM concentrations suggests that atmospheric transport of metal-enriched aerosols plays a crucial role in KD epidemiology.

Paper II expanded on KD research by conducting an age-stratified spatiotemporal analysis of KD incidence across Japan from 2000 to 2019. We discovered an abrupt and unprecedented shift in the seasonal pattern of KD among toddlers (6–24 months) after 2016, with a new peak emerging from August to November. This shift was specific to toddlers and was not associated with changes in traditional environmental factors. The findings highlight the necessity of considering age-specific factors and suggest that novel or previously unrecognized environmental or societal influences may be affecting KD onset in this age group.

In **Paper III**, we investigated the influence of climate factors—specifically temperature and absolute humidity—on the transmission dynamics of COVID-19. Our analysis across global, regional, and local scales consistently revealed significant negative associations between lower temperatures, lower humidity, and increased SARS-CoV-2 transmission rates. By incorporating temperature into epidemiological models, we improved their predictive accuracy, classifying COVID-19 as a seasonal low-temperature infection and emphasizing the importance of climate considerations in public health strategies.

Paper IV presented compelling evidence that viable microorganisms, including potential human pathogens and antibiotic-resistant bacteria (ARB), are present in the troposphere above the planetary boundary layer and can be transported over long distances by atmospheric currents. Through aircraft-based sampling, we detected a rich microbial diversity at altitudes up to 3,000 meters. The presence of known pathogens and ARB at these altitudes challenges traditional notions of microbial survival in harsh atmospheric conditions and suggests that the upper troposphere serves as a conduit for global dispersal of microorganisms. These findings also provide a plausible mechanism for the long-distance transport of agents potentially contributing to KD.

Finally, in **Paper V**, we developed a novel methodology for the real-time detection and classification of bacterial particles in bioaerosols using a modified Rapid-E bioaerosol analyzer equipped with a 266 nm laser. Employing machine learning classifiers, we achieved high accuracy in distinguishing bacterial particles from controls and in species-level classification. This technological advancement offers a rapid, portable solution for monitoring airborne bacteria, with significant implications for environmental monitoring and public health.

Significance of the Findings

Collectively, these studies highlight the critical role of the atmosphere as an active medium influencing human health through the transport of pollutants, climate variables, and microorganisms. The identification of specific metal-enriched ultrafine particles associated with KD incidence highlights the impact of anthropogenic emissions and atmospheric dynamics on disease onset. The discovery of an age-specific seasonal shift in KD among toddlers suggests that unique environmental or societal factors are influencing this vulnerable age group, necessitating targeted research and public health interventions.

The confirmation of climate factors influencing COVID-19 transmission emphasizes the need to integrate environmental data into epidemiological models and public health strategies. Recognizing COVID-19 as a seasonal low-temperature infection similar to influenza informs the timing and nature of interventions aimed at mitigating transmission, particularly measures addressing airborne spread.

The detection of viable pathogens and ARB in the upper troposphere has profound implications for understanding disease ecology and the global spread of infectious agents. It supports

hypotheses linking atmospheric transport to diseases like KD and raises concerns about the dissemination of antibiotic resistance through environmental pathways.

The development of real-time bioaerosol detection technology represents a significant advancement in our ability to monitor the aerobiome. By enabling rapid, accurate detection of airborne bacteria, including potential pathogens, we enhance our capacity for early warning and response to biohazards, contributing to public health and biosecurity efforts.

Outlook

The findings from the research presented in this thesis open several avenues for future investigation and application:

Elucidating KD etiology: The big question that led to the planning of this thesis in the first place still remains not completely answered. Almost 60 years after the recognition of the first cases in Japan, and plenty of research and efforts made towards understanding what causes the onset of this pediatric vasculitis, yet we still do not know the agent (or agents) actually eliciting it. Some steps have been done getting us closer to the solution, but much more research needs to be done to get to a complete understanding. As directions pointed directly by the results shown in this thesis, further research is needed to understand the mechanisms by which metal-enriched ultrafine particles contribute to KD onset. This includes exploring the potential role of oxidative stress, immune modulation, and the possibility that microbes or organic compounds attached to aerosols act as triggers. Understanding what exactly is in the air masses that show strong association with the incidence of the disease should provide some answer. Age-specific studies focusing on toddlers could uncover why this group experienced a seasonal shift in KD incidence, examining factors such as immune system development, exposure patterns, and societal changes.

Atmospheric Microbiology and Public Health: Expanding sampling efforts to different regions, altitudes, and seasons will enhance our understanding of atmospheric microbial dynamics. Investigating the viability, infectivity, and pathogenicity of airborne microorganisms upon deposition is crucial for assessing health risks. Epidemiological studies correlating atmospheric microbial content with health outcomes can inform public health policies, particularly in regions downwind of industrial or agricultural emissions.

Climate considerations in disease transmission: Incorporating climate variables into predictive models for infectious diseases can improve forecasting and intervention strategies. As climate change alters temperature and humidity patterns, understanding their influence on disease dynamics becomes increasingly important. Policies made to mitigate outbreaks and transmission of new and existing infectious diseases should definitely include climate variables in their modelling efforts. Additionally, public health initiatives should emphasize measures that mitigate airborne transmission, especially during periods conducive to increased spread.

Technological advancements in bioaerosol monitoring: Further development of real-time detection technologies, including enhancing device sensitivity and adapting machine learning models to complex environmental conditions, will improve our ability to monitor the aerobiome. This includes collecting diverse environmental samples to train robust models capable of distinguishing bacterial particles in heterogeneous aerosols.

Policy and environmental controls: Recognizing the health impacts of transboundary pollution and atmospheric transport of harmful particles underscores the need for regional cooperation in emission control. Implementing policies to reduce industrial and agricultural emissions of metal-enriched aerosols can have significant public health benefits.

This thesis advances our understanding of the intricate connections between the atmosphere and human health. By exploring how atmospheric particles and conditions influence diseases like KD and COVID-19, we highlight the necessity of integrating environmental considerations into public health frameworks. The atmosphere is not merely a passive medium but an active participant in disease ecology, capable of transporting and dispersing agents that impact health on a global scale.

Our research emphasizes the importance of continuous monitoring and investigation of the aerobiome, particularly in the context of climate change and increasing anthropogenic activities. By developing innovative detection technologies and refining our epidemiological models, we can better anticipate and respond to emerging health threats. As we move forward, it is imperative to foster multidisciplinary collaborations and implement policies that address the environmental determinants of health. Understanding the atmosphere's role in microbial ecology and disease transmission is essential for safeguarding human health and developing effective interventions in an increasingly interconnected world.

Bibliography

1. Burns JC, Glodé MP. Kawasaki syndrome. The Lancet. 2004 Aug 7;364(9433):533–44.

2. Kawasaki T. Kawasaki disease. Int J Rheum Dis. 2014;17(5):597–600.

3. Kawasaki T. [Acute febrile mucocutaneous syndrome with lymphoid involvement with specific desquamation of the fingers and toes in children]. Arerugi Allergy. 1967 Mar;16(3):178–222.

4. Kawasaki T, Kosaki F, Okawa S, Shigematsu I, Yanagawa H. A new infantile acute febrile mucocutaneous lymph node syndrome (MLNS) prevailing in Japan. Pediatrics. 1974 Sep;54(3):271–6.

5. Kawasaki T. Pediatric acute febrile mucocutaneous lymph node syndrome with characteristic desquamation of fingers and toes: my clinical observation of fifty cases. Pediatr Infect J. 2002;21(11):1–38.

6. Burns JC. Commentary: Translation of Dr. Tomisaku Kawasaki's original report of fifty patients in 1967. Pediatr Infect Dis J. 2002 Nov;21(11):993.

7. Nakamura Y, Yashiro M, Uehara R, Oki I, Watanabe M, Yanagawa H. Epidemiologic features of Kawasaki disease in Japan: results from the nationwide survey in 2005-2006. J Epidemiol. 2008;18(4):167–72.

8. Yanagawa H, Nakamura Y, Yashiro M, Uehara R, Oki I, Kayaba K. Incidence of Kawasaki disease in Japan: the nationwide surveys of 1999–2002. Pediatr Int. 2006 Aug;48(4):356–61.

9. Nakamura Y, Yashiro M, Uehara R, Sadakane A, Tsuboi S, Aoyama Y, Kotani K, Tsogzolbaatar EO, Yanagawa H. Epidemiologic features of Kawasaki disease in Japan: results of the 2009–2010 nationwide survey. J Epidemiol. 2012;22(3):216–21.

10. Ae R, Makino N, Kosami K, Kuwabara M, Matsubara Y, Nakamura Y. Epidemiology, Treatments, and Cardiac Complications in Patients with Kawasaki Disease: The Nationwide Survey in Japan, 2017-2018. J Pediatr. 2020;225:23-29.e2.

11. Ae R, Makino N, Kuwabara M, Matsubara Y, Kosami K, Sasahara T, Nakamura Y. Incidence of Kawasaki Disease Before and After the COVID-19 Pandemic in Japan: Results of the 26th Nationwide Survey, 2019 to 2020. JAMA Pediatr [Internet]. 2022 Oct 17 [cited 2022 Oct 20]; Available from: https://doi.org/10.1001/jamapediatrics.2022.3756

12. Burns JC, Kushner HI, Bastian JF, Shike H, Shimizu C, Matsubara T, Turner CL. Kawasaki Disease: A Brief History. Pediatrics. 2000 Aug 1;106(2):e27.

13. Yanagawa H, Nakamura Y, Kawasaki T, Shigematsu I. NATIONWIDE EPIDEMIC OF KAWASAKI DISEASE IN JAPAN DURING WINTER OF 1985-86. The Lancet. 1986 Nov 15;328(8516):1138–9.

14. Yanagawa H, Shigematsu I. Epidemiological Features of Kawasaki Disease in Japan. Pediatr Int. 1983;25(2):94–107.

15. Shih WL, Huang LM, Wu MH, Chang LY. Secular trend of Kawasaki disease and its correlation with viral activity in Taiwan: a nationwide population-based study. BMC Public Health. 2024 Jun 13;24(1):1591.

16. Oh KJ, Lee SY. Decreased incidence of Kawasaki disease in South Korea during the SARS-CoV-2 pandemic. Front Pediatr [Internet]. 2024 Apr 3 [cited 2024 Oct 1];12. Available from: https://www.frontiersin.org/journals/pediatrics/articles/10.3389/fped.2024.1307931/full

17. Holman RC, Christensen KY, Belay ED, Steiner CA, Effler PV, Miyamura J, Forbes S, Schonberger LB, Melish M. Racial/Ethnic Differences in the Incidence of Kawasaki Syndrome among Children in Hawai'i. Hawaii Med J. 2010 Aug;69(8):194–7.

18. Singh S, Vignesh P, Burgner D. The epidemiology of Kawasaki disease: a global update. Arch Dis Child. 2015 Nov 1;100(11):1084–8.

19. Jiao F, Jindal AK, Pandiarajan V, Khubchandani R, Kamath N, Sabui T, Mondal R, Pal P, Singh S. The emergence of Kawasaki disease in India and China. Glob Cardiol Sci Pract. 2017(3):e201721.

20. Mat Bah MN, Alias EY, Razak H, Sapian MH, Foo FH, Abdullah N. Epidemiology, clinical characteristics, and immediate outcome of Kawasaki disease: a population-based study from a tropical country. Eur J Pediatr. 2021 Aug 1;180(8):2599–606.

21. Celis-Seposo AK, Madaniyazi L, Seposo X, Hashizume M, Yoshida LM, Toizumi M. Incidence and seasonality of Kawasaki disease in children in the Philippines, and its association with ambient air temperature. Front Pediatr [Internet]. 2024 Apr 22 [cited 2024 Oct 1];12. Available from: https://www.frontiersin.org/journals/pediatrics/articles/10.3389/fped.2024.1358638/full

22. Maddox RA, Person MK, Joseph LJ, Haberling DL, Steiner CA, Schonberger LB, Belay ED. Abstract O.03: Monitoring the Occurrence of Kawasaki Syndrome in the United States. Circulation. 2015 Apr 28;131(suppl_2):AO03.

23. Piram M. Epidemiology of Kawasaki Disease in Europe. Front Pediatr [Internet]. 2021 May 25 [cited 2024 Oct 1];9. Available from: https://www.frontiersin.org/journals/pediatrics/articles/10.3389/fped.2021.673554/full

24. Saundankar J, Yim D, Itotoh B, Payne R, Maslin K, Jape G, Ramsay J, Kothari D, Cheng A, Burgner D. The epidemiology and clinical features of Kawasaki disease in Australia. Pediatrics. 2014 Apr;133(4):e1009-1014.

25. Burns JC, Cayan DR, Tong G, Bainto EV, Turner CL, Shike H, Kawasaki T, Nakamura Y, Yashiro M, Yanagawa H. Seasonality and Temporal Clustering of Kawasaki Syndrome. Epidemiology. 2005 Mar;16(2):220.

26. Sánchez-Manubens J, Antón J, Bou R, Iglesias E, Calzada-Hernandez J. Paediatric rheumatology Incidence, epidemiology and clinical features of Kawasaki disease in Catalonia, Spain. Clin Exp Rheumatol. 2016;34(97):S139–44.

27. Manlhiot C, O'Shea S, Bernknopf B, LaBelle M, Chahal N, Dillenburg RF, Lai LS, Bock D, Lew B, Masood S, Mathew M, McCrindle BW. Epidemiology of Kawasaki Disease in

Canada 2004 to 2014: Comparison of Surveillance Using Administrative Data vs Periodic Medical Record Review. Can J Cardiol. 2018 Mar 1;34(3):303–9.

28. Hearn J, McCrindle BW, Mueller B, O'Shea S, Bernknopf B, Labelle M, Manlhiot C. Spatiotemporal clustering of cases of Kawasaki disease and associated coronary artery aneurysms in Canada. Sci Rep. 2018 Dec 5;8(1):17682.

29. Burns JC, Herzog L, Fabri O, Tremoulet AH, Rodó X, Uehara R, Burgner D, Bainto E, Pierce D, Tyree M, Cayan D, Consortium for the KDGC. Seasonality of Kawasaki Disease: A Global Perspective. PLOS ONE. 2013 Sep 18;8(9):e74529.

30. Burns JC. The etiologies of Kawasaki disease. J Clin Invest [Internet]. 2024 Mar 1 [cited 2024 Sep 28];134(5). Available from: https://www.jci.org/articles/view/176938#B58

31. Rowley AH. Is Kawasaki disease an infectious disorder? Int J Rheum Dis. 2018;21(1):20–5.

32. Chang LY, Lu CY, Shao PL, Lee PI, Lin MT, Fan TY, Cheng AL, Lee WL, Hu JJ, Yeh SJ, Chang CC, Chiang BL, Wu MH, Huang LM. Viral infections associated with Kawasaki disease. J Formos Med Assoc Taiwan Yi Zhi. 2014 Mar;113(3):148–54.

33. Turnier JL, Anderson MS, Heizer HR, Jone PN, Glodé MP, Dominguez SR. Concurrent Respiratory Viruses and Kawasaki Disease. Pediatrics. 2015 Sep;136(3):e609-614.

34. Hara T, Furuno K, Yamamura K, Kishimoto J, Mizuno Y, Murata K, Onoyama S, Hatae K, Takemoto M, Ishizaki Y, Kanno S, Sato K, Motomura Y, Sakai Y, Ohga S, Yashiro M, Nakamura Y, Hara T. Assessment of Pediatric Admissions for Kawasaki Disease or Infectious Disease During the COVID-19 State of Emergency in Japan. JAMA Netw Open. 2021 Apr 6;4(4):e214475.

35. Kang JM, Jung J, Kim YE, Huh K, Hong J, Kim DW, Kim MY, Jung SY, Kim JH, Ahn JG. Temporal Correlation Between Kawasaki Disease and Infectious Diseases in South Korea. JAMA Netw Open. 2022 Feb 7;5(2):e2147363.

36. Neubauer HC, Lopez MA, Haq HA, Ouellette L, Ramirez AA, Wallace SS. Viral Coinfections in Kawasaki Disease: A Meta-analysis. Hosp Pediatr. 2023 May 12;13(6):e153–69.

37. Burns JC, Hsieh LE, Kumar J, Behnamfar N, Shimizu C, Sivilay N, Tremoulet AH, Franco A. Characterization of circulating immune cells in acute Kawasaki disease suggests exposure to different antigens. Clin Exp Immunol. 2020 Dec;202(3):263–72.

38. Rodó X, Curcoll R, Robinson M, Ballester J, Burns JC, Cayan DR, Lipkin WI, Williams BL, Couto-Rodriguez M, Nakamura Y, Uehara R, Tanimoto H, Morguí JA. Tropospheric winds from northeastern China carry the etiologic agent of Kawasaki disease from its source to Japan. Proc Natl Acad Sci U S A. 2014 Jun 3;111(22):7952–7.

39. Rodó X, Ballester J, Curcoll R, Boyard-Micheau J, Borràs S, Morguí JA. Revisiting the role of environmental and climate factors on the epidemiology of Kawasaki disease. Ann N Y Acad Sci. 2016 Oct 1;1382(1):84–98.

40. Zhao WM, Wang ZJ, Shi R, Zhu YY, Zhang S, Wang RF, Wang DG. Environmental factors influencing the risk of ANCA-associated vasculitis. Front Immunol. 2022;13:991256.

41. Scott J, Hartnett J, Mockler D, Little MA. Environmental risk factors associated with ANCA associated vasculitis: a systematic mapping review. Autoimmun Rev. 2020;19(11):102660.

42. Mulloy KB. Silica exposure and systemic vasculitis. Environ Health Perspect. 2003 Dec;111(16):1933–8.

43. Lane SE, Watts RA, Bentham G, Innes NJ, Scott DGI. Are environmental factors important in primary systemic vasculitis? A case-control study. Arthritis Rheum. 2003 Mar;48(3):814–23.

44. Willeke P, Schlüter B, Sauerland C, Becker H, Reuter S, Jacobi A, Schotte H. Farm Exposure as a Differential Risk Factor in ANCA-Associated Vasculitis. PLOS ONE. 2015 Sep 4;10(9):e0137196.

45. DeHaan LL, Copeland CD, Burney JA, Nakamura Y, Yashiro M, Shimizu C, Miyata K, Burns JC, Cayan DR. Age-Dependent Variations in Kawasaki Disease Incidence in Japan. JAMA Netw Open. 2024 Feb 6;7(2):e2355001.

46. Rodó X, Ballester J, Cayan D, Melish ME, Nakamura Y, Uehara R, Burns JC. Association of Kawasaki disease with tropospheric wind patterns. Sci Rep. 2011 Nov 10;1(1):152.

47. Ballester J, Burns JC, Cayan D, Nakamura Y, Uehara R, Rodó X. Kawasaki disease and ENSO-driven wind circulation. Geophys Res Lett. 2013;40(10):2284–9.

48. Jorquera H, Borzutzky A, Hoyos-Bachiloglu R, García A. Association of Kawasaki disease with tropospheric winds in Central Chile: Is wind-borne desert dust a risk factor? Environ Int. 2015 May 1;78:32–8.

49. Altizer S, Ostfeld RS, Johnson PTJ, Kutz S, Harvell CD. Climate change and infectious diseases: from evidence to a predictive framework. Science. 2013 Aug 2;341(6145):514–9.

50. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, Rohr JR, Ryan SJ, Savage V, Shocket MS, Sippy R, Stewart Ibarra AM, Thomas MB, Villena O. Thermal biology of mosquito-borne disease. Ecol Lett. 2019 Oct;22(10):1690–708.

51. Moriyama M, Hugentobler WJ, Iwasaki A. Seasonality of Respiratory Viral Infections. Annu Rev Virol. 2020 Sep 29;7(1):83–101.

52. Grassly NC, Fraser C. Seasonal infectious disease epidemiology. Proc Biol Sci. 2006 Oct 7;273(1600):2541–50.

53. Taubenberger JK, Morens DM. 1918 Influenza: the mother of all pandemics. Emerg Infect Dis. 2006 Jan;12(1):15–22.

54. Honigsbaum M. Spanish influenza redux: revisiting the mother of all pandemics. Lancet Lond Engl. 2018 Jun 23;391(10139):2492–5.

55. Akin L, Gözel MG. Understanding dynamics of pandemics. Turk J Med Sci. 2020 Apr 21;50(SI-1):515–9.

56. Tamerius J, Nelson MI, Zhou SZ, Viboud C, Miller MA, Alonso WJ. Global influenza seasonality: reconciling patterns across temperate and tropical regions. Environ Health Perspect. 2011 Apr;119(4):439–45.

57. Shaman J, Kohn M. Absolute humidity modulates influenza survival, transmission, and seasonality. Proc Natl Acad Sci U S A. 2009 Mar 3;106(9):3243–8.

58. Vittecoq M, Roche B, Cohen JM, Renaud F, Thomas F, Gauthier-Clerc M. Does the weather play a role in the spread of pandemic influenza? A study of H1N1pdm09 infections in France during 2009-2010. Epidemiol Infect. 2015 Dec;143(16):3384–93.

59. Sooryanarain H, Elankumaran S. Environmental Role in Influenza Virus Outbreaks. Annu Rev Anim Biosci. 2015 Feb 16;3(Volume 3, 2015):347–73.

60. Costa ACC da, Codeço CT, Krainski ET, Gomes MF da C, Nobre AA. Spatiotemporal diffusion of influenza A (H1N1): Starting point and risk factors. PLOS ONE. 2018 Sep 4;13(9):e0202832.

61. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 [Internet]. [cited 2024 Oct 4]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

62. COVID-19 cases | WHO COVID-19 dashboard [Internet]. datadot. [cited 2024 Oct 4]. Available from: https://data.who.int/dashboards/covid19/cases

63. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, Lau YC, Wong JY, Guan Y, Tan X, Mo X, Chen Y, Liao B, Chen W, Hu F, Zhang Q, Zhong M, Wu Y, Zhao L, Zhang F, Cowling BJ, Li F, Leung GM. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020 May;26(5):672–5.

64. Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. Temperature, Humidity and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19 [Internet]. Rochester, NY: Social Science Research Network; 2020 Mar [cited 2021 Sep 20]. Report No.: ID 3550308. Available from: https://papers.ssrn.com/abstract=3550308

65. COVID-19: High temperature and high humidity reduce the transmission of COVID-19. | The Centre for Evidence-Based Medicine [Internet]. [cited 2024 Oct 4]. Available from: https://www.cebm.net/study/covid-19-high-temperature-and-high-humidity-reduce-the-transmission-of-covid-19/

66. Casanova LM, Jeon S, Rutala WA, Weber DJ, Sobsey MD. Effects of air temperature and relative humidity on coronavirus survival on surfaces. Appl Environ Microbiol. 2010 May;76(9):2712–7.

67. Chan KH, Peiris JSM, Lam SY, Poon LLM, Yuen KY, Seto WH. The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. Adv Virol. 2011;2011:734690.

68. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, Huang J, He N, Yu H, Lin X, Wei S, Wu T. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA. 2020 May 19;323(19):1915–23.

69. Merow C, Urban MC. Seasonality and uncertainty in global COVID-19 growth rates. Proc Natl Acad Sci. 2020 Nov 3;117(44):27456–64.

70. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020 May 22;368(6493):860–8.

71. Neher RA, Dyrdak R, Druelle V, Hodcroft EB, Albert J. Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. Swiss Med Wkly. 2020 Mar 9;150:w20224.

72. Prather KA, Wang CC, Schooley RT. Reducing transmission of SARS-CoV-2. Science. 2020 Jun 26;368(6498):1422–4.

73. Kampf G, Todt D, Pfaender S, Steinmann E. Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. J Hosp Infect. 2020 Mar;104(3):246–51.

74. Yu ITS, Li Y, Wong TW, Tam W, Chan AT, Lee JHW, Leung DYC, Ho T. Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. N Engl J Med. 2004 Apr 22;350(17):1731–9.

75. Lowen AC, Steel J. Roles of humidity and temperature in shaping influenza seasonality. J Virol. 2014 Jul;88(14):7692–5.

76. Leung NHL. Transmissibility and transmission of respiratory viruses. Nat Rev Microbiol. 2021 Aug;19(8):528–45.

77. Després VR, Huffman JA, Burrows SM, Hoose C, Safatov AS, Buryak G, Fröhlich-Nowoisky J, Elbert W, Andreae MO, Pöschl U, Jaenicke R. Primary biological aerosol particles in the atmosphere: a review. Tellus B Chem Phys Meteorol [Internet]. 2012 Jan 1 [cited 2024 Aug 19];64(1). Available from: https://b.tellusjournals.se/articles/10.3402/tellusb.v64i0.15598

78. Fröhlich-Nowoisky J, Kampf CJ, Weber B, Huffman JA, Pöhlker C, Andreae MO, Lang-Yona N, Burrows SM, Gunthe SS, Elbert W, Su H, Hoor P, Thines E, Hoffmann T, Després VR, Pöschl U. Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. Atmospheric Res. 2016 Dec 15;182:346–76.

79. Bowers RM, McLetchie S, Knight R, Fierer N. Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. ISME J. 2011 Apr;5(4):601–12.

80. Womack AM, Bohannan BJM, Green JL. Biodiversity and biogeography of the atmosphere. Philos Trans R Soc Lond B Biol Sci. 2010 Nov 27;365(1558):3645–53.

81. Kellogg CA, Griffin DW. Aerobiology and the global transport of desert dust. Trends Ecol Evol. 2006 Nov;21(11):638–44.

82. Cox CS, Wathes CM, editors. Bioaerosols Handbook. Boca Raton: CRC Press; 2020. 656 p.

83. Cole EC, Cook CE. Characterization of infectious aerosols in health care facilities: An aid to effective engineering controls and preventive strategies. Am J Infect Control. 1998 Aug 1;26(4):453–64.

84. Dillon CF, Dillon MB. Multiscale Airborne Infectious Disease Transmission. Appl Environ Microbiol. 2021 Jan 29;87(4):e02314-20.

85. Tang L, Rhoads WJ, Eichelberg A, Hamilton KA, Julian TR. Applications of Quantitative Microbial Risk Assessment to Respiratory Pathogens and Implications for Uptake in Policy: A State-of-the-Science Review. Environ Health Perspect. 2024 May;132(5):056001.

86. Nasir ZA, Campos LC, Christie N, Colbeck I. Airborne biological hazards and urban transport infrastructure: current challenges and future directions. Environ Sci Pollut Res. 2016 Aug 1;23(15):15757–66.

87. D'Amato G, Cecchi L, Bonini S, Nunes C, Annesi-Maesano I, Behrendt H, Liccardi G, Popov T, van Cauwenberge P. Allergenic pollen and pollen allergy in Europe. Allergy. 2007 Sep;62(9):976–90.

88. Schmidt CW. Pollen Overload: Seasonal Allergies in a Changing Climate. Environ Health Perspect. 2016 Apr;124(4):A70–5.

89. von Wintersdorff CJH, Penders J, van Niekerk JM, Mills ND, Majumder S, van Alphen LB, Savelkoul PHM, Wolffs PFG. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. Front Microbiol [Internet]. 2016 Feb 19 [cited 2024 Oct 3];7. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2016.00173/full

90. Burrows SM, Elbert W, Lawrence MG, Pöschl U. Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems. Atmospheric Chem Phys. 2009 Dec 10;9(23):9263–80.

91. Stull RB, editor. An Introduction to Boundary Layer Meteorology [Internet]. Dordrecht: Springer Netherlands; 1988 [cited 2024 Oct 3]. Available from: http://link.springer.com/10.1007/978-94-009-3027-8

92. Jaenicke R. Abundance of cellular material and proteins in the atmosphere. Science. 2005 Apr 1;308(5718):73.

93. Fahlgren C, Hagström Å, Nilsson D, Zweifel UL. Annual Variations in the Diversity, Viability, and Origin of Airborne Bacteria. Appl Environ Microbiol. 2010 May;76(9):3015–25.

94. Cox MM, Battista JR. Deinococcus radiodurans — the consummate survivor. Nat Rev Microbiol. 2005 Nov;3(11):882–92.

95. Smith DJ, Jaffe DA, Birmele MN, Griffin DW, Schuerger AC, Hee J, Roberts MS. Free Tropospheric Transport of Microorganisms from Asia to North America. Microb Ecol. 2012 Nov 1;64(4):973–85.

96. Prospero JM, Mayol-Bracero OL. Understanding the Transport and Impact of African Dust on the Caribbean Basin. Bull Am Meteorol Soc. 2013 Sep 1;94(9):1329–37.

97. Garrison VH, Shinn EA, Foreman WT, Griffin DW, Holmes CW, Kellogg CA, Majewski MS, Richardson LL, Ritchie KB, Smith GW. African and Asian Dust: From Desert Soils to Coral Reefs. BioScience. 2003 May 1;53(5):469–80.

98. Griffin DW. Atmospheric Movement of Microorganisms in Clouds of Desert Dust and Implications for Human Health. Clin Microbiol Rev. 2007 Jul;20(3):459–77.

99. DeLeon-Rodriguez N, Lathem TL, Rodriguez-R LM, Barazesh JM, Anderson BE, Beyersdorf AJ, Ziemba LD, Bergin M, Nenes A, Konstantinidis KT. Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. Proc Natl Acad Sci. 2013 Feb 12;110(7):2575–80.

100. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 1995 Mar;59(1):143–69.

101. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. Annu Rev Genet. 2004;38:525–52.

102. Behzad H, Gojobori T, Mineta K. Challenges and Opportunities of Airborne Metagenomics. Genome Biol Evol. 2015 May 1;7(5):1216–26.

103. Gusareva ES, Acerbi E, Lau KJX, Luhung I, Premkrishnan BNV, Kolundžija S, Purbojati RW, Wong A, Houghton JNI, Miller D, Gaultier NE, Heinle CE, Clare ME, Vettath VK, Kee C, Lim SBY, Chénard C, Phung WJ, Kushwaha KK, Nee AP, Putra A, Panicker D, Yanqing K, Hwee YZ, Lohar SR, Kuwata M, Kim HL, Yang L, Uchida A, Drautz-Moses DI, Junqueira ACM, Schuster SC. Microbial communities in the tropical air ecosystem follow a precise diel cycle. Proc Natl Acad Sci. 2019 Nov 12;116(46):23299–308.

104. Yooseph S, Andrews-Pfannkoch C, Tenney A, McQuaid J, Williamson S, Thiagarajan M, Brami D, Zeigler-Allen L, Hoffman J, Goll JB, Fadrosh D, Glass J, Adams MD, Friedman R, Venter JC. A metagenomic framework for the study of airborne microbial communities. PloS One. 2013;8(12):e81862.

105. Hairston PP, Ho J, Quant FR. Design of an instrument for real-time detection of bioaerosols using simultaneous measurement of particle aerodynamic size and intrinsic fluorescence. J Aerosol Sci. 1997 Apr 1;28(3):471–82.

106. Ho J. Future of biological aerosol detection. Anal Chim Acta. 2002 Apr 15;457(1):125–48.

107. Pan YL. Detection and characterization of biological and other organic-carbon aerosol particles in atmosphere using fluorescence. J Quant Spectrosc Radiat Transf. 2015 Jan 1;150:12–35.

108. Kaye PH, Stanley WR, Hirst E, Foot EV, Baxter KL, Barrington SJ. Single particle multichannel bio-aerosol fluorescence sensor. Opt Express. 2005 May 16;13(10):3583–93.

109. Kawashima S, Clot B, Fujita T, Takahashi Y, Nakamura K. An algorithm and a device for counting airborne pollen automatically using laser optics. Atmos Environ. 2007 Nov 1;41(36):7987–93.

110. Oteros J, Weber A, Kutzora S, Rojo J, Heinze S, Herr C, Gebauer R, Schmidt-Weber CB, Buters JTM. An operational robotic pollen monitoring network based on automatic image recognition. Environ Res. 2020 Dec 1;191:110031.

111. Šaulienė I, Šukienė L, Daunys G, Valiulis G, Vaitkevičius L, Matavulj P, Brdar S, Panic M, Sikoparija B, Clot B, Crouzy B, Sofiev M. Automatic pollen recognition with the Rapid-E

particle counter: the first-level procedure, experience and next steps. Atmospheric Meas Tech. 2019 Jun 28;12(6):3435–52.

112. Matavulj P, Panić M, Šikoparija B, Tešendić D, Radovanović M, Brdar S. Advanced CNN Architectures for Pollen Classification: Design and Comprehensive Evaluation. Appl Artif Intell. 2023 Dec 31;37(1):2157593.

113. Huffman JA, Perring AE, Savage NJ, Clot B, Crouzy B, Tummon F, Shoshanim O, Damit B, Schneider J, Sivaprakasam V, Zawadowicz MA, Crawford I, Gallagher M, Topping D, Doughty DC, Hill SC, Pan Y. Real-time sensing of bioaerosols: Review and current perspectives. Aerosol Sci Technol. 2020 May 3;54(5):465–95.

114. Maya-Manzano JM, Tummon F, Abt R, Allan N, Bunderson L, Clot B, Crouzy B, Daunys G, Erb S, Gonzalez-Alonso M, Graf E, Grewling Ł, Haus J, Kadantsev E, Kawashima S, Martinez-Bracero M, Matavulj P, Mills S, Niederberger E, Lieberherr G, Lucas RW, O'Connor DJ, Oteros J, Palamarchuk J, Pope FD, Rojo J, Šaulienė I, Schäfer S, Schmidt-Weber CB, Schnitzler M, Šikoparija B, Skjøth CA, Sofiev M, Stemmler T, Triviño M, Zeder Y, Buters J. Towards European automatic bioaerosol monitoring: Comparison of 9 automatic pollen observational instruments with classic Hirst-type traps. Sci Total Environ. 2023 Mar 25;866:161220.

115. Pöhlker C, Huffman JA, Pöschl U. Autofluorescence of atmospheric bioaerosols – fluorescent biomolecules and potential interferences. Atmospheric Meas Tech. 2012 Jan 9;5(1):37–71.

116. Gabbarini V, Rossi R, Ciparisse JF, Malizia A, Divizia A, De Filippis P, Anselmi M, Carestia M, Palombi L, Divizia M, Gaudio P. Laser-induced fluorescence (LIF) as a smart method for fast environmental virological analyses: validation on Picornaviruses. Sci Rep. 2019 Aug 29;9(1):12598.

117. Moreno T, Kojima T, Amato F, Lucarelli F, De La Rosa J, Calzolai G, Nava S, Chiari M, Alastuey A, Querol X. Daily and hourly chemical impact of springtime transboundary aerosols on Japanese air quality. Atmospheric Chem Phys. 2013;13(3):1411–24.

118. Vautard R, Yiou P, Ghil M. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. Phys Nonlinear Phenom. 1992 Sep 15;58(1):95–126.

119. Cleveland WS, Graedel TE, Kleiner B, Warner JL. Sunday and workday variations in photochemical air pollutants in new jersey and new york. Science. 1974 Dec 13;186(4168):1037–8.

120. Beirle S, Platt U, Wenig M, Wagner T. Weekly cycle of NO_2 by GOME measurements: a signature of anthropogenic sources. Atmospheric Chem Phys. 2003 Dec 16;3(6):2225–32.

121. Cheng X, Huang Y, Zhang SP, Ni SJ, Long ZJ. Characteristics, Sources, and Health Risk Assessment of Trace Elements in PM10 at an Urban Site in Chengdu, Southwest China. Aerosol Air Qual Res. 2018;18(2):357–70.

122. Mills NL, Donaldson K, Hadoke PW, Boon NA, MacNee W, Cassee FR, Sandström T, Blomberg A, Newby DE. Adverse cardiovascular effects of air pollution. Nat Clin Pract Cardiovasc Med. 2009;6(1):36–44.

123. Davis WT, Fu JS, Godish T. Air Quality. CRC Press; 2021.

124. Hollingsworth JW, Maruoka S, Li Z, Potts EN, Brass DM, Garantziotis S, Fong A, Foster WM, Schwartz DA. Ambient ozone primes pulmonary innate immunity in mice. J Immunol. 2007;179(7):4367–75.

125. Miyata R, van Eeden SF. The innate and adaptive immune response induced by alveolar macrophages exposed to ambient particulate matter. Toxicol Appl Pharmacol. 2011;257(2):209–26.

126. Park M, Joo HS, Lee K, Jang M, Kim SD, Kim I, Borlaza LJS, Lim H, Shin H, Chung KH. Differential toxicities of fine particulate matters from various sources. Sci Rep. 2018;8(1):1–11.

127. Zeft AS, Burns JC, Yeung RS, McCrindle BW, Newburger JW, Dominguez SR, Anderson MS, Arrington C, Shulman ST, Yoon J, Tewelde H, Mix C, Pope CA. Kawasaki Disease and Exposure to Fine Particulate Air Pollution. J Pediatr. 2016 Oct 1;177:179-183.e1.

128. Buteau S, Belkaibech S, Bilodeau-Bertrand M, Hatzopoulou M, Smargiassi A, Auger N. Association between Kawasaki Disease and Prenatal Exposure to Ambient and Industrial Air Pollution: A Population-Based Cohort Study. Environ Health Perspect. 2020 Oct;128(10):107006.

129. Jung CR, Chen WT, Lin YT, Hwang BF. Ambient Air Pollutant Exposures and Hospitalization for Kawasaki Disease in Taiwan: A Case-Crossover Study (2000–2010). Environ Health Perspect. 2017 Apr;125(4):670–6.

130. Giefing-Kröll C, Berger P, Lepperdinger G, Grubeck-Loebenstein B. How sex and age affect immune responses, susceptibility to infections, and response to vaccination. Aging Cell. 2015 Jun;14(3):309–21.

131. Klein SL, Flanagan KL. Sex differences in immune responses. Nat Rev Immunol. 2016 Oct;16(10):626–38.

132. Ozeki Y, Yamada F, Saito A, Kishimoto T, Yashiro M, Makino N, Nakamura Y. Epidemiologic features of Kawasaki disease distinguished by seasonal variation: an age-specific analysis. Ann Epidemiol. 2018 Nov 1;28(11):796–800.

133. Rodríguez-Arias MÀ, Rodó X. A primer on the study of transitory dynamics in ecological series using the scale-dependent correlation analysis. Oecologia. 2004 Mar 1;138(4):485–504.

134. Rodo X, Rodriguez Arias MA. A new method to detect transitory signatures and local time/space variability structures in the climate system: The scale-dependent correlation analysis. Clim Dyn. 2006 Oct 1;27:441–58.

135. Neves JMM, Belo VS, Catita CMS, de Oliveira BFA, Horta MAP. Modeling the Climatic Suitability of COVID-19 Cases in Brazil. Trop Med Infect Dis. 2023 Apr;8(4):198.

136. Nottmeyer L, Armstrong B, Lowe R, Abbott S, Meakin S, O'Reilly KM, von Borries R, Schneider R, Royé D, Hashizume M, Pascal M, Tobias A, Vicedo-Cabrera AM, Lavigne E, Correa PM, Ortega NV, Kynčl J, Urban A, Orru H, Ryti N, Jaakkola J, Dallavalle M, Schneider A, Honda Y, Ng CFS, Alahmad B, Carrasco-Escobar G, Holobâc IH, Kim H, Lee W, Íñiguez C, Bell ML, Zanobetti A, Schwartz J, Scovronick N, Coélho M de SZS, Saldiva PHN, Diaz MH, Gasparrini A, Sera F. The association of COVID-19 incidence

with temperature, humidity, and UV radiation – A global multi-city analysis. Sci Total Environ. 2023 Jan 1;854:158636.

137. Mao N, Zhang D, Li Y, Li Y, Li J, Zhao L, Wang Q, Cheng Z, Zhang Y, Long E. How do temperature, humidity, and air saturation state affect the COVID-19 transmission risk? Environ Sci Pollut Res. 2023 Jan 1;30(2):3644–58.

138. Sakib SN, Islam ARMdT, Azad MdAK, Mallick J, Ahmed Mohd, Pal SC, Islam MdS, Hu Z, Alam E, Malafaia G. Seasonality of meteorological factors influencing the COVID-19 era in coastal and inland regions of Bangladesh. Geocarto Int. 2023 Dec 31;38(1):2203115.

139. Liu Y, Rocklöv J. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. J Travel Med. 2021 Oct 11;28(7):taab124.

140. Català M, Coma E, Alonso S, Andrés C, Blanco I, Antón A, Bordoy AE, Cardona PJ, Fina F, Martró E, Medina M, Mora N, Saludes V, Prats C, Prieto-Alhambra D, Alvarez-Lacalle E. Transmissibility, hospitalization, and intensive care admissions due to omicron compared to delta variants of SARS-CoV-2 in Catalonia: A cohort study and ecological analysis. Front Public Health [Internet]. 2022 Aug 12 [cited 2024 Oct 5];10. Available from: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.961030/full

141. Ren SY, Wang WB, Gao RD, Zhou AM. Omicron variant (B.1.1.529) of SARS-CoV-2: Mutation, infectivity, transmission, and vaccine resistance. World J Clin Cases. 2022 Jan 7;10(1):1–11.

142. Milne G, Hames T, Scotton C, Gent N, Johnsen A, Anderson RM, Ward T. Does infection with or vaccination against SARS-CoV-2 lead to lasting immunity? Lancet Respir Med. 2021 Dec 1;9(12):1450–66.

143. Morris DH, Yinda KC, Gamble A, Rossine FW, Huang Q, Bushmaker T, Fischer RJ, Matson MJ, Van Doremalen N, Vikesland PJ, Marr LC, Munster VJ, Lloyd-Smith JO. Mechanistic theory predicts the effects of temperature and humidity on inactivation of SARS-CoV-2 and other enveloped viruses. Garrett WS, Ogbunugafor CB, Handel A, editors. eLife. 2021 Apr 27;10:e65902.

144. Allen JG, Ibrahim AM. Indoor Air Changes and Potential Implications for SARS-CoV-2 Transmission. JAMA. 2021 May 25;325(20):2112–3.

145. Rodó X, San-José A, Kirchgatter K, López L. Changing climate and the COVID-19 pandemic: more than just heads or tails. Nat Med. 2021 Apr;27(4):576–9.

146. Zhou Z, Shuai X, Lin Z, Yu X, Ba X, Holmes MA, Xiao Y, Gu B, Chen H. Association between particulate matter (PM)2·5 air pollution and clinical antibiotic resistance: a global analysis. Lancet Planet Health. 2023 Aug 1;7(8):e649–59.

147. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. Handb Brain Theory Neural Netw. 1995;3361(10):1995.