Temperature, Air Pollution, and Perinatal Health in the Context of Climate Change : Exposure and Impacts in Sensitive Populations

Ian Hough

THÈSE pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Préparée dans le cadre d'une cotutelle entre l'*Université Grenoble Alpes (UGA)* & l'*Université Ben-Gurion du Néguev (BGU)*

Spécialités : **Modèles, méthodes et algorithmes en biologie, santé et environnement & Geographic information technologies**

Arrêté ministériel : 25 mai 2016

Présentée par

Ian HOUGH

Codirigée par **Johanna LEPEULE** et **Itai KLOOG**

Préparée au sein des **Institut pour l'Avancée des Biosciences** & **Département de Géographie et du Développement Environnemental**

Dans les Écoles Doctorales **Ingénierie pour la Santé la Cognition et l'Environnement** & **École Kreitman des Études Supérieures**

Température, pollution atmosphérique et santé périnatale dans le contexte du changement climatique : exposition et impacts dans les populations sensibles

Thèse soutenue publiquement le 16 décembre 2021,

devant le jury composé de :

M. Kees DE HOOGH
Directeur de groupe, Swiss TPH, Rapporteur

Mme Mathilde PASCAL
Chargée de projets scientifiques, Santé Publique France, Rapportrice

Mme Maayan YITSHAK SADE
Professeure assistante, École de Médecine Icahn au Mont Sinai, Membre

M. Jean-Luc JAFFREZO
Professeur, Université Grenoble Alpes, Membre

Mme Adeline LECLERCQ SAMSON
Professeure, Université Grenoble Alpes, Présidente

M. Massimo STAFOGGIA
Épidémiologiste sénior, Service de Santé de la Région Lazio, Invité

M. Itai KLOOG
Professeur, Université Ben-Gurion du Néguev, Co-directeur de thèse

Temperature, Air Pollution, and Perinatal Health in the Context of Climate Change: Exposure and Impacts in Sensitive Populations

Thesis submitted in partial fulfilment
of the requirements for the degree of
“Doctor of Philosophy”

by

Ian Hough

2021-12-16

Submitted jointly to

The Kreitman School of Advanced Graduate Studies
Ben-Gurion University of the Negev

The Doctoral School of Health, Cognition, and Environmental Engineering
Université Grenoble Alpes

Prepared under the supervision of

Dr. Itai Kloog
Department of Geography and Environmental Development
Ben-Gurion University of the Negev, Be’er Sheva, Israel

Dr. Johanna Lepeule
Institute for Advanced Biosciences, INSERM, CNRS
Université Grenoble Alpes, Grenoble, France

Research-Student's Affidavit when Submitting the Doctoral Thesis for Judgment

I **Ian Hough**, whose signature appears below, hereby declare that

[X] I have written this Thesis by myself, except for the help and guidance offered by my Thesis Advisors.

[X] The scientific materials included in this Thesis are products of my own research, culled from the period during which I was a research student.

[ ] This Thesis incorporates research materials produced in cooperation with others, excluding the technical help commonly received during experimental work. Therefore, I am attaching another affidavit stating the contributions made by myself and the other participants in this research, which has been approved by them and submitted with their approval.

Date: **2021-10-19** Student's name: **Ian Hough** Signature: 

Acknowledgements

This work would not have been possible without the guidance and support of my supervisors, Itai and Johanna. I am very grateful for their patience and advice. I would also like to thank my advisory committee, Allan Just, Mauro Dalla-Mura, Jean-Luc Jaffrezo, Evyatar Erell, and Victor Novak, for their helpful discussions. And thanks to Ben-Gurion University of the Negev and Université Grenoble Alpes for their financial and administrative support.

Many thanks to Alex, Bin, and Ron of the Environmental Exposure Assessment Lab for their enriching discussions and friendship, and Michael, Noam, Oron, Vlad, and the staff of the BGU geography department for their warm welcome. Merci également au bureau 103 (Alicia, Dorothy, Pau, Stephan, Solène, ...) for the camaraderie and divertissement. Et merci à tout l'équipe 12 pour les bons moments passés ensemble des petits déjs aux défis pirate.

Finally, thanks to Manon, Moira, Janet, John, and all the friends and family that inspired and encouraged me on this journey.

Abstract

Studies suggest that exposure to heat, cold, or particulate matter (PM) during pregnancy may decrease birth weight and shorten the duration of gestation, but we do not fully understand these associations. This is concerning because low birth weight and preterm birth increase risk for numerous morbidities and chronic diseases throughout childhood and into adulthood. The health burdens of temperature and PM are expected to increase with climate change and global urbanization. Improving our understanding requires accurate exposure assessment because error can bias associations with health outcomes. A scalable way to assess exposure over large areas and long time periods is to model the relationship between monitored temperature or PM and satellite-derived variables such as land surface temperature (LST) or aerosol optical depth (AOD). Improved calibration methods might increase accuracy, and finer spatial resolutions could be useful in urban areas.

This PhD developed novel approaches to increase the accuracy and spatial resolution of satellite-based exposure models, applied them to reconstruct daily temperature and PM in France over the past two decades, and used the temperatures to identify periods during pregnancy when exposure to cold or heat may increase the risk of preterm birth.

First, we found that allowing the relationship between monitored ambient temperature and satellite-derived LST to vary over space as well as time improves the accuracy of mixed models. We used this approach to model diurnal air temperature (mean [T_{mean}], minimum [T_{min}], and maximum [T_{max}]) over France at 1 km resolution from 2000 to 2018. We also showed that an ensemble of random forests and gradient boosting machines can downscale daily temperature over urban areas by incorporating thermal satellite data at a high spatial but low temporal resolution. We used this approach to model daily T_{mean}, T_{min}, and T_{max} at 200 m resolution over 103 cities.

Second, we found that Gaussian Markov random fields can reconstruct PM based on gap-filled AOD more accurately than widely used mixed models or random forests, and that an ensemble of all three algorithms performs even better. We used this approach to model daily mean $PM_{2.5}$ and PM_{10} (PM <2.5 and <10 microns in diameter, respectively) concentration over France at 1 km resolution from 2000 to 2019 and showed that alleviating the sparsity of $PM_{2.5}$ monitors by imputing $PM_{2.5}$ at more common PM_{10} monitors increased accuracy.

Third, we conducted a survival analysis using Cox proportional hazards models with distributed lags to assess the time-varying association between residence-based ambient temperature estimated by our exposure model and the risk of preterm birth among 5347 singleton live births (4.3% preterm) from prospective pregnancy cohorts in France. We simultaneously examined chronic exposure (weekly from conception) and acute exposure (daily during the 30 days before delivery), and we explored the effects of temperature variability and acclimation to location and season during heatwaves. Cold and night-time heat increased the risk of preterm birth, with susceptibility starting as early as conception and continuing through portions of trimesters 1 and 2. Cold around 7 days before delivery also increased the risk of preterm birth.

Our approaches could improve satellite-based exposure models in other areas and our multi-decadal dataset will be of interest to epidemiologists, climatologists, planners, policymakers, and the public. In the context of rising temperatures and more frequent weather hazards, our findings on the risks that heat and cold pose to pregnant women and their infants should inform public health policies to reduce the growing burden of preterm birth.

Résumé (French abstract)

Bien que plusieurs études suggèrent que l'exposition à la chaleur, au froid ou aux particules en suspension (PM) pendant la grossesse pourraient diminuer le poids de naissance et raccourcir la durée de la grossesse, ces effets ne sont pas complètement compris. Ceci est préoccupant car la prématurité et un faible poids à la naissance augmentent le risque de nombreuses morbidités et maladies chroniques tout au long de l'enfance et à l'âge adulte. Ces fardeaux sanitaires devraient s'accroître avec le changement climatique et l'urbanisation mondiale. Mieux comprendre ces relations nécessite une évaluation précise de l'exposition, car l'erreur peut affaiblir les associations avec la santé. Une approche efficace consiste à modéliser la relation entre des mesures de la température ou des PM et des variables dérivées des satellites. Des nouvelles approches pourraient réduire l'erreur, et des résolutions spatiales plus fines pourraient être utiles, notamment dans les zones urbaines.

Ce doctorat a développé de nouvelles approches pour augmenter la précision et la résolution spatiale des modèles d'exposition, les a appliqués pour reconstruire la température quotidienne et les PM en France au cours des deux dernières décennies, et a utilisé les températures pour identifier des périodes pendant la grossesse où l'exposition au froid ou à la chaleur peut augmenter le risque de naissance prématurée.

Premièrement, nous avons démontré que le fait de prendre en compte la variabilité spatiale ainsi que temporelle de la relation entre la température de l'air mesurée et les variables dérivées de satellites améliore la précision de modèles de prédiction. Nous avons également démontré qu'un ensemble de forêts aléatoires et de machines de renforcement de gradient permet d'améliorer la résolution spatiale des estimations quotidiennes pour les zones urbaines en incorporant des données satellitaires thermiques à haute résolution spatiale mais à faible résolution temporelle. Avec ces approches, nous avons modélisé la température quotidienne moyenne, minimale et maximale de 2000 à 2018 à une résolution de 1 km en France et de 200 m sur 103 zones urbaines.

Deuxièmement, nous avons démontré que les champs aléatoires gaussiens de Markov peuvent estimer les PM plus précisément que les modèles mixtes ou les forêts aléatoires, et qu'un ensemble des trois est encore plus performant. Avec cette approche, nous avons modélisé les concentrations moyennes quotidiennes des PM en France à une résolution de 1 km de 2000 à 2019.

Troisièmement, nous avons mené une analyse de survie avec un modèle de Cox à effets distribués dans le temps pour évaluer l'association entre la température résidentielle issu de notre modèle d'exposition et le risque de naissance prématurée parmi 5347 naissances (4,3% prématurés) suivies dans le cadre de cohortes prospectives en France. Nous avons examiné de manière concomitante des fenêtres d'exposition chroniques (les 26 semaines suivant la conception) et aigues (les 30 jours avant l'accouchement) et avons exploré les effets de la variabilité de la température et des vagues de chaleur en tenant compte de l'acclimatation. Le froid et la chaleur nocturne augmentaient le risque d'accouchement prématuré ; les femmes étaient sensibles dès la conception et pendant les secondes parties des trimestres 1 et 2. Le froid augmentait également le risque de naissance prématurée environ sept jours plus tard.

Nos approches novatrices pourraient permettre d'améliorer les modèles d'exposition dans d'autres régions et notre jeu de données multi-décennale sera utile aux épidémiologistes, climatologues, planificateurs, décideurs et au public. Dans le contexte de la hausse des températures et des aléas météorologiques plus fréquents, nos découvertes sur les risques que la chaleur et le froid représentent pour les femmes enceintes et leurs bébés devraient éclairer les politiques de santé publique pour réduire le fardeau croissant de la naissance prématurée.

Table of Contents

List of abbreviations

AIC – Akaike information criterion

AOD – Aerosol optical depth

BMI – Body mass index

CI – Confidence interval

CLC – Corine Land Cover

CV – Cross-validation

DLNM – Distributed lag nonlinear model

EAC4 – ECMWF Atmospheric Composition Reanalysis, version 4

ECMWF – European Centre for Medium-Range Weather Forecasts

EDEN – Étude des déterminants pré et post natals du développement et de la santé de l'enfant

EHF – Excess Heat Factor

ERA5 – ECMWF Reanalysis, version 5

ETM+ – Enhanced Thematic Mapper Plus

ESA – European Space Agency

GAM – Generalized additive model

GMRF – Gaussian Markov random field

GOES – Geostationary Operational Environmental Satellite

HR – Hazard ratio

IDW – Inverse distance weighting

INLA – Integrated Nested Laplace Approximations

IQR – Interquartile range

LMM – Linear mixed model

LMP – Last menstrual period

LST – Land surface temperature

MAE – Mean squared error

MAIAC – Multi-Angle Implementation of Atmospheric Correction

MERRA-2 – Modern-Era Retrospective Reanalysis, version 2

MODIS – Moderate Resolution Imaging Spectroradiometer

MSG – Meteosat Second Generation

MTSAT – Multi-Functional Transport Satellite

NASA – National Aeronautics and Space Administration

NDVI – Normalized difference vegetation index

OR – Odds ratio

PELAGIE – Perturbateurs Endocriniens : étude Longitudinale sur les Anomalies de la Grossesse, l'Infertilité, et l'Enfance

PM – Particulate matter

PM_{10} – Particulate matter with a diameter <10 microns

$PM_{2.5}$ – Particulate matter with a diameter <2.5 microns

R^2 – Coefficient of determination

RF – Random forest

RMSE – Root mean squared error

RR – Relative risk

R_{Tmax} – Residual of predicted maximum temperature

R_{Tmean} – Residual of predicted mean temperature

R_{Tmin} – Residual of predicted minimum temperature

SD – Standard deviation

SEPAGES – Suivi de l'Exposition à la Pollution Atmosphérique durant le Grossesse et Effet sur la Santé

SEVIRI – Spinning Enhanced Visible and Infra-red Ranger Instrument

Suomi NPP – Suomi National Polar-orbiting Partnership satellite

T_a – 2 m air temperature

T_b – Top-of-atmosphere brightness temperature

TIRS – Thermal Infrared Sensor

TM – Thematic Mapper

T_{max} – Maximum temperature

T_{mean} – Mean temperature

T_{min} – Minimum temperature

T_{SD} – Standard deviation of temperature

USGS – United States Geological Survey

VIIRS – Visible Infrared Imaging Radiometer Suite

WRF – Weather Research and Forecasting

XGB– Extreme gradient boosting (XGBoost)

Introduction

Background

Ambient temperature and air pollution are increasingly recognized as major environmental health risks. Heat, cold, and variable temperature have been linked to increased morbidity and mortality across regions and climates (Cheng et al., 2014; Gasparrini et al., 2015; Guo et al., 2014; Song et al., 2017), and this burden is expected to grow with climate change (Gasparrini et al., 2017; Wang et al., 2018). Meanwhile, ambient particulate matter (PM) air pollution has become one of the fastest growing and leading risk factors for death and disability worldwide (Murray et al., 2020). Particulate matter <2.5 microns in diameter ($PM_{2.5}$) causes cardiovascular disease, pulmonary disease, and lung cancer (Landrigan et al., 2017), and both $PM_{2.5}$ and particulate matter <10 microns in diameter (PM_{10}) are associated with increased mortality, even at concentrations below current air quality guidelines (Liu et al., 2019), which the World Health Organization lowered in September 2021 to an annual mean of 5 $\mu g/m^3$ for $PM_{2.5}$ and 15 $\mu g/m^3$ for PM_{10} (World Health Organization, 2021). Recent work suggests that PM may even interact with heat, producing synergistic effects (J. Li et al., 2017; Qiu et al., 2020; Sun et al., 2020a).

Ambient temperature and PM are pervasive: everyone is exposed, and individuals have limited ability to reduce their exposure. This means that even small effects can have a substantial impact on population health. Cities can compound the impacts of temperature and PM as they are often warmer and more polluted than the surrounding countryside due to concentrated emissions from human activities, increased heat accumulation, slower heat diffusion, and decreased dispersion due to obstructed air movement (Arnfield, 2003). Urban areas are now home to more than half the world's population, and this share is projected to increase to almost 70% by 2050 (UN DESA Population Division, 2018). Understanding the health effects of ambient temperature and PM is essential to developing public health policies for a warming, urbanizing world.

Exposure assessment

To understand how ambient temperature and PM affect health, epidemiologists need to assess the air temperature or PM concentration to which a study population was exposed. Accurate exposure assessment is essential because exposure error tends to bias effect estimates towards the null, meaning studies may fail to detect harmful impacts (Zeger et al., 2000). Many studies, especially of temperature, have estimated the average exposure across a city or region based on measurements from one or a few meteorological or air quality monitoring stations. This approach captures temporal variation in exposure quite well, as monitors often take

measurements every hour or more frequently. But it is much less effective at capturing spatial variation in exposure. This can be problematic in rural areas where monitors tend to be sparse, leading many studies to focus exclusively on urban populations, which could limit the generalizability of their findings. Ignoring spatial variation in exposure can also be problematic in cities, as complex built environments can cause temperature and PM concentration to vary over fine spatial scales. Temperature exposure assessment in cities is further complicated by the fact that meteorological stations are often located in parks or peripheral areas which are usually cooler than denser central neighbourhoods.

Many researchers consider the "gold standard" of exposure assessment to be personal exposure measured by a portable monitor. This includes both indoor and ambient exposure, with error mostly determined by the accuracy of the monitoring equipment and the quality of the measurement protocol (for example, a temperature monitor worn on the wrist may be biased by body heat). But the costs of purchasing and maintaining monitors, recruiting volunteers, carrying the monitors, and processing the raw measurements limits widespread use of this approach, and it cannot be retrospectively applied to existing health data.

An alternative is to model the complete spatiotemporal distribution of temperature or PM over a study area. Individuals can then be assigned exposure based on their home address, or personal movement data if it is available. A few studies have used numerical weather prediction models and chemical transport models that simulate the formation, dispersion, and deposition of PM based on emissions, meteorology, and atmospheric chemistry (Ha et al., 2017a, 2017b; Sellier et al., 2014). These models are effective at both large and small spatial scales and can be used to forecast future exposure levels, but are limited by the accuracy of the input data and the completeness of their representation of atmospheric processes (Zhang et al., 2012a, 2012b). They are also computationally complex, which limits the spatiotemporal resolution, geographical area, and period that can be considered. For example, van Donkelaar et al. (2016) used four nested chemical transport models to estimate daily $PM_{2.5}$ at 250 to 60 km resolution worldwide and then used statistical downscaling to estimate annual mean $PM_{2.5}$ at 1 km resolution. And a recent study in France estimated hourly PM concentration only in 2010 and 2011 by combining the output of 51 different chemical transport models: a national model at 4 km resolution, 7 regional models at 3 to 4 km resolution, and 43 urban models at 10 to 200 m resolution (Riviere et al., 2019).

A less computationally intensive alternative is to use a geostatistical model that calibrates monitored temperature or PM concentration with spatial or spatiotemporal variables such as elevation, land use, and population density. Early studies developed land use regression models that captured the typical spatial distribution of temperature or PM (e.g. annual mean level) but struggled to capture day to day variation (Beckerman et al., 2013; Eeftens et al., 2012; Hoek et al., 2008; Pedersen et al., 2013; Sellier et al., 2014). A major advance was to incorporate daily satellite measurements of land surface temperature or aerosol optical depth as indicators of the spatiotemporal distribution of temperature or PM concentration, and to use statistical techniques that account for the fact that the relationship between satellite-derived variables and exposure levels often varies over space and time (Kloog et al., 2011; Lee et al., 2011). This allows accurate estimation of temperature or PM concentration over large areas, often daily at a 1 km resolution, with the main limitations being the spatial distribution of monitoring stations and the spatiotemporal resolution and extent of the satellite data. Satellite-based geostatistical models currently represent a good trade-off between modelling effort, spatiotemporal resolution, scalability, and accuracy.

Modelling ambient temperature

Satellite-based temperature models often use satellite-derived land surface temperature (LST) as an indicator of the spatiotemporal distribution of ambient temperature. LST, also known as skin temperature or radiometric surface temperature, is calculated from the measured thermal radiation of the earth's surface using Planck's Law after adjusting for atmospheric effects and surface emissivity (Li et al., 2013b). Thermal radiation can be measured at any time of day or night but requires clear-sky conditions because it is blocked by clouds. Elevation is also an important predictive variable since air temperature usually decreases with elevation.

One of the most widely used satellite instruments that measures LST is the Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS is carried on NASA's Terra and Aqua satellites which launched in December 1999 and May 2002, respectively. It has two thermal bands with a spatial resolution of 1 km, and each satellite overpasses twice daily (equator crossing around 10:30 and 22:30 for Terra; 13:30 and 1:30 for Aqua) resulting in four LST measurements per day. A validated LST product is freely available for each satellite (Wan, 2014). The relatively high temporal resolution (four measures per day) coupled with a spatial resolution of 1 km and availability of a pre-calculated LST product make MODIS well suited for temperature exposure models.

Air temperature at 2 m above the ground (T_a) is closely related to LST, but it is not straightforward to derive one from the other (Jin & Dickinson, 2010; Leaf & Erell, 2018; Vancutsem et al., 2010; Voogt & Oke, 1997). LST drives T_a through the surface energy balance. During the day, the earth's surface is exposed to solar radiation (mainly shortwave) and longwave radiation (mainly from the atmosphere). The amount of radiation at the surface depends on insolation and the transmittance and emissivity of the atmosphere, and the surface albedo determines the fraction of the radiation that it absorbs. The absorbed energy is stored as heat, increasing LST. It may then be conducted deeper into the ground (ground heat flux), transferred to the atmosphere as latent heat through evapotranspiration (latent heat flux), emitted as thermal radiation (radiative heat flux), or transferred to the atmosphere as sensible heat through convective heat flux, increasing T_a. Ground heat flux depends on the soil's temperature profile and its thermal conductivity and heat capacity, both of which vary with the soil moisture. Latent heat flux depends on soil moisture, land cover (especially vegetation type and abundance), ambient temperature, humidity, and atmospheric stability. Radiative heat flux depends on LST and emissivity, and sensible heat flux depends on ambient temperature, humidity, wind speed, and atmospheric stability. The equilibrium between these processes determines LST and T_a.

At night LST is often close to T_a (Vancutsem et al., 2010), but during the day the two diverge and more complex methods are needed to estimate T_a from LST. A very simple method is the temperature-vegetation index, which estimates T_a based on LST and normalized difference vegetation index (NDVI) (Prihodko & Goward, 1997; Zhu et al., 2013). Since this method does not require direct T_a measurements, it is useful for areas where temperature monitors are very sparse. But it struggles in areas with strong variation in topography, land cover, or soil moisture (Vancutsem et al., 2010). In better-monitored areas, studies have used LST to estimate daily T_a via regression, often in combination with variables such as elevation and land cover (Fu et al., 2011; Ho et al., 2016; Kestens et al., 2011; Weiss et al., 2014) and kriging (Florio et al., 2004; Hengl et al., 2012; Kilibarda et al., 2014). These models generally achieve root mean squared error (RMSE) of 2-3°C.

To improve performance over large, heterogeneous areas, Kloog et al. (2012a) applied a method first developed for PM modelling (Kloog et al., 2011; Lee et al., 2011) that uses linear mixed models to allow for day to day variation in the LST–T_a relationship. Importantly, the method includes a second gap-filling stage that estimates T_a for day-locations where LST is unavailable based on information from nearby stations and modelled T_a at the location on days when LST

is available (Kloog et al., 2012a; Lee et al., 2012). The approach has been used to estimate daily mean T_a at 1 km resolution from MODIS LST and other variables over a variety of large study areas including the northeastern United States (R^2 = 0.95; RMSE = 2.2°C) (Kloog et al., 2014b), the southeastern United States (R^2 = 0.95; RMSE = 1.7°C) (Shi et al., 2016), and France (R^2 = 0.95; RMSE 1.7°C) (Kloog et al., 2017). It was recently extended to estimate daily minimum, maximum, and mean T_a in Israel (T_{min} R^2 = 0.97, RMSE 1.1°C; T_{mean} R^2 = 0.99, RMSE = 0.7°C; T_{max} R^2 = 0.97, RMSE = 1.1°C T_{max}) (Rosenfeld et al., 2017) and has been applied to Landsat data to estimate mean summer T_a at 60 m resolution in Tel Aviv (R^2 = 0.92; RMSE = 1.6°C) (Pelta & Chudnovsky, 2017). Other recent studies have used climatology-anomaly geographically weighted regression (Oyler et al., 2015; Parmentier et al., 2015) or machine learning algorithms (Noi et al., 2017; Yoo et al., 2018) as an alternative to mixed models. Recent work has also focused on filling gaps in satellite-derived LST to allow LST-based T_a prediction for all days and locations (X. Li et al., 2018a, 2018b).

Modelling ambient particulate matter

Satellite-based particulate matter models can use aerosol optical depth (AOD) as an indicator of the spatiotemporal distribution of PM. AOD, also known as aerosol optical thickness, measures light extinction due to absorption and scattering by particles suspended in the atmosphere. It is related to the total quantity of aerosols in the entire atmospheric column and can be retrieved by satellites during the daytime over cloud-free areas. The MODIS instrument described above provides 1 km AOD at 0.47 and 0.55 µm derived via the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm from both the Aqua and Terra satellites twice per day at the equator and at higher frequencies towards the poles (up to four times per day over France) (Lyapustin et al., 2018). Other satellite AOD sources include the older MODIS deep blue/dark target AOD product which has lower spatial resolution (3 km) and is less accurate over bright surfaces, Suomi NPP's VIIRS instrument (daily 750 m at nadir), Sentinel-3's OLCI instrument (daily 1 km), and MSG's SEVIRI instrument (15 minutes 3 km at nadir).

PM modelling is challenging because AOD reflects aerosols throughout the entire atmospheric column and so is less closely related to ground level PM concentration than LST is to air temperature. Boundary layer height and other meteorological variables can be used as indicators of the fraction of AOD that represents ground level particles, and indicators of proximity to emissions sources such as land cover and road density can also be important predictive

variables. AOD is also more frequently missing than LST since it cannot be retrieved at night, and it can be contaminated by glint off snow or water or gridding issues along coastlines. Some studies estimate PM only when AOD was available, resulting in incomplete coverage, while others accommodate AOD missingness by calibrating two relationships: one to estimate PM based on AOD and another to estimate PM when AOD was not available (Hu et al., 2014; Kloog et al., 2011). More recently, methods have been developed to fill gaps in satellite AOD based on modelled AOD from chemical transport models or atmospheric reanalyses, allowing continuous AOD-based prediction (Di et al., 2019, 2016; Goldberg et al., 2019; L. Li et al., 2020; Stafoggia et al., 2019; Xiao et al., 2017).

PM models use similar statistical techniques as the temperature models described above to account for spatiotemporal variation in the relationship between PM and predictive variables. Mixed models have been used to predict daily 1 km PM based on MAIAC AOD over parts of the United States (Chudnovsky et al., 2014; Hu et al., 2014; Kloog et al., 2014a, 2012c, 2011; Lee et al., 2016), Mexico City (Just et al., 2015), parts of China (Meng et al., 2016; Xiao et al., 2017; Xie et al., 2015; Xue et al., 2020; Zhang et al., 2018; Zheng et al., 2016), Italy (Nordio et al., 2013; Stafoggia et al., 2017), Switzerland (de Hoogh et al., 2018), and Israel (Kloog et al., 2015b). These models generally achieved R^2 of about 0.65 to 0.85, often with better performance for $PM_{2.5}$ than PM_{10}. The mixed model approach has recently been extended to estimate PM in past years when there were no PM monitors (Liang et al., 2018) and to estimate twice-daily PM concentration (at the time of satellite overpass) in Israel, with R^2 of about 0.89 (Shtein et al., 2018).

Other studies have used geographically weighted regression, which allows for spatial variation in the relationship between PM and predictive variables, achieving R^2 of about 0.65 to 0.80 for daily PM in China (Ma et al., 2014; Song et al., 2014; Wu et al., 2016; You et al., 2015) and the United States (Hu et al., 2014). Geographically and temporally weighted regression extends this approach to allow temporal as well as spatial variation in the PM-predictors relationship, improving performance compared to geographically weighted regression (Yuanxi Guo et al., 2017; He et al., 2021b, 2020; He & Huang, 2018). Most recently, many studies have focused on applying machine learning algorithms that can capture complex relationships between PM and predictive variables such as random forests (Chen et al., 2019; Hu et al., 2017; Jiang et al., 2021; Meng et al., 2021; Schneider et al., 2020; Shao et al., 2020; Stafoggia et al., 2020, 2019; Sun et al., 2021; Wei et al., 2020, 2019), gradient boosting machines (Chen et al., 2019; Gui et al., 2020; Just et al., 2020), and neural networks (Chen et al., 2019; Di et al., 2016; Park et al.,

2020; Yan et al., 2020). These techniques can improve accuracy, although the resulting models may be difficult to interpret. Accuracy can be further improved by combining the predictions from multiple algorithms in an ensemble model that takes into account each algorithm's relative performance over different areas, periods, or PM concentrations (Di et al., 2019; Li et al., 2018; L. Li et al., 2017; Murray et al., 2019; Pu & Yoo, 2021; Shtein et al., 2019; Zhai & Chen, 2018).

Estimating model performance

It is important to evaluate the accuracy of satellite-based exposure models on independent data to estimate how well they generalize to unmonitored locations. This is particularly important when using highly flexible techniques such as machine learning algorithms, since they are vulnerable to overfitting (Just et al., 2020; Sarafian et al., 2019). A common approach is cross-validation (CV): data are repeatedly split into training and test sets, the model is calibrated using only the training data, and its predictions are compared to the held-out test data. Since PM concentrations are often spatiotemporally autocorrelated, the splitting should ensure that test data are far in space and time from training data. Recent studies have used spatial blocking, holding out all data from groups of randomly selected or clustered monitors (Just et al., 2020; Meng et al., 2021; Murray et al., 2019; Park et al., 2020; Pu & Yoo, 2021; Schneider et al., 2020; Shtein et al., 2019; Stafoggia et al., 2019; Xiao et al., 2020) or temporal blocking, holding out all data during one year (He et al., 2021a, 2020; Meng et al., 2021; Pu & Yoo, 2021; Xiao et al., 2020; Yan et al., 2020). Ensemble models require special care: data held out to test the ensemble should not have been used to train the base learners, and ensembles should be calibrated using CV predictions (predictions for held-out test data), as these reflect each base learner's ability to generalize to new areas (Shtein et al., 2019).

Perspectives for ambient temperature and particulate matter modelling

Improving accuracy

One of the main challenges in satellite-based exposure modelling, particularly of PM, is that the relationship between an exposure such as PM and predictive variables such as AOD can be complex and vary over space and time. As described above, various techniques have been proposed to address this issue including mixed models (Kloog et al., 2011), geographically and temporally weighted regression (He & Huang, 2018), and machine learning algorithms (Chen et al., 2019). Recent work shows that different techniques can complement each other, improving accuracy when their predictions are combined (Di et al., 2019; Shtein et al., 2019; Zhou et al., 2020). Gaussian Markov random fields (GMRF) are a technique from the

geostatistical literature that could improve performance alone or in an ensemble. GMRF allow approximating a Gaussian random field, a fundamental model for continuous spatiotemporal processes, in a way that is computationally tractable for large datasets (Lindgren et al., 2011; Rue et al., 2009). Recent work has shown that GMRFs are applicable to PM modelling (Beloconi et al., 2018; Cameletti et al., 2013), and a recent study in the northeastern United States showed that GMRF predicted daily 1 km $PM_{2.5}$ more accuratly than a mixed model (Sarafian et al., 2019). But GMRF have not yet been compared to high-performing machine learning algorithms or incorporated into ensemble models.

Increasing spatiotemporal resolution

The daily 1 km resolution of many satellite-based temperature and PM models represents a great improvement over station-based exposure assessment and suffices for many applications. But higher spatiotemporal resolutions could be helpful for epidemiological studies in urban areas, where temperature and PM can vary at fine spatial scales. It could also improve exposure assessment for the growing number of studies that have access to address or even personal movement data. One approach to achieve higher resolutions could be to combine data from multiple satellite instruments, some with high spatial but low temporal resolution and others with high temporal but low spatial resolution.

A number of geostationary satellites (GOES, MSG, MTSAT) carry instruments that capture thermal radiation or aerosol properties at very high temporal resolutions (e.g. every 15 minutes), although they have lower spatial resolutions (2 km to 5 km at nadir), a constant field of view, and possibly greater bias and inaccuracy (Freitas et al., 2013). Two recent studies used LST from the SEVIRI instrument on the MSG satellites to estimate hourly T_a over Germany via multiple linear regression (Bechtel et al., 2017) or via inverse distance weighted interpolation combined with a regional climate model (Krähenmann et al., 2018). And a study in Israel combined LST from MODIS and SEVIRI using an ensemble of random forests and gradient boosting machines to achieve an hourly 1 km resolution (Zhou et al., 2020).

Other satellites instruments provide thermal data at a higher spatial but lower temporal resolution than MODIS. The longest dataset comes from the USGS Landsat 5 Thematic Mapper, Landsat 7 Enhanced Thematic Mapper Plus, and Landsat 8 Thermal Infrared Sensor, which together provide a satellite record from March 1984 to the present. Landsat 5 and 7 have a spatial resolution of 120 m and 60 m, respectively, but only a single thermal band, which limits the accuracy of LST retrieval (Chen et al., 2017; Li et al., 2013a). Landsat 8 has two

thermal bands with a spatial resolution of 100 m, but stray light contamination of the second thermal band makes it unusable for LST retrieval. The USGS is developing a pre-calculated Landsat LST product (Malakar et al., 2018), but it is not yet available worldwide. In the meantime, a variety of methods have been developed to calculate LST from the freely available data (Jiménez-Muñoz et al., 2014, 2009; Li et al., 2013a; Wang et al., 2015; Yu et al., 2014), and one LST dataset is available for limited download (Parastatidis et al., 2017).

The main limitation of Landsat data is its low temporal resolution. Each satellite has a return time of 16 days, with the orbits staggered such that one of the satellites overpasses every 8 days. However, the satellites do not continuously image all areas overpassed so there may be several weeks between consecutive images of a particular location. Because of this, the few studies that used Landsat thermal data to model T_a over an urban area estimated only the typical spatial pattern of T_a in a season (Ho et al., 2014) or only considered days with imagery (Pelta & Chudnovsky, 2017). No study has yet used Landsat data to model gap-free daily T_a at high spatial resolution over large areas and long periods.

Temperature, particulate matter, and perinatal health

A growing body of research suggests that exposure to heat, cold, or PM during pregnancy may decrease birth weight and shorten the duration of gestation (Bekkar et al., 2020; Chersich et al., 2020; Jacobs et al., 2017; C. Li et al., 2020; X. Li et al., 2017; Zhang et al., 2017). This is particularly concerning because adverse birth outcomes are a growing health problem. Low birth weight (<2500 g at term) is the leading health risk for children under 10 worldwide (Murray et al., 2020) while preterm birth (delivery at <37 weeks amenorrhea) is the leading cause of under-5 mortality (Murray et al., 2020), accounting for one million deaths in 2015 (Liu et al., 2016). About 11% of births worldwide are preterm, and this rate is increasing in many countries (Chawanpaiboon et al., 2019). In metropolitan France, the rate of preterm birth among singletons increased steadily from 4.5% in 1995 to 6.0% in 2016 while the rate of low birth weight increased from 4.6% to 5.7% (Blondel et al., 2017).

Low birth weight and preterm birth impose substantial medical, economic, and psychosocial burdens on infants, mothers, and their families (Harrison & Goldenberg, 2016; McCormick et al., 2011; Saigal & Doyle, 2008). These burdens can continue throughout childhood and into adulthood, with increased risk for asthma and other respiratory diseases, cardiovascular disease, cerebral palsy, visual impairment, hearing loss, epilepsy, and learning difficulties (Barker,

2004; Belbasis et al., 2016; McCormick et al., 2011; Saigal & Doyle, 2008). The study of these effects falls under the framework of the developmental origins of health and disease (DOHaD), which hypothesizes that environmental exposures during foetal or post-natal development can influence the risk of later health problems (Sinclair et al., 2007). A large corpus of research in animal models supports the DOHaD hypothesis (Hanson & Gluckman, 2011). In humans, DOHaD remains an active field of research, as links between exposures during the developmental period and health events in childhood and adulthood remain incompletely understood.

Ambient temperature and adverse birth outcomes

Although several studies have examined associations between ambient temperature and adverse birth outcomes, and most have reported harmful effects (Bekkar et al., 2020; Chersich et al., 2020; Zhang et al., 2017), the diversity of exposure metrics and windows considered makes it difficult to synthesize the findings. For example, studies have considered entire pregnancy, entire trimester, or entire month exposure, weekly exposure throughout pregnancy, weekly exposure during the last 1, 2, or 4 weeks of pregnancy, and daily exposure during the last 7, 10, 15, or 30 days of pregnancy. Heat has been defined as the 75th, 90th, 95th, or 99th percentile of the annual or hot season distribution of daily mean or maximum air temperature or apparent temperature. And most studies estimated exposure for entire cities or regions based on one or a few weather stations, largely ignoring spatial temperature contrasts.

Table I-1 lists the average effects of heat on birth weight and preterm birth reported by a recent review and meta-analysis of heat and adverse birth outcomes (Chersich et al., 2020). Compared to lower temperature, high temperature throughout an entire trimester or pregnancy decreased birth weight by a median of 25.5 g (range: -39.4 to -15.0 g; 4 studies) and had median odds ratio for low birth weight of 1.09 (range 1.01 to 2.49; 8 studies), but the diversity of study designs precluded meta-analysis. Entire trimester or entire pregnancy heat had mean odds ratio for preterm birth of 1.15 (95% confidence interval [CI] 1.13 to 1.18; 8 studies). For heat during the last days or weeks of pregnancy, the mean odds ratio for preterm birth was 1.05 (95% CI 1.04 to 1.05; 19 studies), and a 1°C increase during the last days of pregnancy increased the odds of preterm birth by 5% (95% CI 3% to 7%). Heat throughout pregnancy or during the last few days of pregnancy also increased the risk of stillbirth (Chersich et al., 2020).

Table I-1. Effects of heat on adverse birth outcomes from a review and meta-analysis by Chersich et al. (2020).

Outcome	Exposure	Effect	N[a]	I^2
Birth weight	High vs. low temperature (entire trimester or pregnancy)	-25.5 g (range -39.4 to -15.0 g)[b]	4	-
Birth weight	High vs. low temperature (entire trimester or pregnancy)	OR 1.09 (range 1.01 to 2.49)[b]	8	-
Preterm birth	High vs. low temperature (entire trimester or pregnancy)	OR 1.15 (95% CI 1.13 to 1.18)	8	65.2%
Preterm birth	High vs. low temperature (last weeks of pregnancy)	OR 1.05 (95% CI 1.04 to 1.05)	19	83.6%
Preterm birth	1°C increase (last days of pregnancy)	OR 1.05 (95% CI 1.03 to 1.07)	6	87.7%

[a] Number of studies

[b] Median & range of effect across studies; methodological differences precluded meta-analysis

Adapted with permission from "Associations between high temperatures in pregnancy and risk of preterm birth, low birth weight, and stillbirths: Systematic review and meta-analysis" by Chersich, M.F., Pham, M.D., Area, A., Haghighi, M.M., Manyuchi, A., Swift, C.P., Wernecke, B., Robinson, M., Hetem, R., Boeckmann, M., & Hajat, S. (2020). The BMJ, 371, 1–13. Copyright 2020 by "BMJ Publishing Group Ltd."

There is also growing evidence that cold may increase the risk of adverse birth outcomes. Cold throughout pregnancy or during trimesters 2 or 3 was associated with decreased birth weight or increased risk of term low birth weight in New York City (Ngo & Horton, 2016), Massachusetts (Yitshak-Sade et al., 2021), multiple cities in the United States (Ha et al., 2017b), and Israel (Kloog et al., 2018). Cold during the last days of pregnancy increased the risk of preterm birth in Shenzhen (Liang et al., 2016) and Guangzhou (He et al., 2016), while cold earlier in or throughout pregnancy increased preterm birth risk in Uppsala (Bruckner et al., 2014), Guangzhou (He et al., 2016), Brisbane (Li et al., 2018), multiple cities in the United States (Ha et al., 2017a), and multiple cities in China (Y.-Y. Wang et al., 2020). However, other studies reported no association, mixed results, or a protective effect for cold during the last days of pregnancy (Cox et al., 2016; Schifano et al., 2013; Sun et al., 2019) or earlier in pregnancy (Arroyo et al., 2016; Avalos et al., 2017; Giorgis-Allemand et al., 2017; Guo et al., 2018; Kloog et al., 2015a; Liu et al., 2020). Some of this inconsistency may be due to differences between climates and populations' adaptation to local conditions. For example, the study in Brisbane found that heat was less hazardous for preterm birth in 2013 than it had been in 1994 while cold was more hazardous, suggesting adaptation to the changing climate.

Particulate matter and adverse birth outcomes

Many studies have examined associations between PM_{10}, $PM_{2.5}$, and birth outcomes, with quite mixed results. While several studies reported that PM increased the risk of low birth weight (Pedersen et al., 2013) and preterm birth (Dadvand et al., 2013; Ha et al., 2014; Kloog et al., 2012b; Pedersen et al., 2013), others found no effect in Europe (Giorgis-Allemand et al., 2017; Lee et al., 2008) or even a protective effect (Stieb et al., 2016), which they attributed to bias or residual confounding.

Table I-2 lists the average effects of PM on birth weight and preterm birth reported by two recent meta-analyses (Ju et al., 2021; C. Li et al., 2020). A 10 $\mu g/m^3$ increase in entire pregnancy $PM_{2.5}$ increased the risk of low birth weight by 8% (95% CI 4% to 12%) (C. Li et al., 2020) and the risk of preterm birth by 7% (95% CI 5% to 10%) (Ju et al., 2021). $PM_{2.5}$ was also harmful during trimesters 2 and 3. Effects were slightly weaker for 10 $\mu g/m^3$ increase in entire pregnancy PM_{10}: it increased the risk of low birth weight by 5% (95% CI 3% to 8%) (C. Li et al., 2020) and the risk of preterm birth by 3% (95% CI 1% to 6%) (Ju et al., 2021). PM_{10} was also harmful during trimester 2. However, there was substantial heterogeneity between effect estimates from different studies. A meta-analysis focusing exclusively on the United States reported somewhat larger effects (Bekkar et al., 2020), and a fourth meta-analysis concluded that while PM is likely harmful, the magnitude of its effect remains uncertain due to unexplained heterogeneity between studies and lack of standardized, high-quality exposure assessment (Uwak et al., 2021).

Table I-2. Effects of PM on adverse birth outcomes from meta-analyses by C. Li et al. (2020) and Ju et al. (2021). Only statistically significant effects are reported here.

Outcome	Exposure[a]	Period	Relative risk (95% CI)	N[b]	I^2
Low birth weight	$PM_{2.5}$	Pregnancy	1.08 (1.04 to 1.12)	29	86%
Low birth weight	$PM_{2.5}$	Trimester 3	1.05 (1.01 to 1.10)	20	92%
Low birth weight	PM_{10}	Pregnancy	1.05 (1.03 to 1.08)	23	70%
Low birth weight	PM_{10}	Trimester 2	1.01 (1.01 to 1.02)	13	28%
Preterm birth	$PM_{2.5}$	Pregnancy	1.07 (1.05 to 1.10)	31	89%
Preterm birth	$PM_{2.5}$	Trimester 2	1.03 (1.00 to 1.07)	23	97%
Preterm birth	$PM_{2.5}$	Trimester 3	1.02 (1.00 to 1.04)	23	93%
Preterm birth	PM_{10}	Pregnancy	1.03 (1.01 to 1.06)	15	92%

[a] 10 $\mu g/m^3$ increase; [b] Number of studies

Biological mechanisms

The biological mechanisms by which temperature and PM could affect intrauterine growth and the timing of delivery are incompletely understood. For temperature, the main hypothesized mechanism are hemodynamic effects related to thermoregulation and hormonal imbalance due to thermal stress. Pregnant women may be particularly sensitive to temperature because their increased fat deposition, decreased surface area to volume ratio, and increased metabolic heat production (due to increased mass and the foetus' metabolic rate) complicate thermoregulation (Dadvand et al., 2011; Strand et al., 2011a). Heat exposure can cause peripheral vasodilation, which may decrease blood flow to the placenta (Basu et al., 2010), dehydration, which decreases blood volume and increases blood viscosity (Stan et al., 2013), and the release of heat shock proteins, which may cause placental inflammation (Dadvand et al., 2011). Cold may similarly decrease blood flow to the placenta via vasoconstriction and increased blood viscosity (Beltran et al., 2013; Bruckner et al., 2014). These might trigger labour or alter transplacental oxygenation and nutrition, disrupting foetal growth and development. They could also contribute to gestational hypertension and preeclampsia, which are major risk factors for intrauterine growth restriction and preterm birth (Beltran et al., 2013; Shashar et al., 2020). Heat stress and dehydration may increase levels of antidiuretic hormone, prostaglandin, and oxytocin, which are involved in labour (Dadvand et al., 2011; Schifano et al., 2013; Stan et al., 2013), and cold stress may also alter hormonal balance (Bruckner et al., 2014). Finally, cold may be associated with increased exposure to risk factors such as smoke and infectious agents due to greater time spent indoors (He et al., 2016).

For PM, the main hypothesized mechanisms are oxidative stress, inflammation, hemodynamic effects, endothelial dysfunction, and endocrine disruption. Pregnant women may be particularly sensitive because the need to supply blood to the placenta increases stress on the cardiovascular system and endothelium (Pedersen et al., 2014). PM-induced inflammation in the lungs may reduce maternal blood oxygenation, which could in turn affect transplacental oxygen delivery (Kannan et al., 2006). In the placenta, PM-induced inflammation and oxidative DNA damage might alter transplacental exchanges or induce labour (Kannan et al., 2006; Schifano et al., 2013), as might epigenetic changes (Abraham et al., 2018; Lin et al., 2016; Slama et al., 2008). The same processes may occur in the foetus if PM is transported across the placenta (Abraham et al., 2018; Kannan et al., 2006; Lin et al., 2016; Slama et al., 2008). Inflammation may also alter the mother's immune function, increasing vulnerability to infection (Pedersen et al., 2014; Slama et al., 2008) and possibly exacerbating its effects (Kannan et al., 2006). PM can cause

vasoconstriction, increase blood viscosity, alter blood coagulability, and cause endothelial dysfunction, all of which might disrupt transplacental oxygenation and nutrition or contribute to hypertensive disorders (Kannan et al., 2006; Pedersen et al., 2014; Slama et al., 2008). Finally, PM may interfere with foetal growth and development via endocrine disruption (Slama et al., 2008) or binding of placental growth factor receptors (Kannan et al., 2006).

Perspectives for perinatal health studies of temperature and PM

Recent health studies have increasingly used spatiotemporally resolved exposure estimates and advanced statistical methods to limit bias and confounding, but findings of temperature and PM's effects on perinatal health remain mixed (Giorgis-Allemand et al., 2017; Ha et al., 2017b, 2017a; Kloog et al., 2015a, 2012b; Pedersen et al., 2013; Qiu et al., 2020; Stieb et al., 2016; Yitshak-Sade et al., 2020). A major unsettled question is how the timing of exposure during pregnancy influences health effects. Windows of susceptibility can be difficult to detect because exposure tends to be correlated across trimesters (Wilson et al., 2017) and temperature follows a seasonal pattern to which women may acclimatize. Windows of susceptibility may not align with clinical trimesters, but few studies considered narrower windows. Furthermore, when studying preterm birth, it is important to account for the fact that the risk of delivery increases with gestational age and that preterm infants have shorter periods of prenatal exposure than term infants.

In relation to birth weight, heat and cold seem most dangerous during trimester 3 (Basu et al., 2018; Ha et al., 2017b; Kloog et al., 2015a; Yitshak-Sade et al., 2021, 2020), although exposure during trimesters 1 and 2 may also be harmful (Arroyo et al., 2016; Ha et al., 2017a; Ngo & Horton, 2016; Yitshak-Sade et al., 2021, 2020). The relationship between the timing of temperature exposure and preterm birth is less clear. Many studies examined only the last days of pregnancy and found that heat may trigger early delivery, usually with a lag time of 1 to 7 days (Auger et al., 2014; Avalos et al., 2017; Basu et al., 2017; Cox et al., 2016; Schifano et al., 2016; Spolter et al., 2020; Sun et al., 2019; Vicedo-Cabrera et al., 2014). Fewer studies have examined temperature earlier in pregnancy, but they suggest that pregnant women may be susceptible to heat and cold during the weeks before and after conception and the first half of trimester 2 (Guo et al., 2018; Ha et al., 2017a; Liu et al., 2020). PM may be most harmful during trimester 2 or 3, decreasing birth weight and shortening the duration of gestation, but findings have been inconsistent (Ju et al., 2021; C. Li et al., 2020b; Uwak et al., 2021).

Another unsettled question is which temperature metric is most biologically relevant. This may depend on the local climate, the population's acclimation and adaptability, and the health outcome of interest (Basu et al., 2008; Laaidi et al., 2012; Murage et al., 2017). A few studies have reported stronger associations for night-time heat (Cox et al., 2016; Zhong et al., 2018), while others have linked temperature variability to adverse birth outcomes (Jakpor et al., 2020; Li et al., 2018; Molina & Saldarriaga, 2017).

Objectives and outline

The objectives of this PhD were to develop novel approaches to improve high spatiotemporal resolution satellite-based geostatistical exposure models, apply them to model daily ambient temperature and PM over France since 2000, and use the exposure estimates to study the association between ambient temperature during pregnancy and preterm birth risk.

Chapter 1 describes how we modelled daily minimum, mean, and maximum air temperature from 2000 to 2018 at a 1 km resolution over continental France and a 200 m resolution over 103 urban areas. A major contribution was a spatial downscaling technique for satellite-based air temperature models.

Chapter 2 describes how we modelled daily mean $PM_{2.5}$ and PM_{10} concentration from 2000 to 2019 at a 1 km spatial resolution over continental France. A major contribution was the finding that Gaussian Markov random fields may be more accurate than widely used mixed models and random forests.

Chapter 3 describes how we comprehensively assessed the effects of ambient temperature throughout pregnancy on the risk of preterm birth in France. A major contribution was the finding that pregnant women in a temperate climate may be susceptible to cold during the second half of trimester 1 and about 7 days before delivery, and to night-time heat during the first half of trimester 1, the second half of trimester 2, and the beginning of trimester 3.

The final section discusses these findings and directions for further research.

Chapter 1

Modelling daily ambient temperature in France

Ian Hough[a,b], Allan C. Just[c], Bin Zhou[b,d], Michael Dorman[b], Johanna Lepeule[a], Itai Kloog[b]

[a] *Institute for Advanced Biosciences, Université Grenoble Alpes, INSERM, CNRS, La Tronche, France*

[b] *Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Be'er Sheva, Israel*

[c] *Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, United States*

[d] *Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany*

Portions of this chapter were previously published as:

Hough, I., Just, A. C., Zhou, B., Dorman, M., Lepeule, J., & Kloog, I. (2020). A multi-resolution air temperature model for France from MODIS and Landsat thermal data. *Environmental Research*, *183*, 109244. https://doi.org/10.1016/j.envres.2020.109244

1.1 Introduction

The first objective of this PhD was to develop a satellite-based model of ambient temperature to estimate the exposure of participants in epidemiological studies in France. The model needed to estimate daily air temperature (T_a) over continental France starting from 2000 to 2018. We knew of no existing temperature model covering continental France during this period at a daily 1 km spatial resolution or better. The main existing models were:

- ERA-20C (Poli et al., 2016), an atmospheric reanalysis with worldwide coverage from 1900 to 2010 at a 3-hourly 125 km resolution.
- ERA-Interim (Balsamo et al., 2015; Dee et al., 2011), an atmospheric reanalysis with worldwide coverage from 1979 to 2010 at a 3-hourly 80 km resolution.
- SAFRAN (Habets et al., 2008; Quintana-Seguí et al., 2008), a hydrometeorological model covering France from 1960 to present at a daily 8 km resolution.
- A geostatistical model (Kloog et al., 2017) of daily mean T_a using satellite-derived land surface temperature (LST) over France from 2000 to 2011 at a daily 1 km resolution.

We advanced the state of the art by extending the geostatistical T_a model of Kloog et al. (2017) to cover 2000 to 2018 and adding three key enhancements.

First, we estimated daily minimum (T_{min}) and maximum T_a (T_{max}) in addition to daily mean T_a (T_{mean}) by calibrating a separate model for each, using daytime LST as a predictor of T_{max} and night-time LST as a predictor of T_{min} and T_{mean}. T_{min} and T_{max} can serve as proxies for night-time and daytime temperature, which could be useful for studying urban heat islands which exhibit different spatial patterns and intensities during day vs. night (Arnfield, 2003). Night-time and daytime T_a may also have different health effects (Basu et al., 2008; Laaidi et al., 2012; Murage et al., 2017), as may diurnal temperature range ($T_{max} - T_{min}$) (Cheng et al., 2014; Wu et al., 2019).

Second, we increased accuracy by allowing the T_a–LST relationship to vary over space on each day by using linear mixed models with a nested random effect for each day and each of 8 climatic regions.

Third, we developed a downscaling technique to increase the spatial resolution over urban areas. This involved an ensemble of machine learning algorithms that predict the daily residuals of the 1 km model based on 30 m thermal data from the Landsat 5, 7, and 8 satellites and other

high-resolution variables. The increased spatial resolution should reduce exposure error in cities, where temperature can vary at fine spatial scales due to the complex built environment, and where weather stations are often located at a peripheral airport or in parks which may be cooler than dense central neighbourhoods.

Our final model reconstructs daily T_{min}, T_{mean}, and T_{max} from 2000 to 2018 at a 1 km spatial resolution over continental France and at a 200 m spatial resolution over 103 urban areas with population 50 000 or more. The model performs very well, with mean 1 km R^2 of 0.92 (T_{min}), 0.97 (T_{mean}), or 0.95 (T_{max}) and root mean squared error (RMSE) of 1.9, (T_{min}), 1.3 (T_{mean}), or 1.8 (T_{max}). Over urban areas, the 200 m estimates have R^2 of 0.79 (T_{min}), 0.79 (T_{mean}), and 0.84 (T_{max}). Our downscaling technique may improve temperature models in other areas, and our temperature dataset is available to assess the exposure of participants in epidemiological studies in France.

1.2 Materials

1.2.1 Study domain

Continental France covers a roughly hexagonal area of 542 973 km^2 in western Europe bounded by the Atlantic Ocean to the west and the Mediterranean Sea to the southeast. Most of the terrain is at low elevation, but the Pyrenees in the southwest rise to over 3000 m and in the southeast the Alps reach 4 809 m. Joly et al. (2010) identified eight distinct climatic regions, each with a characteristic magnitude, variability, and seasonality of temperature and precipitation (Figure 1-1). The north and west coasts have a wet, temperate oceanic climate, which transitions to a drier, cooler modified oceanic climate in the north centre. The mountainous east, south centre, and southwest have variable montane and semi-continental climates with cold winters. In the southeast, the Mediterranean coast has hot, dry summers with mild wet winters; the inland southeast and isolated segments of the west coast are similar but cooler. The southwest basin resembles the inland southeast but with drier winters.

Continental France has a population of approximately 64.5 million. About 80% of the population is urban, and this share is projected to grow to 88% by 2050 (UN DESA Population Division, 2018). The largest urban area, Paris, has a population of 12.5 million (20% of the total) and the six next largest urban areas have a population of one to 2.3 million (combined 14% of total). A further 10% of the population lives in cities of one half to one million, and

37% live in urban areas with fewer than 500 000 residents (Figure A1-1). For this study, we considered the period of January 1, 2000 through December 31, 2018.

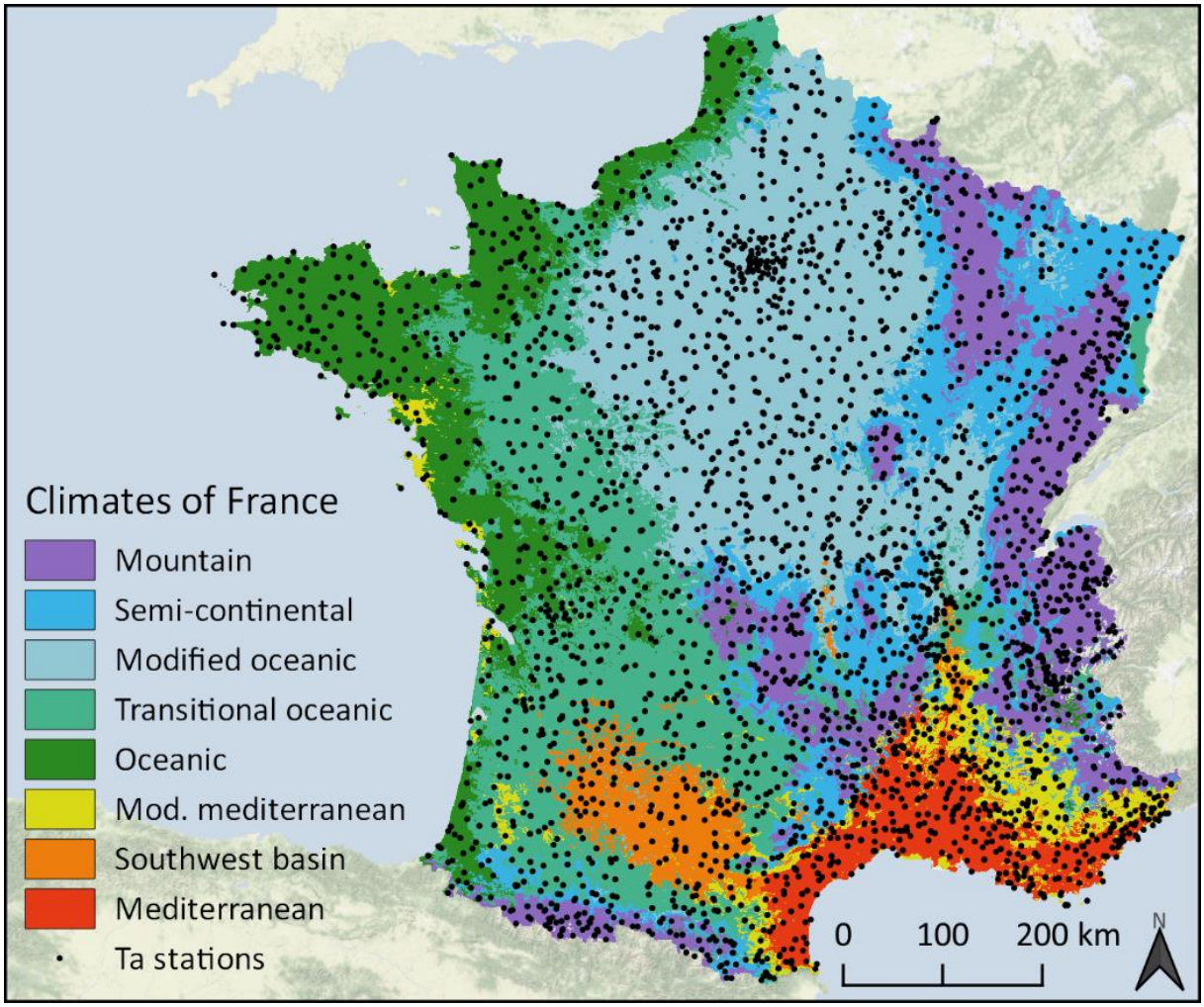

Figure 1-1. Climatic regions of France as defined by Joly et al. (2010) and METEO FRANCE weather monitoring stations (T_a stations) used in the model.

1.2.2 Meteorological observations

We used daily weather station observations from Météo France, the French national meteorological service. About 64% of the observations came from stations managed by Météo France; the remaining stations were managed by other entities. All observations were quality controlled by Météo France. We excluded stations with no metadata or that did not record hourly T_a, and for each month during the study period we excluded stations that were active for fewer than 21 days in the month. This left 1710 to 2314 stations on each day. The stations were distributed over the entire study region, but were denser in populous areas (e.g. Paris, the southeast) and the Alps (where many ski resorts, hydroelectric dams, and avalanche monitors

collect temperature data) (Figure 1-1). Just 3% of the stations were located within large urban areas (defined in section 1.2.7), 7% were in peri-urban areas (within 5 km of an urban area), and the remaining 90% were rural.

The stations registered daily T_{min} as the lowest T_a measured from 18 UTC the previous day until 18 UTC on the day; daily T_{max} was the highest T_a measured from 6:00 UTC on the day until 6:00 UTC the following day. Most stations calculated T_{mean} as the mean of all (at least 24) T_a observations from 0 UTC on the day until 0 UTC the following day. However, about 40% of the T_{mean} observations were calculated as the average of T_{min} and T_{max}. We excluded these observations, meaning our final dataset had fewer observations for T_{mean} than for T_{min} or T_{max}. Daily T_a at the included stations during the study period ranged from T_{min} of -31.2°C to T_{max} of 44.1°C; mean T_{mean} was 11.3°C with a standard deviation of 7.1°C (Table A1-1).

1.2.3 Land surface temperature and emissivity

We used version 6 of the MODIS daily 1 km land surface temperature and emissivity product from the Terra and Aqua satellites (MOD11A1 and MYD11A1, respectively) (Table 1-1). These products include daytime and night-time LST derived using a split-window algorithm and land use classification-based emissivity and have been masked for clouds and validated to ± 2 K in clear-sky conditions across 47 sites on all seven continents (Wan, 2014). We used the quality assessment band to exclude pixels with LST error > 2 K. As LST retrieval error increases over snow and water, we also excluded pixels with NDVI < 0 or where the corresponding 1 km grid cell had land cover of > 33% water.

Table 1-1. Satellite instruments used in the geostatistical air temperature model.

Instrument	Satellite	Resolution	Revisit time	Overpass*	Period
MODIS	Terra	1 km	12 hours	10:00 22:00	2000-02-02 to present
MODIS	Aqua	1 km	12 hours	13:00 01:00	2002-07-04 to present
TM	Landsat 5	120 m†	16 days	10:00	1984-03-01 to 2011-11-18
ETM+	Landsat 7	60 m†	16 days	10:00	1999-04-15 to present
TIRS	Landsat 8	100 m†	16 days	10:00	2013-02-11 to present

*Approximate local solar time; †Resampled to 30 m

1.2.4 Top-of-atmosphere brightness temperature

For large urban areas, we composited 30 m top-of-atmosphere brightness temperature (T_b) from the Landsat 5, 7, and 8 satellites (Table 1-1). T_b is the kinetic temperature a perfect blackbody

would have if it emitted the quantity of thermal radiation measured by the satellite instrument. Converting T_b to LST requires correcting for atmospheric effects and accounting for the emissivity of the earth's surface. This is difficult in the case of the Landsat satellites because Landsat 5 and 7 have only a single thermal band and the USGS Landsat 8's second thermal band is contaminated by stray light, precluding the use of the split-window algorithm (Li et al., 2013b). A global Landsat LST product was under development but not yet available (Malakar et al., 2018), so for this study we used T_b from the USGS Landsat Collection 1 Level-2 surface reflectance products (USGS, 2018a, 2018b).

The 16-day revisit time of the Landsat satellites meant that T_b was unavailable for many locations on many days. Cloud cover and sensor malfunctions also contribute to these data gaps and can increase error in T_b retrieval. To reduce error, we discarded all scenes with cloud cover > 75%. We also discarded all scenes captured during periods of instrument malfunction, which we identified by checking summary statistics of each scene for unrealistic values (e.g. mean T_b > 100°C). We then trimmed the edges of Landsat 5 scenes by 2.5 km to remove abnormalities (Robinson et al., 2017) and masked pixels identified as high- or medium-confidence cloud in the pixel quality assessment band. We masked any remaining pixels where $T_b \leq$ -25°C or $T_b \geq$ 50°C. Finally, for each calendar month we composited all T_b retrievals during the entire study period (e.g. every January from 2000 to 2018). This yielded 12 gap-free T_b datasets representing the 19-year mean T_b of each pixel in each month.

1.2.5 NDVI

We used version 6 of the MODIS monthly composite 1 km NDVI product from the Terra and Aqua satellites (MOD13A3 and MYD13A3, respectively). For large urban areas we also composited 30 m NDVI from the Landsat 5, 7, and 8 Collection 1 Level-2 surface reflectance products. We used a similar quality assurance and compositing procedure as for T_b, first discarding all scenes with greater than 75% cloud cover or during periods of thermal sensor malfunction (as this results in unreliable cloud confidence scores in the pixel quality assessment band). We then trimmed the edges of Landsat 5 scenes by 2.5 km and adjusted NDVI from Landsat 5 and Landsat 7 to match Landsat 8 following Robinson et al. (2017). For each calendar month, we created two 17-year mean composites, one using pixels marked as clear in the pixel quality assurance band (i.e. not cloud, cloud shadow, snow, or water) and a second using pixels marked as snow or water. Finally, we mosaiced the two composites preferring the clear pixels composite.

1.2.6 Elevation, Population, Land Cover, and Climatic Regions

We used version 1.1 of the European Digital Elevation Model (EU-DEM) from the Copernicus Land Monitoring Service. These data have a 25 m spatial resolution and vertical RMSE of ±7 m (Tøttrup, 2014). We also used 200 m gridded 2010 population from INSEE, the French national statistics agency (Insee, 2016). We used the 100 m Corine Land Cover (CLC) inventory for 2000, 2006, and 2012. The 2000 edition has been validated to better than 85% thematic accuracy. We aggregated the land cover classes into four groups: artificial, vegetation, bare, and water (Table A1-2). Finally, we used the eight climatic regions of Joly et al. (2010), which are based on temperature and precipitation patterns (Figure 1-1).

1.2.7 Model grids

For the 1 km model, we created a grid covering continental France by making a 1 km square buffer around the centroid of each MODIS 1 km LST pixel in the ETRS89-LAEA Europe (EPSG:3035) equal-area projection. We associated each 1 km grid cell with the MODIS LST and NDVI pixel having the same centroid and calculate the mean elevation, total population, percent area of each land cover group, and climate region with greatest spatial overlap.

For the 200 m model, we created a grid covering large urban areas. Starting from a 200 m grid in the ETRS89-LAEA Europe (EPSG:3035) equal-area projection, we selected all cells in continental France containing “Urban fabric” or “Industrial or commercial units” in the 2012 CLC inventory. We associated each cell with the corresponding INSEE gridded population and selected cells with 50 or more inhabitants as well as the eight surrounding cells (i.e. including diagonal neighbours). We defined urban areas as four-wise contiguous (i.e. excluding diagonal neighbours) groups of cells and summed the population of all cells in each urban area. Finally, we eliminated urban areas with population < 50 000, leaving 103 large urban areas ranging from greater Paris (9.4 million inhabitants) to Armentières (50 260 inhabitants). For each 200 m grid cell in a large urban area or that contained a weather station we calculated the mean 17-year composite Landsat T_b and NDVI for each calendar month, mean elevation, and percent area of each land cover group.

1.3 Methods

We used a four-stage approach to predict T_a: stages 1 and 2 predicted daily 1 km T_a across continental France and stages 3 and 4 predicted daily 200 m T_a within large urban areas. We

considered each year during the study period (2000 to 2018) and each T_a measure (T_{min}, T_{max}, and T_{mean}) separately. Briefly, in stage 1, we calibrated T_a at each station as a function of daily 1 km LST and other variables. We used a linear mixed model to allow the T_a–LST relationship to vary by day within each climatic region and used the calibrated relationship to predict 1 km T_a (T_{ap}) for all cell-days where LST was available.

In stage 2, we filled gaps in T_{ap} where 1 km LST was not available by calibrating T_{ap} as a function of daily 1 km inverse distance weighting interpolated observed T_a (T_{IDW}). We used a linear mixed model to allow the T_{ap}–T_{IDW} relationship to vary by location and used the calibrated relationship to fill gaps in T_{ap}. This produced the final 1 km T_a model: gap-free daily 1 km predicted T_a (T_{1km}).

In stage 3, we trained random forest (RF) and extreme gradient boosting (XGB) models to predict the daily 200 m residuals of the 1 km T_a model (the difference between the model's prediction and monitored temperature) based on 200 m composite T_b and other high-resolution variables. We used the RF and XGB to predict the residual for all 200 m cell-days (R^{RF}_{200m} and R^{XGB}_{200m}, respectively).

In stage 4, we calibrated a generalized additive model that ensembles R^{RF}_{200m} and R^{XGB}_{200m} using a tensor product smooth that allows the relative performance of the RF and XGB to vary by location and with the magnitude of the predicted residual. Finally, we added the ensemble predictions to T_{1km} to get the final 200 m T_a model: daily 200 m predicted T_a over large urban areas (T_{200m}).

1.3.1 Stage 1: predicting 1 km T_a from LST

In stage 1, we predicted T_a for all 1 km grid cells and days where MODIS 1 km LST was available. Our method was similar to that used in Kloog et al. (2017) with the addition of some explanatory variables and nesting of daily random effects within climatic regions. First, we associated each weather station T_a observation with the nearest 1 km grid cell for which LST was available on the day of the observation, up to a maximum distance of 1.5 km. The number of T_a observations matched with LST varied by year (Table A1-3 to Table A1-5); the average was about 354 thousand for T_{min}, 205 thousand for T_{mean}, and 324 thousand for T_{max}. We used these to calibrate a mixed model with the equation:

$$T_{a_{ij}} = (\alpha + \mu_{jk}) + (\beta_1 + \upsilon_{jk})LST_{ij} + \beta_2 Emis_{ij} + \beta_3 NDVI_{im} + \sum_{l=1}^{4} \beta_{4l} LC_{ily} + \beta_5 Elev_i + \beta_6 Pop_i + \varepsilon_{ij} \quad \text{Eq. (1-1)}$$

where $T_{a_{ij}}$ is the observed ambient temperature associated with 1 km grid cell *i* on day *j*; α is a fixed intercept and μ_{jk} is a random intercept on day *j* for the climatic region *k* that contains cell *i*; β_1 is a fixed and υ_{jk} is a random coefficient for LST on day *j* for the climatic region *k* that contains cell *i*; LST_{ij} is the MODIS 1 km land surface temperature of cell *i* on day *j*. β_2–β_6 are fixed coefficients of the other explanatory variables; $Emis_{ij}$ is the emissivity of cell *i* on day *j*; $NDVI_{im}$ is the MODIS NDVI of cell *i* in the month *m* that contains day *j*; β_l is a fixed slope for each of the *l* land cover groups and LC_{ily} is the fraction of cell *i* occupied by land cover group *l* in the CLC inventory year *y* closest to day *j*; $Elev_i$ is the mean elevation of cell *i*; Pop_i is the population of cell *i*; and ε_{ij} is the error for cell *i* on day *j*.

We used the R package lme4 (Bates et al., 2015) to estimate a single value for the fixed intercept α and each fixed coefficient β_1–β_6 as well as a value in each climatic region on each day for the random intercept μ and the random coefficient υ using maximum likelihood. The random intercept and coefficient allowed the relationship between LST and T_a to vary by day and between climatic regions, improving model fit. We then applied backwards stepwise regression, removing predictors that did not reduce the Akaike information criterion (AIC) by at least 5, and refit the final model using restricted maximum likelihood. We repeated this process for each of the four LST measures (Aqua daytime, Aqua night-time, Terra daytime, and Terra night-time) and selected the model with the lowest 10-fold cross-validated RMSE. We used the final stage 1 model to predict T_a for all 1 km grid cell-days with LST.

1.3.2 Stage 2: predicting 1 km T_a where LST was unavailable

In stage 2, we predicted T_a for the 1 km grid cell-days where LST was not available (usually due to cloud cover). We started by using inverse distance weighting to interpolate daily observed T_a from all weather stations across continental France. We then used all 1 km cell-days with LST to calibrate a mixed model with the equation:

$$T_{ap_{ij}} = (\alpha + \mu_i) + (\beta + \upsilon_i)T_{IDW_{ij}} + \varepsilon_{ij} \quad \text{Eq. (1-2)}$$

where $T_{ap_{ij}}$ is the stage 1 predicted T_a of 1 km grid cell *i* on day *j*; α and β are the fixed intercept and slope, respectively; μ_i and ν_i are the random intercept and slope, respectively, for cell *i*; $T_{IDW_{ij}}$ is the inverse distance weighted T_a of cell *i* on day *j*; and ε_{ij} is the error for cell *i* on day *j*. The random intercept and slope allowed the relationship between T_{ap} and T_{IDW} to vary between grid cells, improving model fit. We used the calibrated model to predict T_a for 1 km grid cell-days where LST was unavailable, then combined the predictions from stage 1 and stage 2 to get daily 1 km predicted T_a (T_{1km}) across the entire study domain.

1.3.3 Stage 3: increasing spatial resolution over large urban areas

In stage 3, we increased the spatial resolution of our predictions over large urban areas. We started by associating each 200 m grid cell with T_{1km} (T_a predicted in stage 2 by the final 1 km model) from the 1 km grid cell that contained the 200 m grid cell. Next, we calculated the residuals (for all 200 m grid cell-days with a weather station T_a observation by subtracting observed T_a from T_{1km}. The number of cell-days with a weather station observation varied by year; on average there were about 462 thousand for T_{mean} and 789 thousand for each of T_{min} and T_{max}. We used these cell-days to train a random forest (RF) and an extreme gradient boosting (XGB) model with the equation:

$$R_{200m_{ij}} = f\begin{pmatrix} T_{1km_{ij}}, T_{b_{im}}, NDVI_{im}, \text{Land Cover}_{ily}, \\ \text{Climate}_i, \text{Elevation}_i, \text{Population}_i, x_i, y_i, j \end{pmatrix} + \varepsilon_{ij} \qquad \text{Eq. (1-3)}$$

where$R_{200m_{ij}}$ is the residual of the 1 km T_a model associated with 200 m grid cell *i* on day *j*; *f* designates the random forest or extreme gradient boosting function; $T_{1km_{ij}}$ is the 1 km T_a model prediction associated with 200 m grid cell *i* on day *j*; $T_{b_{im}}$ is the Landsat top-of-atmosphere brightness temperature of cell *i* for the calendar month *m* in which day *j* falls; $NDVI_{im}$ is the Landsat NDVI of cell *i* for the calendar month *m* in which day *j* falls; Land Cover_{ily} is the fraction of cell *i* occupied by each land cover group *l* in the CLC inventory year *y* closest to day *j*; Climate_i is the climatic region of cell *i*; Elevation_i is the elevation of cell *i*; Population_i is the population of cell *i*; x_i and y_i are the geographical coordinates of cell *i*; j is the Julian day; and ε_{ij} is the error for cell *i* on day *j*.

We used the R packages ranger (Wright & Ziegler, 2017), XGBoost (Chen & Guestrin, 2016), and mlr (Bischl et al., 2016) to train the RF and XGB models. We tuned the models using the sequential model-based optimization of package mlrMBO (Bischl et al., 2017). Briefly,

mlrMBO estimates optimal hyperparameter values by iteratively training and evaluating a model using hyperparameter values that are chosen based on the performance of previous iterations. We used a fixed number of iterations and evaluated performance as the mean RMSE of two random 80% holdouts (i.e. we trained the model on a 20% random sample of the data, predicted and calculate RMSE for the held-out 80%, repeated, and took the mean of the two RMSEs). Initial exploration showed that this resampling approach produced stable estimates of RMSE at a lower computational cost than cross-validation.

For the RF, we used 400 trees and a minimum of 5 observations per node, and tuned mtry (the number of variables to consider for each split) from 3 to 12 (25% to 100% of the explanatory variables) using 6 mlrMBO iterations. Initial exploration showed that using more than 400 trees only marginally increased performance and had a high computational cost. For the XGB model, we used the gbtree booster with 100 rounds and set gamma (the minimum loss reduction for a split) to 5. We used 24 mlrMBO iterations to tune eta (the learning rate) from 0.1 to 0.3, the maximum tree depth from 5 to 20, the minimum number of observations per node from 3 to 30, and the fraction of features used per tree from 0.75 to 1.

We evaluated the performance of the stage 3 models using 5-fold cross-validation with nested tuning. We used the final stage 3 RF and XGB models to predict the residual of the 1 km T_a model (R^{RF}_{200m} and R^{XGB}_{200m}, respectively) for all 200 m cell-days.

1.3.4 Stage 4: improving 200 m predictions

In stage 4, we improved the stage 3 predictions by ensembling. We use all 200 m grid cell-days with a weather station T_a observation to calibrate a generalized additive model (GAM) with the formula:

$$R_{200m_{ij}} = t(x_i, y_i,)R^{RF}_{200m_{ij}} + t(x_i, y_i)R^{XGB}_{200m_{ij}} + \varepsilon_{ij} \quad \text{Eq. (1-4)}$$

where $R_{200m_{ij}}$ is the residual of the 1 km T_a model associated with 200 m grid cell *i* on day *j*; $t(x_i, y_i)$ is a tensor product smooth of the x and y coordinates of cell *i*; $R^{RF}_{200m_{ij}}$ and $R^{XGB}_{200m_{ij}}$ are the predicted residuals of the 1 km T_a model from the stage 3 RF and XGB model, respectively, for cell *i* on day *j*; and ε_{ij} is the error for cell *i* on day *j*. The GAM averages the RF and XGB predicted residuals using weights that vary both by location and with the magnitude of each model's predicted residual. Finally, we added the ensemble-predicted residuals for all 200 m

grid cells to T_{1km} (T_a predicted in stage 2 by the final 1 km model) to obtain daily 200 m predicted T_a (T_{200m}) over large urban areas.

1.3.5 Performance assessment

We estimated the models' accuracy using 10-fold out-of-sample cross-validation. For the RF and XGB model we used nested tuning (i.e. within each cross-validation fold we tuned the model as described in section 1.3.3). To evaluate the models' ability to capture both spatial and temporal patterns in T_a, we also calculated the spatial and temporal components of the errors. The spatial component is the difference at each station between the annual mean of daily observed T_a ($\overline{T_a}$) and the annual mean of daily predicted T_a ($\overline{T_{ap}}$). The temporal component is the difference at each station between ΔT_a and ΔT_{ap} where ΔT_a is the difference between daily observed T_a and $\overline{T_a}$; ΔT_{ap} is the difference between daily predicted T_a and $\overline{T_{ap}}$.

We used Google Earth Engine (Gorelick et al., 2017) to quality assure and composite Landsat T_b and NDVI and aggregate them to the 200 m grid cells. For all other data processing and analyses we used R version 3.4.4 (R Core Team, 2018).

1.4 Results

Table 1-2 presents the mean 10-fold cross-validated performance of the stage 1 models (predicting daily 1 km T_a from LST) across all years. The models performed very well, with R^2 of 0.92 or higher, RMSE of less than 2°C, and mean absolute error (MAE) of less than 1.5°C. All models had very low bias: the slope of observed *vs.* predicted T_a was 1.00 while the intercept ranged from 0.01 to 0.02. The T_{mean} models performed best overall (MAE 0.94), followed by the T_{max} (MAE 1.35) and T_{min} (MAE 1.43) models. The models captured both spatial and temporal variation in T_a and showed little variation in performance between years, although overall T_{mean} performance decreased slightly after 2010, possibly reflecting degradation of the Terra MODIS instrument (Table A1-3). Consistent with previous studies (Oyler et al., 2016; Rosenfeld et al., 2017; Yoo et al., 2018), night-time LST was the best predictor of T_{mean} (Table A1-3) and T_{min} (Table A1-4) while daytime LST was the best predictor of T_{max} (Table A1-5). Aqua LST was a better predictor of T_{min} and T_{max} while Terra LST was a better predictor of T_{mean}. This was expected as the Aqua overpasses (approximately 1:30 and 13:30 local solar time) are closer to the time at which T_{min} and T_{max} typically occur in France. However, Aqua LST was only available since July 2002, so we used Terra LST for all models prior to 2003.

Table 1-2. Stage 1 model (predicting daily 1 km T_a from LST): 10-fold cross-validated performance across all years (2000 to 2018), overall, spatial, and temporal components.

		Overall			Spatial			Temporal		
	N*	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
T_{min}	348	0.92	1.89	1.42	0.89	1.11	0.81	0.94	1.61	1.19
T_{mean}	204	0.97	1.29	0.94	0.95	0.83	0.57	0.97	1.15	0.84
T_{max}	319	0.95	1.81	1.35	0.88	1.23	0.89	0.96	1.51	1.12

* N = mean thousands of observations used to fit each annual model

Table 1-3 presents the 10-fold cross-validated performance of the stage 1 models across all years by calendar month and season. The T_{min} and T_{mean} models performed slightly less well in winter months, possibly due to higher LST missingness from more frequent cloud cover. The T_{max} model performed best in late winter, early spring, and fall. The models performed less well in the mountain, semi-continental, and modified Mediterranean climates (Table A1-6). Those climates occur in mountainous areas where large contrasts in topography and land cover make modelling particularly challenging; other factors not included in the model may also have reduced performance in those areas. The models performed slightly better in peri-urban areas than in urban and rural areas (Table A1-6), possibly due to the higher density of weather stations (peri-urban areas had the most stations per km^2).

Table 1-3. Stage 1 model performance (predicting daily 1 km T_a from LST): 10-fold cross-validated performance across all years (2000 to 2018), by month and season.

	T_{min}			T_{mean}			T_{max}		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Jan	0.83	2.16	1.59	0.89	1.55	1.11	0.86	1.87	1.37
Feb	0.84	2.02	1.50	0.91	1.36	0.99	0.89	1.74	1.28
Mar	0.80	1.92	1.46	0.91	1.23	0.91	0.89	1.71	1.28
Apr	0.78	1.82	1.40	0.91	1.16	0.86	0.87	1.75	1.32
May	0.81	1.74	1.32	0.92	1.19	0.85	0.85	1.84	1.38
Jun	0.81	1.73	1.31	0.92	1.23	0.90	0.84	1.93	1.45
Jul	0.79	1.70	1.30	0.91	1.19	0.88	0.84	1.90	1.44
Aug	0.79	1.77	1.35	0.92	1.18	0.89	0.86	1.88	1.43
Sep	0.79	1.83	1.41	0.92	1.13	0.84	0.87	1.70	1.29
Oct	0.83	1.95	1.48	0.91	1.27	0.93	0.88	1.66	1.25
Nov	0.82	2.02	1.51	0.89	1.42	1.03	0.88	1.69	1.25
Dec	0.82	2.16	1.60	0.87	1.67	1.20	0.84	1.94	1.39
Winter	0.83	2.12	1.57	0.89	1.54	1.10	0.87	1.85	1.35
Spring	0.86	1.83	1.39	0.94	1.19	0.87	0.91	1.77	1.33
Summer	0.80	1.73	1.32	0.92	1.20	0.89	0.85	1.90	1.44
Fall	0.86	1.92	1.46	0.95	1.26	0.92	0.93	1.68	1.26

Figure 1-2 shows the spatial pattern of the daily 1 km T_a predictions of the stage 2 model on selected winter and summer days. For the cold winter day of Feb 18, 2003, predictions range from T_{min} of -17°C in parts of the Alps, the Massif Central, and the Pyrenees to T_{max} of 11°C on the Mediterranean coast. The urban heat island of Paris is faintly visible in the north centre of the T_{min} and T_{mean} maps but disappears on the T_{max} map. Spatial contrasts corresponding to terrain features are well resolved, and the spatial pattern of T_{min} vs. T_{mean} vs. T_{max} varies most in the north, northeast, and southwest.

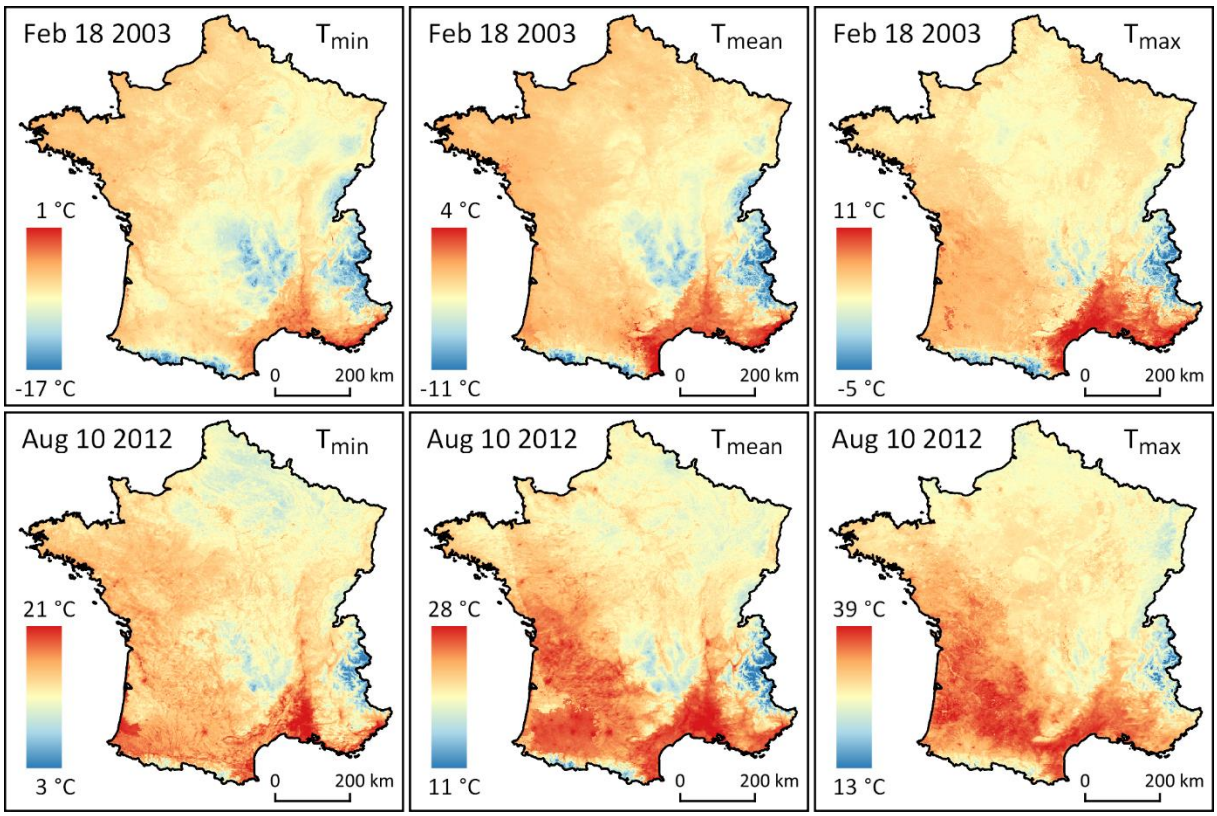

Figure 1-2. Predicted 1 km T_a from the stage 2 model on selected days: Feb 18, 2003 (top row) and Aug 10, 2012 (bottom row).

For the hot summer day of Aug 10, 2012, predictions range from a T_{min} of 3°C in parts of the Alps to a T_{max} of 39°C in the southeast and southwest. On the T_{min} map, the southwestern cities of Toulouse and Bordeaux stand out as hotspots, while Paris and Rouen are faintly visible as warm spots in the north. The north is colder than the Vosges mountains in the northeast and the Pyrenees in the southwest are warmer than the alps. The warmest areas are the southern Rhone river valley in the southeast and a patch of the southwestern Atlantic coast. On the T_{mean} map, Paris and Rouen are still visible, Lyon stands out in the east, and a few northwestern cities appear. Much of the southwest is as warm as the southeast, and the southwestern cities are harder to distinguish from the countryside. On the T_{max} map, Lyon, Rouen, and some

northwestern cities remain faintly visible, Pau and Tarbes appear in the southwest, and the north is warmer than the Vosges.

Table 1-4 presents the 10-fold cross-validated performance of the stage 4 models (predicting daily 200 m residuals of the 1 km model using an ensemble) across all years and by month and season. The stage 4 models also performed well, with overall R^2 of 0.79 to 0.85, RMSE of 0.41 to 0.63, and MAE of 0.26 to 0.39 (residual scale). As with the stage 1 models, the R_{Tmean} predictions were slightly better than the R_{Tmin} or R_{Tmax} predictions and the models performed least well in the mountain, semi-continental, and modified Mediterranean climates (Table A1-7). The R_{Tmin} model performed slightly worse in late summer; otherwise performance was quite consistent across months and seasons. The models had low bias, with a slope of observed *vs.* predicted of 1.00 and intercept of zero for every year. Performance was consistent across years except for the R_{Tmin} model, which performed slightly better in 2000 to 2002, and the R_{Tmean} model, which performed best in 2004 (Table A1-8).

Table 1-4. Stage 4 model performance (predicting daily 200 m residuals with an ensemble): 10-fold cross-validated performance across all years (2000 to 2018), overall and by month and season (residual scale).

	R_{Tmin}			R_{Tmean}			R_{Tmax}		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Overall	0.79	0.63	0.40	0.79	0.41	0.26	0.84	0.52	0.31
Jan	0.84	0.57	0.34	0.82	0.40	0.24	0.85	0.48	0.27
Feb	0.81	0.59	0.36	0.81	0.39	0.24	0.84	0.49	0.28
Mar	0.79	0.63	0.39	0.78	0.40	0.26	0.83	0.50	0.30
Apr	0.77	0.65	0.41	0.76	0.39	0.25	0.83	0.52	0.32
May	0.77	0.60	0.37	0.76	0.38	0.24	0.84	0.51	0.31
Jun	0.77	0.63	0.40	0.78	0.39	0.25	0.86	0.53	0.34
Jul	0.75	0.66	0.44	0.77	0.42	0.28	0.86	0.55	0.35
Aug	0.76	0.68	0.44	0.78	0.42	0.28	0.86	0.54	0.35
Sep	0.76	0.70	0.46	0.75	0.43	0.29	0.84	0.54	0.35
Oct	0.77	0.66	0.42	0.76	0.43	0.27	0.81	0.53	0.32
Nov	0.79	0.61	0.37	0.78	0.42	0.25	0.81	0.51	0.29
Dec	0.83	0.60	0.37	0.83	0.43	0.27	0.83	0.52	0.30
Winter	0.83	0.59	0.36	0.82	0.41	0.25	0.84	0.50	0.29
Spring	0.78	0.63	0.39	0.77	0.39	0.25	0.84	0.51	0.31
Summer	0.76	0.66	0.43	0.78	0.41	0.27	0.86	0.54	0.34
Fall	0.78	0.66	0.42	0.76	0.42	0.27	0.82	0.52	0.32

Spatial location and elevation were generally the most important features in the RF (Figure A1-2**)** and XGB models (Figure A1-3). Day of year and predicted 1 km T_a were equally or even

more important in some models but less important in others. Landsat T_b and NDVI and population also contributed to the models, particularly for R_{Tmean}. The land cover and climatic region variables were the least important.

Figure 1-3 shows the spatial pattern of predicted 1 km T_{min} from the stage 2 model and predicted 200 m T_{min} from the stage 4 model for the Paris metropolitan area (northern France, population 12.5 million), the Toulouse metropolitan area (southwestern France, Population 1.3 million), and the Nancy metropolitan area (northeastern France, population 250 000) on the cold winter day of Feb 18, 2003. Figure A1-4, Figure A1-5, and Figure A1-6 illustrate the difference between the 1 km and 200 m predictions for Paris, Toulouse, and Nancy on selected hot and cold days. On Feb 18, 2003, an urban heat island is clearly visible over the large urban core of Paris where T_{min} is about 5°C warmer than the rural surroundings. The 200 m predictions are slightly higher than the 1 km predictions in the peripheral built-up areas and capture fine details such as the warmer Seine river and cooler parks.

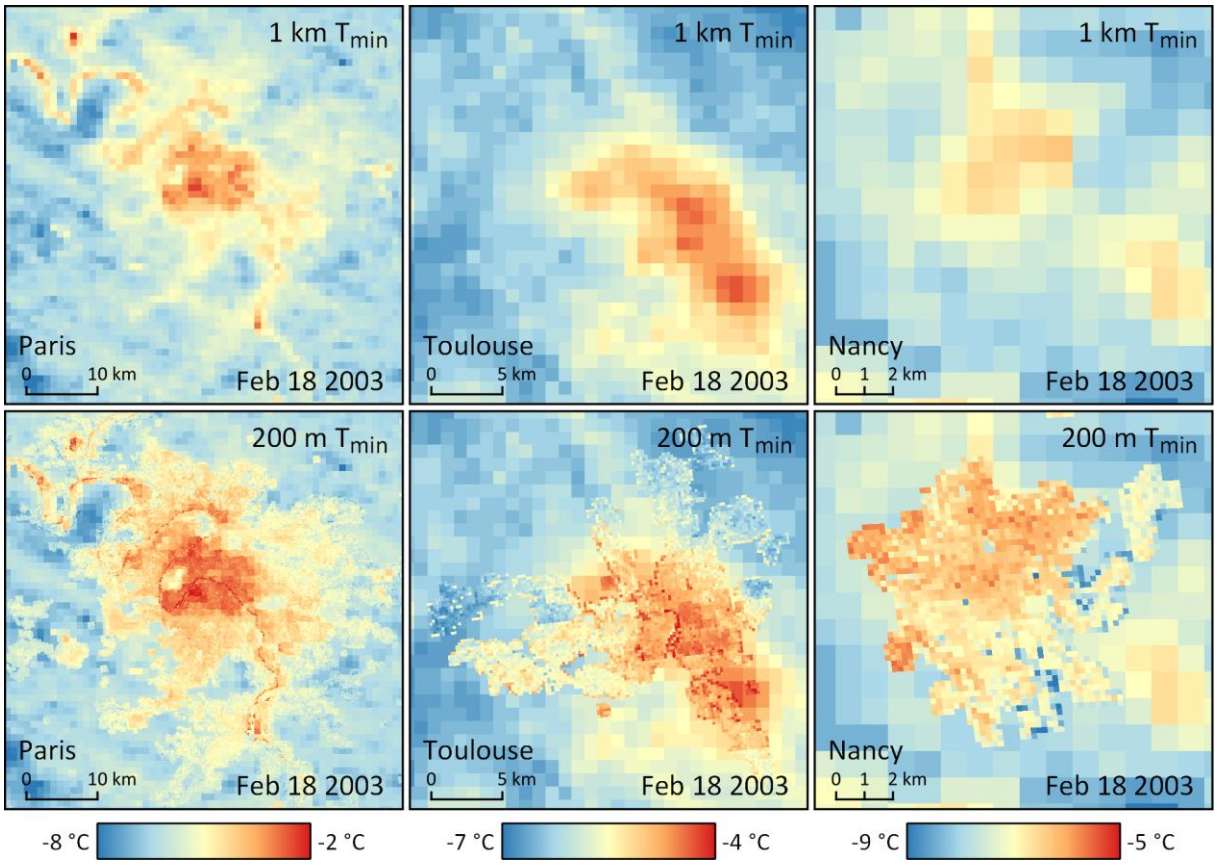

Figure 1-3. Predicted 1 km T_{min} from the stage 2 model alone (top row) and with predicted 200 m T_{min} from the stage 4 model overlaid (bottom row) on Feb 18, 2003 over the Paris, Toulouse, and Nancy metropolitan areas.

In the midsize city of Toulouse, the 1 km predictions capture an urban heat island over the dense city centre and the suburbs to the northwest and southeast, with T_{min} about 3°C warmer than the rural surroundings. The 200 m predictions show warm T_{min} in the southwestern suburbs where 1 km T_{min} was cool and capture the Garonne river in the centre. The northwestern and northeastern suburbs have greater contrast with some areas slightly cooler than in the 1 km predictions and others slightly warmer.

In the small city of Nancy, at 1 km both the city centre and an area of ponds to the southeast have T_{min} about 2°C warmer than the surroundings. The 200 m predictions show warmer T_{min} throughout most of the built-up area with sharp contrasts between built and open areas: compared to the 1 km predictions, T_{min} is up to 2°C higher in the centre, north, and west of the built-up area and up to 2°C lower over parks and over fields abutting the eastern edges of the city.

1.5 Discussion

Spatiotemporally resolved T_a at high resolutions is essential to understanding, monitoring, and managing the health effects of T_a, a pressing issue in a warming, urbanizing world. To our knowledge, our satellite-based geostatistical temperature model covers a longer period (2000 to 2018) at a higher spatial resolution (1 km) than any other model designed for public health research in France. Furthermore, our model provides an unprecedented spatial resolution of 200 m over 103 large urban areas.

A key feature of our model is its ability to capture spatial variation in T_a. Previous epidemiological research in France linked geographical variation in mortality risk to both typical (Laaidi et al., 2006) and extreme T_a (Le Tertre et al., 2006) using weather stations. Recent studies in the U.S. showed that a daily 1 km T_a dataset similar to ours was needed to detect associations with low birth weight (Kloog et al., 2015a) and mortality (Shi et al., 2015). Our model will allow future studies in France to include participants in rural areas far from weather stations and will also improve exposure estimates in urban areas.

Another key feature is our model's 200 m spatial resolution over urban areas. Estimating T_a exposure in cities is particularly challenging due to complex built environments and the scarcity of representative T_a measurements, as weather stations tend to be located outside cities (e.g. at airports) or in large parks. Consequently, few epidemiological studies have examined intra-urban variation in T_a. In Milan, Italy, de'Donato et al. (2008) found that on hot summer days

temperature measured at a nearby airport tended to be higher and more strongly associated with mortality than temperature measured in the city centre, but in Turin and Rome there was little difference in temperature or its association with mortality between the city centre and a nearby airport. In Paris, France, Laaidi et al. (2012) used 1 km LST as a proxy for T_a and found an association between minimum LST and mortality during the August 2003 heatwave. In Brisbane, Australia, Guo et al. (2013) found no significant difference in the mortality–T_a relationship when estimating T_a exposure using a central weather station *vs.* kriging, although they noted that there was little spatial variation in temperature across the city. In Seattle, USA, Ho et al. (2017) found a significant association between spatial variation in mortality on extremely hot days and modelled humidex (a measure of both T_a and humidity). Our model will help future studies clarify the health effects of intra-urban T_a variation.

Our model's unique combination of lower spatial resolution (1 km) predictions over a large geographical extent and higher spatial resolution (200 m) predictions over more densely populated areas will be particularly helpful for epidemiological studies. Broad geographical coverage is essential to including rural residents which have often been excluded from epidemiological studies, especially in France where the 103 largest urban areas covered by our 200 m T_a model contain less than half of the population. At the same time, high spatial resolution is important in dense urban areas where T_a can vary at fine spatial scales and the effect of spatial T_a variation is less well understood. Limiting the 200 m resolution predictions to large urban areas reduces computational effort while still covering a large portion of the population.

A fourth feature of our model is its ability to predict daily T_{min}, T_{mean}, and T_{max}. While T_{mean} suffices for many health studies (Barnett et al., 2010), certain research questions may benefit from having T_{min} and T_{max}. For example, heatwave studies may wish to use heatwave definitions that refer to T_{min} or T_{max} or explore whether certain populations are sensitive to T_{min} or night-time T_a (Laaidi et al., 2012; Murage et al., 2017). T_{max} might also be of interest because it tends to occur in the afternoon when people are more likely to be outside and active (Yuming Guo et al., 2017). T_{min} and T_{max} also allow calculating diurnal T_a range for studies of T_a variability and delineating diurnal and nocturnal urban heat islands for urban climate studies.

We demonstrated that allowing the relationship between 1 km LST and T_a to vary by climatic region as well as by day slightly improves performance: our stage 1 T_{mean} model achieves overall R^2 of 0.97 with RMSE of 1.29 whereas a previous model achieved R^2 of 0.96 with

RMSE of 1.52 (Kloog et al., 2017). We also demonstrated that a GAM ensemble of machine learning models can use higher spatial resolution predictors including Landsat thermal data to account for some of the residual error in our daily 1 km T_a predictions. Adding this local stage both increased the spatial resolution of our model and improved performance.

One limitation of our method is its reliance on historical satellite thermal data. Our model is restricted to the MODIS period of record, which starts in 2000. Older thermal data is available from other satellites (e.g. Landsat), but not with a twice-daily revisit time. In the U.S., Oyler et al. (2015) showed that an anomaly-climatology approach could model daily T_{min} and T_{max} since 1948 from 8-day composite MODIS LST, although their approach may have smoothed spatiotemporal T_a trends.

Our model can estimate past T_a but, unlike numerical weather prediction models, cannot forecast future T_a. However, our model is much simpler, which allows us to run it at relatively high spatial resolutions (1 km and 200 m) over large areas and long periods. In comparison, Météo France's weather prediction model has run at a spatial resolution of 1.3 km only since 2015, and the ECMWF's most recent ERA5 reanalysis has a spatial resolution of just 30 km. Also, recent studies suggest that incorporating LST from geostationary satellites might allow us to estimate close to real-time T_a (Bechtel et al., 2017; Keramitsoglou et al., 2016) or possibly even forecast next-day T_a from present-day MODIS LST (Yoo et al., 2018).

Another limitation of our approach is the temporal misalignment between observations of LST and T_a in the stage 1 model: the satellite overpass does not always coincide with the time that T_{min} or T_{max} occurs. Our model's low MAE (typically less than 1.5°C) suggests that it produces good T_a estimates despite this; incorporating high temporal resolution (e.g. hourly) LST from geostationary satellites might improve performance.

A fourth limitation of our model is the need to fill gaps in satellite thermal data. This can introduce error and may make modelling impossible in some areas or periods. Landsat data is particularly challenging due to the satellites' 16-day revisit time; parts of France have no usable Landsat observations during some winters. The few previous studies that used Landsat thermal data to model T_a limited their analysis to days and locations where Landsat data was available (Pelta & Chudnovsky, 2017) or used a few scenes that were deemed typical of hot summer days (Ho et al., 2016, 2014; Wicki et al., 2018). We fill gaps in Landsat T_b by compositing all scenes for each calendar month across 17 years. This smooths spatial patterns and means we rely

entirely on MODIS to capture short-term temporal variation in LST. Combining data from Landsat 5, 7, and 8 may also introduce error as the sensors operate at different wavelengths and spatial resolutions (Table 1-1). Future studies may benefit from the forthcoming Landsat Surface Temperature product which might be more consistent, and would allow using LST as a predictor rather than brightness temperature.

Future studies could also make use of high spatial resolution LST from forthcoming satellites. Landsat 9 will have a spatial resolution and revisit time similar to the previous Landsat satellites but should offer better LST retrieval thanks to the correction of the stray light contamination that affects Landsat 8 (Hair et al., 2018). HyspIRI aims to provide a 60 m spatial resolution with a revisit time of 5 days, while MISTIGRI aims for 50 m spatial resolution with a daily revisit, but with coverage only within 15 ground tracks. If these satellites improve LST retrieval and reduce missingness then they could improve our method's ability to capture T_a over urban areas.

MODIS LST also contains gaps, which we do not fill. Rather, we predict daily 1 km T_a only where MODIS LST is available and fill gaps in the predictions based on nearby T_a observations. Li et al. (2018b) achieved similar performance (RMSE 2.1°C T_{min}, 1.9°C T_{max}) for urban and surrounding areas in the U.S. by first filling gaps in MODIS LST using spatiotemporally nearby LST observations and then predicting daily T_a using geographically weighted regression. These approaches both assume that the spatial distribution of T_a or LST is similar on clear and cloudy days. Zhu et al. (2017) used the MODIS atmospheric profile and cloud cover products to estimate instantaneous T_a in parts of China and the U.S.. Their approach had the additional advantage of not requiring any weather station T_a observations to calibrate the model, but it produced larger errors (RMSE 3.4°C China, 2.9, U.S.).

Despite these limitations, our model provides very good predictions of historical daily T_a for continental France at a 1 km or finer spatial resolution. These predictions may help compare rural and urban populations, identify and monitor urban heat islands, and better understand health effects. More broadly, our methodology and predictions may be useful in other geographical areas and for any application where T_a is a key variable.

Chapter 2

Modelling daily ambient PM concentration in France

Ian Hough[a,b], Ron Sarafian[b,c], Alexandra Shtein[b], Bin Zhou[b,d], Johanna Lepeule[a], Itai Kloog[b]

[a] *Institute for Advanced Biosciences, Université Grenoble Alpes, INSERM, CNRS, La Tronche, France*

[b] *Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Be'er Sheva, Israel*

[c] *Department of Industrial Engineering, Ben-Gurion University of the Negev, Be'er Sheva, Israel*

[d] *Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany*

Portions of this chapter were previously published as:

Hough, I., Sarafian, R., Shtein, A., Zhou, B., Lepeule, J., & Kloog, I. (2021). Gaussian Markov random fields improve ensemble predictions of daily 1 km PM2. 5 and PM10 across France. *Atmospheric Environment*, 118693. https://doi.org/10.1016/j.atmosenv.2021.118693

2.1 Introduction

The second objective of this PhD was to develop a satellite-based model of ambient particulate matter air pollution to estimate the exposure of participants in epidemiological studies in France. The model needed to estimate daily $PM_{2.5}$ and PM_{10} concentration over continental France from 2000 to 2019. We knew of no existing PM dataset covering continental France during this period at a daily 1 km resolution or better; the main existing datasets were atmospheric composition reanalyses with up to hourly 60 km resolution. While this PhD was underway, a study was published that estimated hourly PM concentration in France only for the years 2010 and 2011 by combining 51 chemical transport models: a national model at 4 km resolution, 7 regional models at 3 to 4 km resolution, and 43 urban models at 10 to 200 m resolution (Riviere et al., 2019).

We advanced the state of the art by creating the first satellite-based geostatistical model of daily 1 km $PM_{2.5}$ and PM_{10} over continental France. Our model has three key features. First, we improved the accuracy of $PM_{2.5}$ estimates by imputing $PM_{2.5}$ at monitors that only measured PM_{10}, increasing the number of monitors available to train the $PM_{2.5}$ model. Second, we created a gap-free AOD dataset by imputing missing MAIAC AOD based on modelled AOD from the atmospheric composition reanalyses. This allowed us to predict PM for all days and locations based on satellite-derived AOD. Third, we improved overall accuracy by developing an approach combining Gaussian Markov random fields (GMRF) and random forests in an ensemble framework.

GMRFs address a limitation of mixed models, which only allow the PM-AOD relationship to vary between predefined hierarchical blocks of data (e.g. a block for each day, within which there is a block for each region). This can result in sharp contrasts (e.g. the regression coefficient for AOD may change abruptly at the border between two regions), which may be unrealistic since the physical processes that determine temperature and PM concentration often vary smoothly over space and time. Mixed models also struggle to incorporate potentially useful information about the distance between blocks (e.g. 1 January is closer in time to 1 March than to 1 May). GMRFs can calibrate continuous spatiotemporal processes in a way that is computationally feasible for large datasets (Lindgren et al., 2011; Rue et al., 2009). Recent work has shown that GMRFs are applicable to PM modelling (Beloconi et al., 2018; Cameletti et al., 2013), and a study in the northeastern United States found that GMRF predicted daily 1 km $PM_{2.5}$ more accuratly than a mixed model (Sarafian et al., 2019). Ours was the first study

to incorporate a GMRF in an ensemble PM model and directly compare GMRF to random forest, a popular machine learning algorithm that is known to have good performance.

Our final model reconstructs daily $PM_{2.5}$ and PM_{10} from 2000 to 2019 at a 1 km spatial resolution over continental France using a four-stage process (Figure 2-1). It performs well, with mean cross-validated R^2 of 0.76 ($PM_{2.5}$) or 0.71 (PM_{10}) and MAE of 2.72 ($PM_{2.5}$) or 4.26 (PM_{10}). It also provides the first evidence that GMRFs may predict daily PM concentration more accurately than random forests. Our approach may advance particulate matter modelling in other areas, and our daily PM dataset is available to assess the exposure of participants in epidemiological studies in France and thus improve our understanding of PM health effects.

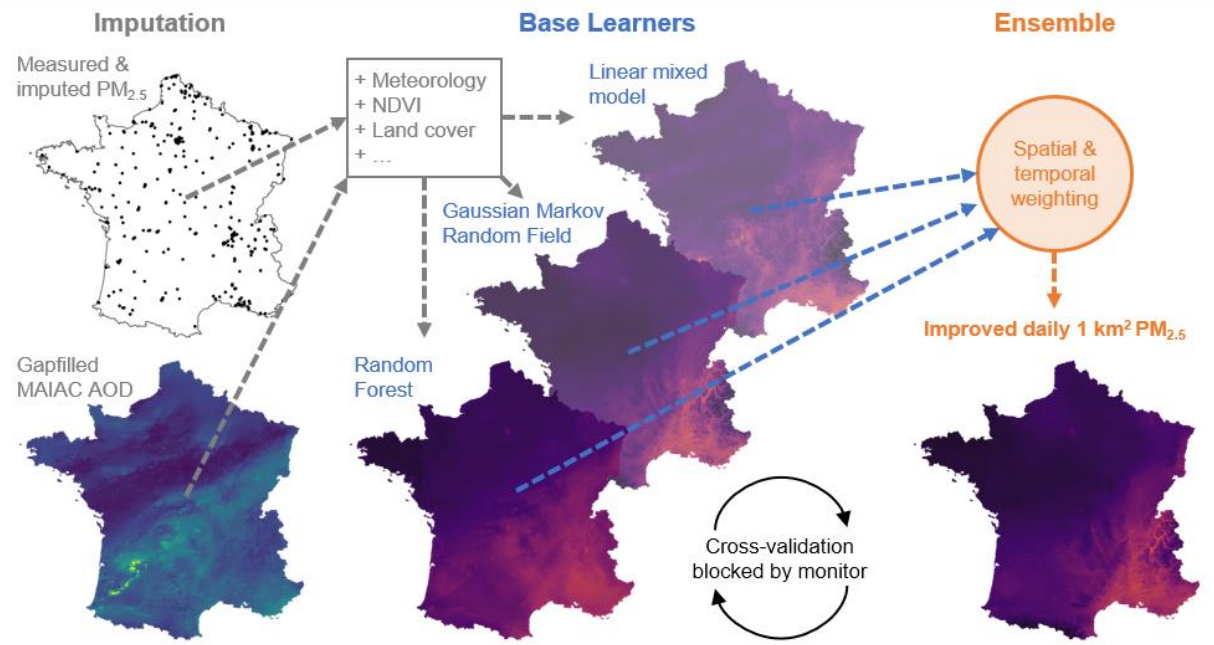

Figure 2-1. Overview of the four-stage modelling process used in this study.

2.2 Materials

2.2.1 Study domain

As described in Chapter 1, section 1.2.1, continental France covers a roughly hexagonal area of 542,973 km^2 in western Europe bounded by the Atlantic Ocean to the west and the Mediterranean Sea to the southeast (Figure 2-2). The population is approximately 64.5 million, of which 20% live in the urban area of Paris and 19% in rural areas. We defined a grid of 632,571 approximately 1 km^2 cells covering continental France coincident with the pixels of the satellite AOD data (see section 2.2.3). We considered the 7245 days from 1 March 2000 through 31 December 2019, giving a total study domain of 4.58×10^9 cell-days.

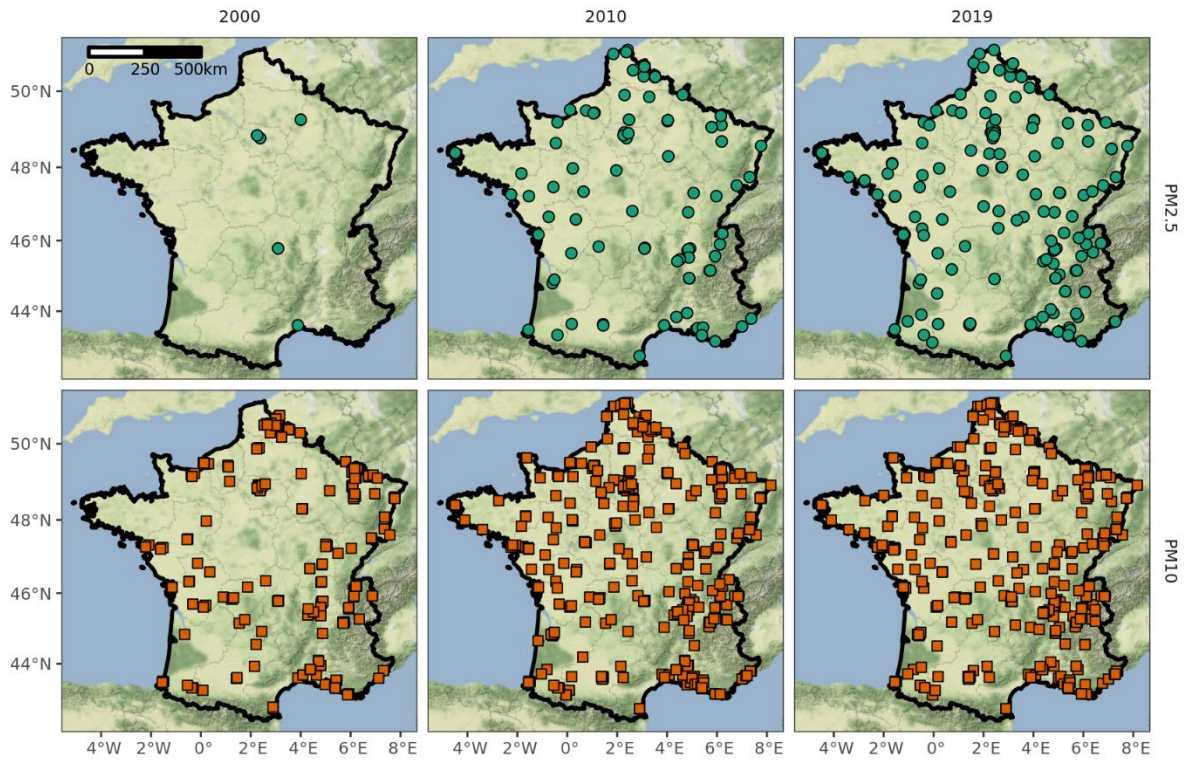

Figure 2-2. Spatial distribution of $PM_{2.5}$ monitors (top row) and PM_{10} monitors (bottom row) in continental France from 2000 to 2019. Basemap by Stamen Design, under CC BY 3.0.

2.2.2 Air quality monitoring data

We obtained hourly $PM_{2.5}$ and PM_{10} ($\mu g/m^3$) measurements from 12 regional air quality monitoring networks (federated by ATMO France) through the French Central Air Quality Monitoring Laboratory. Monitors were mostly clustered in urban areas; the number of PM_{10} monitors increased from 222 to 330 and the number of $PM_{2.5}$ monitors increased from 9 to 142 over the course of the study period (Figure 2-2). Prior to 2007, all monitors measured non-volatile particles via tapered element oscillating microbalances or beta gauge monitors. In 2007, all PM_{10} monitors were equipped with filter dynamics measurement systems or regulated sampling tubes to additionally measure semi-volatile particles, thus increasing measured PM concentrations. All $PM_{2.5}$ monitors were upgraded from 2008 to 2009. To limit the impact of instrument malfunctions and rare events, we excluded hourly $PM_{2.5}$ concentrations > 200 $\mu g/m^3$ and hourly PM_{10} concentrations > 300 $\mu g/m^3$ (0.003% of all observations). We indexed each monitor to the containing 1 km grid cell and calculated daily mean PM for days with at least 18 hourly observations.

2.2.3 Aerosol optical depth

We obtained satellite-derived 0.469 µm AOD at approximately 1 km spatial resolution from the Moderate Resolution Imaging Spectroradiometer (MODIS) Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD product (MCD19A2v006) (Lyapustin et al., 2018). MCD19A2v006 provides AOD up to four times per day (between 9:00 and 15:00 UTC in France). We used the quality assurance band to identify all "best quality" observations; we also included "land; research quality" and "clear; within 2 km of coast" as these represent potentially useable observations over urban areas and coasts where there are few "best quality" observations. We indexed these observations to the 1 km grid (whose cells were defined to coincide with the MAIAC AOD pixels) and calculated daily mean AOD.

To fill gaps in MAIAC AOD (mostly due to cloud cover), we obtained modelled 3-hourly 0.469 µm AOD at approximately 80 km spatial resolution from the Copernicus Atmospheric Monitoring Service EAC4 Reanalysis (Inness et al., 2019). Since EAC4 begins on 1 March 2003, for 1 March 2000 to 28 February 2003 we obtained modelled hourly 0.55 µm AOD for 08:30 to 15:30 UTC at approximately 60 km spatial resolution from the MERRA2 reanalysis (Randles et al., 2017). We bilinearly interpolated EAC4 and MERRA2 AOD to the 1 km grid, giving 8 values per cell-day (0 UTC, 3 UTC, …, 21 UTC for EAC4; 08:30 UTC, 09:30 UTC, …, 15:30 UTC for MERRA2).

2.2.4 Meteorology

Meteorological parameters such as wind, rain, temperature, and the height of the planetary boundary layer affect surface PM concentrations and indicate the extent to which AOD represents aerosols near the surface or higher in the atmosphere. We obtained hourly meteorological parameters at approximately 30 km spatial resolution from the Copernicus Climate Change Service ERA5 reanalysis (Hersbach et al., 2020). We bilinearly interpolated the parameters to the 1 km grid and calculated 10 daily values: boundary layer height at 0:00 and 12:00 UTC, total precipitation, mean and standard deviation of 2m air temperature, mean 2m dewpoint temperature, mean surface pressure, mean u- and v-components of 10m wind speed, and mean cloud cover.

2.2.5 Normalized difference vegetation index

Vegetation may influence PM dispersion and the density of PM sources. We obtained monthly composite NDVI at approximately 1 km^2 spatial resolution from the MODIS MOD13A3v006

product (Didan et al., 2015), which is spatially coincident with the MAIAC AOD data. We indexed NDVI to the 1 km grid, filled rare missing values with the Gaussian kernel mean of nearby cells, and used the same value for every day of each month.

2.2.6 Spatial predictors

In addition to the previous spatiotemporal predictors, we used impervious surfaces, land cover, road and railway density, elevation, population, climatic region, distance to coast, and $PM_{2.5}$ and PM_{10} emissions as indicators of the typical spatial distribution of PM. Since these data are time invariant, we used the value from the closest reference year for every day of each year. Table 2-1 describes the 19 spatial predictors we derived from these data.

Table 2-1. Spatial predictors used as indicators of the typical spatial distribution of PM.

Dataset	Description	N[a]	Reference year(s)	Spatial Resolution	Source
Impervious surfaces	Fraction of grid cell covered by impervious surfaces	1	2006 2009 2012 2015 2018	100 m	Copernicus Land Monitoring Service
Land cover	Fraction of grid cell covered by 5 land cover classes[b]	5	2000 2006 2012 2018	100 m	Corine Land Cover v20u1
Roads	Length within grid cell for each of 5 road classes[c]	5	2015	Vector	Jedlička et al. (2016)
Railways	Length within grid cell for each of 2 railway classes[d]	2	2016	Vector	BD-TOPO v2.2
Emissions	Monthly & annual $PM_{2.5}$ and PM_{10} emissions[e]	2	2007	750 m	INERIS
Population	Total population of grid cell	1	2010	200 m	Insee (2016)
Elevation	Mean elevation of grid cell	1	2016	25 m	EU-DEM v1.1
Coasts	Distance from coast to grid cell	1	2009	Vector	EU-Hydro v1.1
Climate	Climatic region of grid cell[f]	1	2010	250 m	Joly et al. (2010)

[a] Number of predictors
[b] Buildings; urban greenspace; transportation facilities; vegetation; water
[c] Highways; primary; secondary; tertiary; local
[d] Diesel; electric
[e] $PM_{2.5}$ (PM_{10}) models used only $PM_{2.5}$ (PM_{10}) emissions
[f] Montane; semi-continental; transitional oceanic; modified oceanic; oceanic; modified Mediterranean; Mediterranean; southwest basin

2.3 Methods

We used a four-stage process to predict $PM_{2.5}$ and PM_{10} for the 4.58×10^9 1 km grid cell-days in the study domain (Figure 2-3). Briefly, we: 1) alleviated the sparsity of $PM_{2.5}$ monitors by training a random forest (RF) to impute daily $PM_{2.5}$ at monitors that only measured PM_{10}; 2) filled gaps in MAIAC AOD data by training monthly RFs to impute missing MAIAC AOD based on co-located EAC4 or MERRA2 AOD; 3) trained three base learners for each year

(linear mixed models [LMM], Gaussian Markov random fields [GMRF], and RF) to predict daily 1 km PM based on gap-filled AOD, meteorology, NDVI, and spatial predictors; 4) increased accuracy by ensembling the base learner predictions with annual generalized additive models (GAM) that weight the base learners according to spatiotemporal variations in their performance. We performed all data processing and statistical analyses in R 3.6.3 (R Core Team, 2020) using the packages lme4 for LMM (Bates et al., 2015), R-INLA for GMRF (Bakka et al., 2018), ranger for RF (Wright & Ziegler, 2017) with mlr and mlrMBO for tuning via model-based optimization (Bischl et al., 2017, 2016), and mgcv for GAM (Wood, 2017).

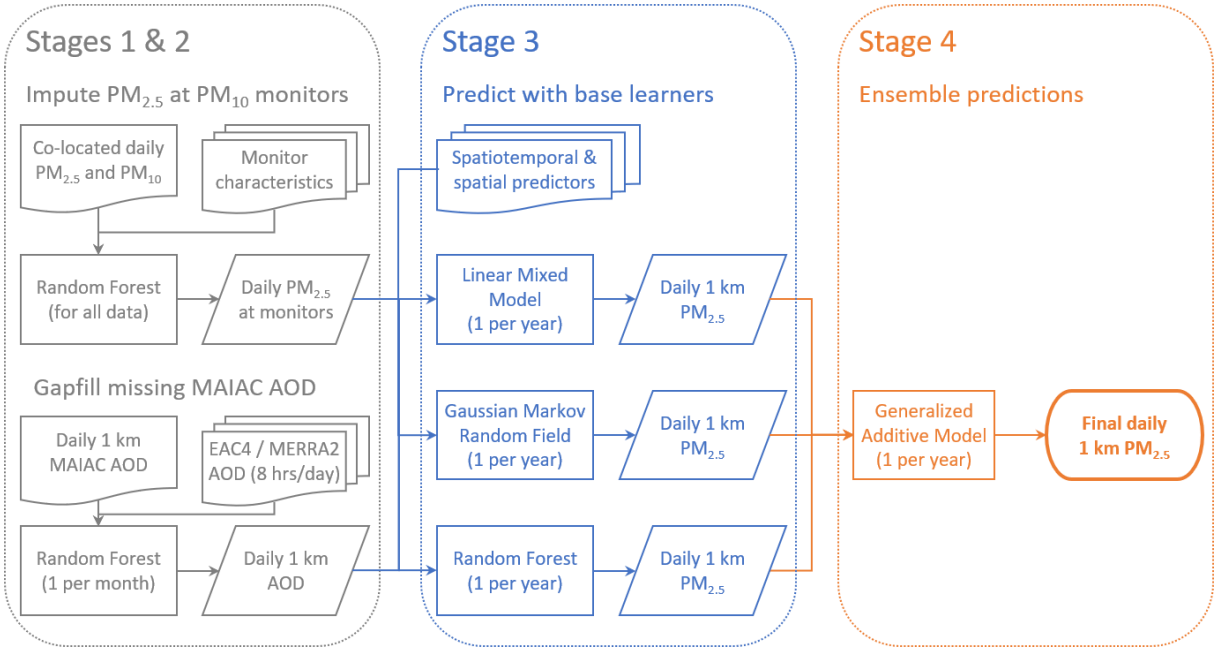

Figure 2-3. Flowchart of four-stage process to predict daily 1 km $PM_{2.5}$.

2.3.1 Stage 1: imputing $PM_{2.5}$ at PM_{10} monitors

Most monitors in France measured PM_{10} but not $PM_{2.5}$. To mitigate the sparsity of $PM_{2.5}$ monitors, we applied a method proposed by Stafoggia et al. (2019). We used all co-located daily measures of $PM_{2.5}$ and PM_{10} (n = 474 761) to tune and train a RF of 500 trees to predict $PM_{2.5}$ based on measured PM_{10} and monitor characteristics:

$$PM_{2.5_{mt}} = f\begin{pmatrix} PM_{10_{mt}}, vol_{mt}, loc_{mt}, infl_{mt}, \\ lat_m, lon_m, wday_t, yday_t, date_t \end{pmatrix} + \varepsilon_{mt} \qquad \text{Eq. (2-1)}$$

where $PM_{2.5_{mt}}$ and $PM_{10_{mt}}$ are, respectively, the $PM_{2.5}$ and PM_{10} measured by monitor m (1, …, 205) on day t (1, …, 7245); vol_{mt} indicates whether on day t monitor m excluded, included, or included an estimate of the semi-volatile fraction of PM_{10}; loc_{mt} and $infl_{mt}$ are, respectively,

the locale (rural, suburban, or urban) and predominant influence (traffic, industrial, or background) of monitor *m* on day *t*; lat_m and lon_m are the latitude and longitude of monitor *m*; $wday_t$, $yday_t$, and $date_t$ are, respectively, the day of week (to capture trends related to commuting or business activity), day of year (to capture seasonal trends), and date (to capture long-term trends); and ε_{mt} is the error at monitor *m* on day *t*. To reduce bias in the variable importance estimates, we sampled 63.2% of observations without replacement for each tree and estimated importance by permutation (Strobl et al., 2007). We tuned mtry (the number of variables to consider at each split) to minimize mean absolute error via model-based optimization and estimated accuracy using 5-fold CV blocked by monitor (section 2.3.5). We used the RF to impute $PM_{2.5}$ for the 1.71×10^6 monitor-days where only PM_{10} was measured.

2.3.2 Stage 2: filling gaps in MAIAC AOD

Clouds and snow cover often prevented MAIAC AOD retrieval over part of the study area. To fill these gaps, we trained RFs to predict MAIAC AOD based on co-located modelled AOD from atmospheric reanalysis. For computational reasons, we used 96 trees per forest, tuned using one spatiotemporally blocked 50% subsample of the data (section 2.3.5), and trained one RF for each month in the study period (mean observations per month $\approx 4.36\times10^6$):

$$AOD_{st}^{M} = f\left(R_{1_{st}}, \ldots, R_{8_{st}}, x_s, y_s, wday_t, yday_t\right) + \varepsilon_{st} \qquad \text{Eq. (2-2)}$$

where for each month *M* (1, …, 238), AOD_{st}^{M} is the MAIAC AOD observed at 1 km grid cell *s* (1, …, 632571) on day *t* (1, …, number of days in month *M*); $R_{h_{st}}$ is the AOD from atmospheric reanalysis (MERRA2 before 1 January 2003; EAC4 otherwise) at cell *s* on day *t* at each of eight times (8:30 UTC, 9:30 UTC, …, 15:30 UTC for MERRA2; 0 UTC, 3 UTC, …, 21 UTC for EAC4); x_s and y_s are the spatial coordinates of cell *s*; $wday_t$ and $yday_t$ are, respectively, the day of week and day of year; and ε_{st} is the error at cell *s* on day *t*. We estimated accuracy using 5-fold CV with spatiotemporal blocking (see section 2.3.5) and used the RFs to predict AOD for the 3.54×10^9 1 km grid cell-days without MAIAC AOD.

2.3.3 Stage 3: predicting daily 1 km PM using three base learners

In stage 3, we trained LMMs, GMRFs, and RFs to predict $PM_{2.5}$ (from stage 1) and PM_{10} based on gap-filled MAIAC AOD (from stage 2), 11 spatiotemporal predictors (sections 2.2.4 and 2.2.5), and 19 spatial predictors (section 2.2.6). We scaled the predictors to have similar range, and for the LMMs and GMRFs we log-transformed PM to approximate normality and prevent

negative predictions. We trained each base learner for each of $PM_{2.5}$ and PM_{10} in each year, yielding 120 base models (mean observations per year = 111 610; range 54 353 to 126 544). We estimated the accuracy of each base model using multi-stage CV blocked by monitor (section 2.3.5) and used the base models to predict PM for the 4.58×10^9 1 km grid cell-days of the study domain.

2.3.3.1 Linear mixed models

For each PM size fraction in each year, we calibrated a LMM with a random effect that allowed the PM-AOD relationship to vary daily for each of 8 climatic regions:

$$\log(PM_{st}^{FY}) = (\alpha^{FY} + \mu_{tr}^{FY}) + (\beta_{AOD}^{FY} + \nu_{tr}^{FY})AOD_{st} + \sum_{p=1}^{32}\left(\beta_p^{FY} X_{p_{st}}\right) + \varepsilon_{st}^{FY} \qquad \text{Eq. (2-3)}$$

where for each PM size fraction F (2.5 or 10) and year Y (2000, …, 2019), $\log(PM_{st}^{FY})$ is the log-transformed PM concentration at 1 km grid cell s (1, …, 632571) on day t (1, …, number of days in year Y); α^{FY} is the fixed intercept and μ_{tr}^{FY} is the random intercept on day t for the climatic region r that contains cell s; β_{AOD}^{FY} is the fixed slope of AOD and ν_{tr}^{FY} is the random slope of AOD on day t for the climatic region r that contains cell s; AOD_{st} is the AOD at cell s on day t. β_p^{FY} is the coefficient and $X_{p_{st}}$ the value at cell s on day t for each of the 11 spatiotemporal predictors, 19 spatial predictors, and sine and cosine transforms of the day of week; and ε_{st}^{FY} is the error at cell s on day t.

2.3.3.2 Gaussian Markov random fields

For each PM size fraction in each year, we calibrated a GMRF with a spatiotemporal random effect that varied smoothly over space on each day:

$$\log(PM_{st}^{FY}) = \alpha^{FY} + \beta_{AOD}^{FY} AOD_{st} + \sum_{p=1}^{32}\left(\beta_p^{FY} X_{p_{st}}\right) + \omega_{st}^{FY} + \varepsilon_{st}^{FY} \qquad \text{Eq. (2-4)}$$

where F, Y, s, t, $\log(PM_{st}^{FY})$, α^{FY}, β_{AOD}^{FY}, AOD_{st}, β_p^{FY}, $X_{p_{st}}$, and ε_{st}^{FY} are as in Eq. (2-3), and ω_{st}^{FY} is the spatiotemporal random effect at cell s on day t. We assumed that the error was independent and identically distributed (i.i.d.) following $\mathcal{N}(0, \sigma_\varepsilon^2)$ and the spatiotemporal random effect was temporally i.i.d. with Matérn spatial covariance:

$$Cov(\omega_{st}, \omega_{s't'}) = \begin{cases} 0, & t \neq t' \\ \sigma_\omega^2 \mathcal{C}(d_{ss'}; \rho_\omega), & t = t' \end{cases} \quad \text{Eq. (2-5)}$$

where σ_ω^2 is the variance of the spatiotemporal random effect, $\mathcal{C}$ is the Matérn function, $d_{ss'}$ is the Euclidean distance between locations *s* and *s'*, and ρ_ω is a hyperparameter that governs the range (distance at which the correlation falls to less than about 10%). We assigned penalized complexity priors to σ_ε^2, σ_ω^2, and ρ_ω that shrank the spatiotemporal random effect towards the null (Fuglstad et al., 2019; Simpson et al., 2017) and fit the model using INLA.

2.3.3.3 Random forests

For each PM size fraction in each year, we trained a RF with the equation:

$$PM_{st}^{FY} = f\left(AOD_{st}, X_{1_{st}}, \ldots, X_{30_{st}}, x_s, y_s, wday_t, yday_t\right) + \varepsilon_{st}^{FY} \quad \text{Eq. (2-6)}$$

where *F*, *Y*, *s*, *t*, PM_{st}^{FY}, AOD_{st}, and ε_{st}^{FY} are as in Eq. (2-3); $X_{1_{st}}, \ldots, X_{30_{st}}$ are, respectively, the value for each of the 11 spatiotemporal predictors and 19 spatial predictors at cell *s* on day *t*; x_s and y_s are the spatial coordinates of cell *s*; and $wday_t$ and $yday_t$ are the day of week and day of year. We used 250 trees and fixed mtry at 5 because exploratory tuning suggested that mtry > 5 provided little benefit and risked overfitting.

2.3.4 Stage 4: ensembling predictions to improve accuracy

In stage 4, we calibrated a GAM to ensemble the predictions of the stage 3 base learners. We used predictions for held-out monitors to calibrate the GAMs because these reflect accuracy at unmonitored locations (section 2.3.5). We fit a GAM for each PM size fraction in each year (20 GAMs total; mean observations per year = 111 610; range 54 353 to 126 544) using tensor product smooths that allowed the coefficient for each base learner's predictions to vary smoothly over space and time:

$$PM_{st}^{FY} = te(x_s, y_s, t)LMM_{st}^{FY} + te(x_s, y_s, t)GMRF_{st}^{FY} + te(x_s, y_s, t)RF_{st}^{FY} + \varepsilon_{st}^{FY} \quad \text{Eq. (2-7)}$$

where for each PM size fraction *F* (2.5 or 10) and year *Y* (2000, …, 2019), PM_{st}^{FY} is the PM concentration at 1 km grid cell *s* (1, …, 632571) on day *t* (1, …, number of days in year *Y*); $te(x_s, y_s, t)$ is the tensor product of penalized cubic regression splines of the spatial coordinates of cell *s* (x_s and y_s) and the temporal index *t*; LMM_{st}^{FY}, $GMRF_{st}^{FY}$, *and* RF_{st}^{FY} are, respectively, the CV prediction at cell *s* on day *t* from a LMM, GMRF, and RF that were trained while holding out all data from the fold that contains cell *s*; and ε_{st}^{FY} is the error at cell *s* on day *t*. We

estimated accuracy using multi-stage CV blocked by monitor (section 2.3.5) and used the GAMs to predict AOD for the 4.58×10^9 1 km grid cell-days of the study domain.

2.3.5 Cross-validation

To limit bias due to spatiotemporal autocorrelation and avoid information leakage between the base learners and the ensemble, we extended the blocked CV scheme described by Shtein et al. (2019). In stage 1 (imputing $PM_{2.5}$ at PM_{10} monitors), we estimated accuracy at monitors that never measured $PM_{2.5}$ using 5-fold CV blocked by monitor: we randomly assigned each monitor that measured both $PM_{2.5}$ and PM_{10} to one of 5 folds. This ensured that no observations from test monitors were in the training set.

In stage 2 (filling gaps in MAIAC AOD data), we estimated accuracy at locations far from same-day MAIAC AOD observations (because MAIAC AOD tends to be missing in spatial clumps) using 5-fold CV with spatiotemporal blocking: we split the study area into 50 regions and randomly assigned MAIAC AOD in each day-region to one of 5 folds (Figure 2-4). This ensured that no same-day observations from test regions were in the training set.

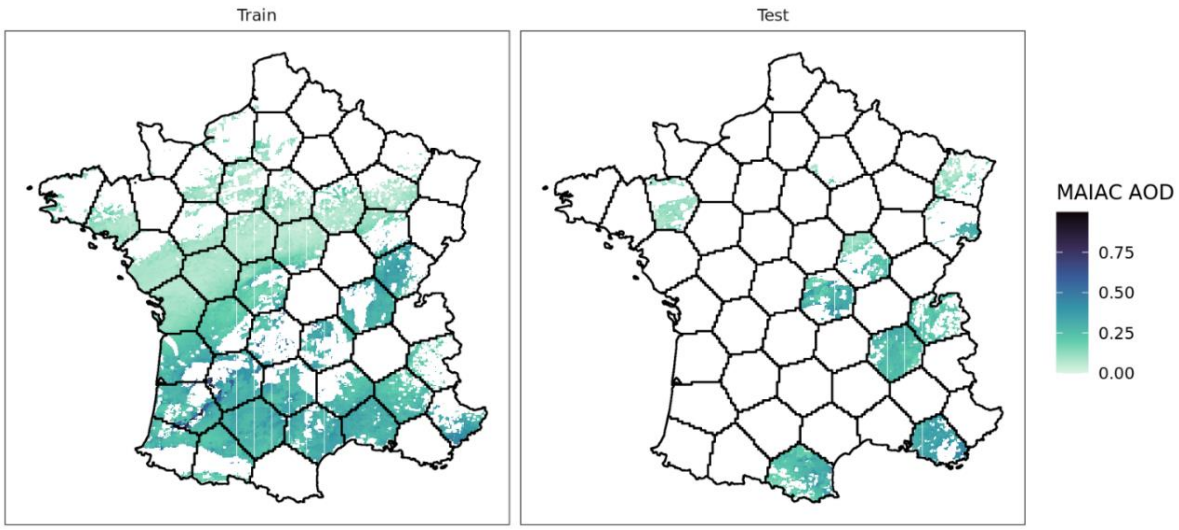

Figure 2-4. Example of spatiotemporal blocking used to cross-validate stage 2 (filling gaps in MAIAC AOD). Left: one day of MAIAC AOD used to train a stage 2 random forest; right: MAIAC AOD on the same day held out to estimate accuracy. Black lines delineate the 50 spatial regions used for blocking. Areas that are white in both images are gaps due to cloud cover that stage 2 will fill. Note that the gaps are large, meaning that most locations where MAIAC AOD is missing are not adjacent to a location where it is available.

In stages 3 and 4 (predicting daily 1 km PM with three base learners and ensembling the predictions), we estimated accuracy at unmonitored locations using the multi-stage CV scheme blocked by monitor illustrated in Figure 2-5. This scheme ensured that 1) observations from test monitors were never used for training; 2) ensembles were regularized (trained on base learner predictions for held-out monitors); and 3) monitors held out to test an ensemble were never used to train that ensemble's base learners.

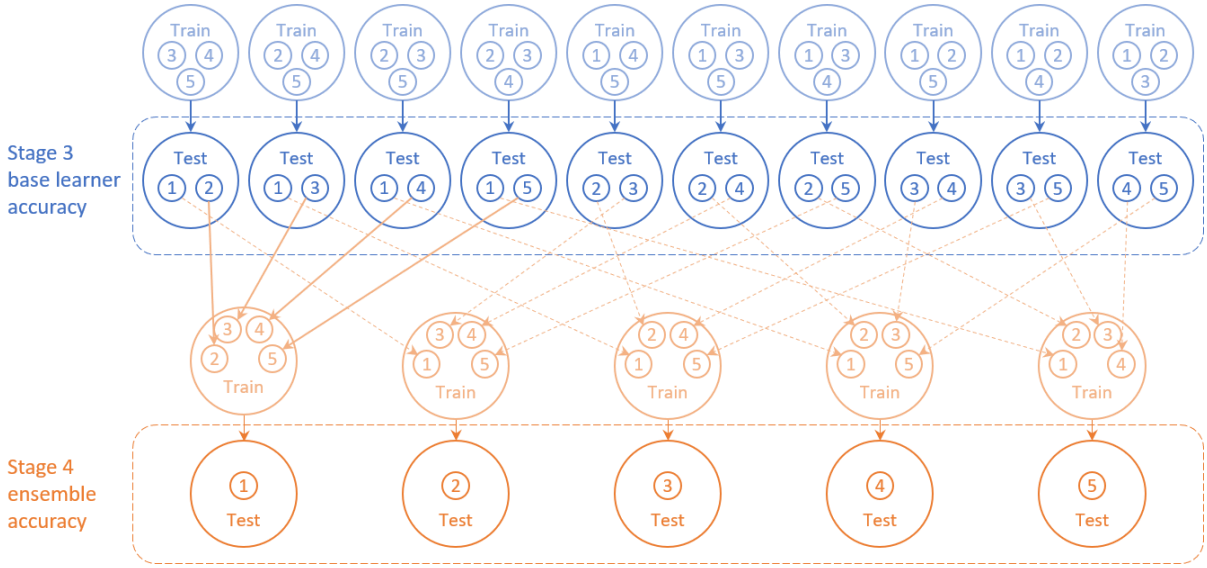

Figure 2-5. Cross-validation scheme for stage 3 (predicting daily 1 km PM with three base learners) and stage 4 (ensembling the base learner predictions to improve accuracy). First, we randomly assigned each PM monitor to one of five folds (small circles labelled 1-5). For each unique combination of three folds (top row; light blue), we trained the three base learners and predicted for the two held-out test folds (second row; dark blue). We used these predictions to evaluate the accuracy of the base learners. Then, for each fold, we trained an ensemble using the test predictions from all base learners that held out the fold (third row; light orange; solid orange arrows show iteration where fold 1 is the test fold). We predicted for the test fold (bottom row; dark orange) and used these predictions to evaluate the accuracy of the ensemble.

2.3.6 Performance metrics

We evaluated the models using mean absolute error (MAE), which reflects the typical difference between a model's predictions and measured PM, R^2, which reflects the fraction of spatiotemporal variation in PM captured by a model, and root mean squared error (RMSE), which can be compared with the standard deviation (SD) of measured PM to see by how much a model improves upon a naïve prediction of the mean. We also split each of these metrics into a spatial and temporal component as described by Kloog et al. (2011).

2.4 Results

2.4.1 Stage 1: imputing $PM_{2.5}$ at PM_{10} monitors

Mean $PM_{2.5}$ was 13.7 µg/m^3 (standard deviation [SD] 9.9 µg/m^3) and mean PM_{10} was 21.6 µg/m^3 (SD 12.5 µg/m^3) across all monitors in continental France for 2000 to 2019 (Figure A2-1). PM concentrations declined over the study period and were generally highest in winter and lowest in summer (Figure A2-2). The cross-validated predictions of the stage 1 RF showed good correspondence with observed $PM_{2.5}$ at monitors that were not used to train the RF (R^2 = 0.87, MAE = 2.48 µg/m^3) with little bias (mean error = -0.157) but a tendency to underestimate very high concentrations (Figure 2-6). Performance was good even in early years when there were few $PM_{2.5}$ monitors (R^2 >= 0.82 in every year except 2008; MAE < 2.5 µg/m^3 in most years) (Table A2-1). The drop in performance in 2007 and 2008 coincided with a change in monitor technology that increased measured PM_{10} concentrations and likely complicated the relationship between PM_{10} and $PM_{2.5}$. PM_{10} concentration was by far the most important predictor of $PM_{2.5}$ concentration (Figure A2-3).

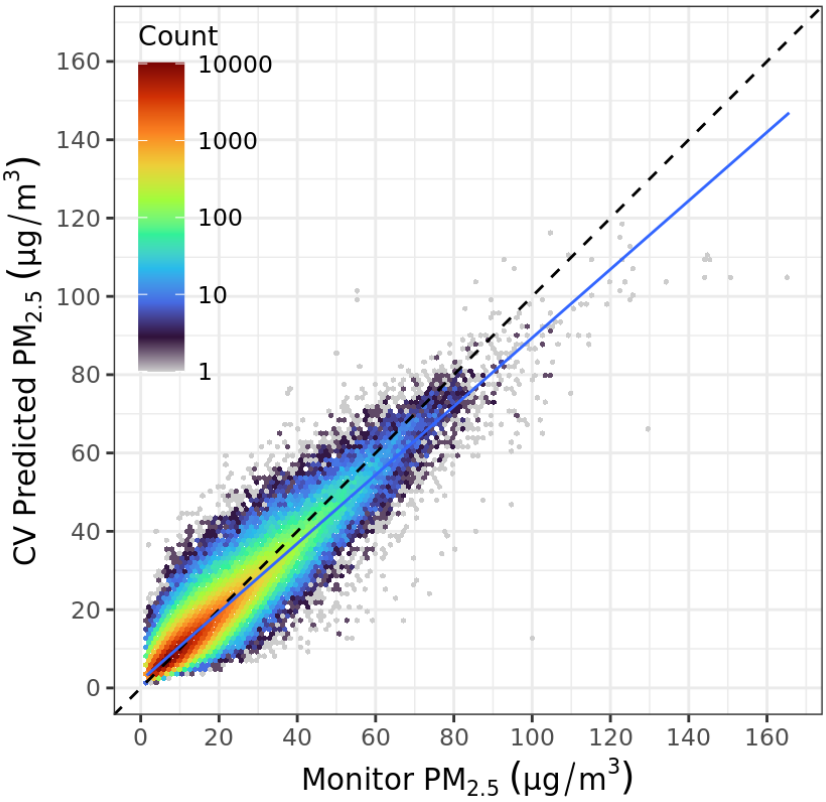

Figure 2-6. Cross-validation (CV) predicted vs. observed daily PM2.5 concentrations from the stage 1 random forest. Dashed black line shows 1:1 relationship; solid blue line shows actual relationship (R^2 = 0.873; MAE = 2.48 µg/m3; mean error = -0.157).

2.4.2 Stage 2: filling gaps in MAIAC AOD

Mean MAIAC AOD over continental France for 2000 to 2019 was 0.126 (SD 0.084); MAIAC AOD was missing for 77% of the 1 km cell-days in the study domain (Figure A2-4), similar to other areas (Di et al., 2019; Schneider et al., 2020; Stafoggia et al., 2019). Cross-validated R^2 for the stage 2 RFs typically ranged from about 0.55 in winter to about 0.78 in summer (Figure 2-7), coinciding with fewer MAIAC observations in winter and more in summer. MAE typically ranged from about 0.025 in fall to about 0.034 in summer, coinciding with lower AOD in fall and higher AOD in summer. Performance was similar between periods that used modelled AOD from MERRA2 vs. EAC4 as predictors. There was a slight tendency to overestimate high AOD (slope = 0.94), but average performance was good (mean R^2 = 0.70; mean MAE = 0.030) (Table A2-2). Prior to 2003, the most important predictors of MAIAC AOD were modelled AOD at 12 UTC, the day of year, and the spatial y coordinate (Figure A2-5). From 2003 on, modelled AOD at 15 UTC was the second most important predictor. This likely related to the mid-2002 launch of the Aqua satellite, which passes over continental France around 13 UTC; previously, MAIAC AOD was only available around 11 UTC from the Terra satellite. Figure A2-6 shows an example of gapfilled AOD.

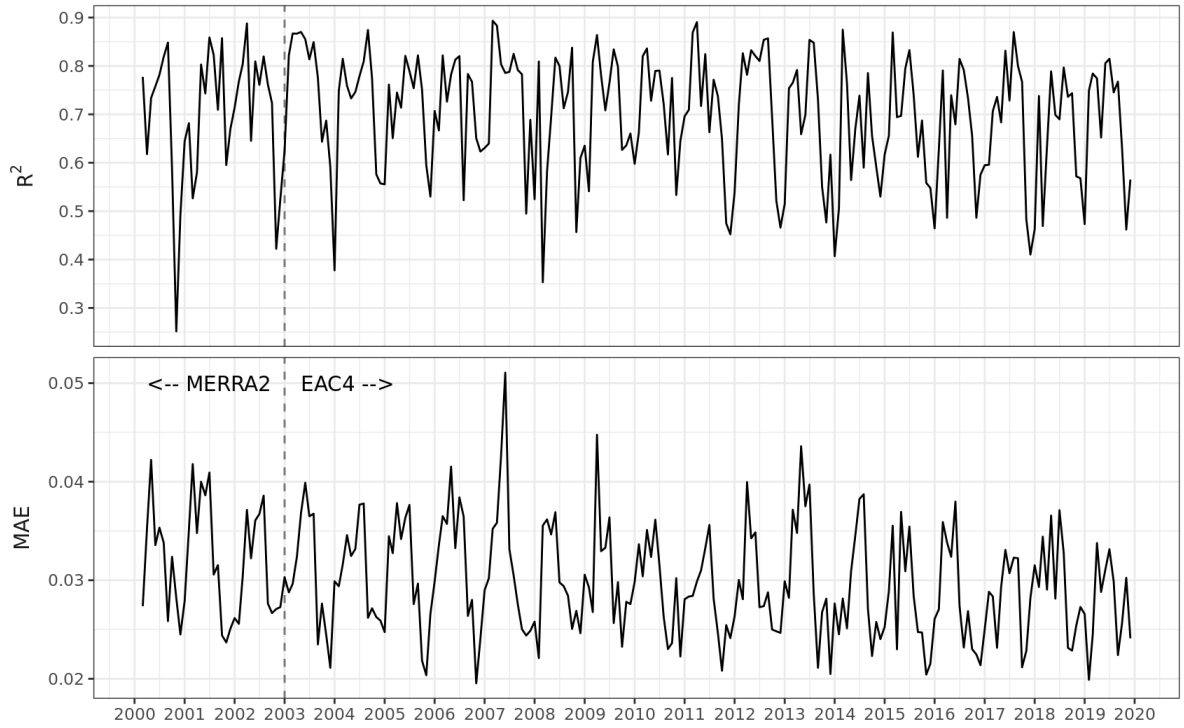

Figure 2-7. Monthly cross-validated R^2 (top) and MAE (bottom) of the stage 2 random forests (filling gaps in MAIAC AOD). Vertical dashed line separates periods when modelled AOD from MERRA2 vs. EAC4 was used to predict MAIAC AOD.

2.4.3 Stages 3 and 4: predicting daily 1 km PM_{10} and $PM_{2.5}$ with three base learners and ensembling the predictions to increase accuracy

Table 2-2 shows the average cross-validated performance of the stage 3 base learners and stage 4 GAM ensemble. GMRF was the most accurate base learner (mean $PM_{2.5}$ R^2 = 0.75, MAE = 2.72 µg/m^3; mean PM_{10} R^2 = 0.70, MAE = 4.26 µg/m^3), followed by RF, with LMM the least accurate. The stage 4 GAM ensembles slightly improved performance (mean $PM_{2.5}$ R^2 = 0.76, MAE = 2.72 µg/m^3; mean PM_{10} R^2 = 0.71, MAE = 4.26 µg/m^3), almost eliminating the GMRFs' slight bias and increasing spatial R^2 compared to both the GMRFs and RFs. The relative importance of the base learners in the GAM ensemble varied over space and time (Figure A2-7), but GMRF predictions usually had the highest weight, consistent with their high cross-validated accuracy.

Table 2-2. Cross-validated (CV) performance (mean 2000-2019) of the stage 3 base learners (LMM, GMRF, RF) and stage 4 ensemble (GAM) predicting daily 1 km PM (µg/m3).

				Multi-stage CV Performance								
	Observed PM		Model	Total					Spatial		Temporal	
	Mean	SD[a]		RMSE	Bias[b]	Slope[c]	R^2	MAE	R^2	MAE	R^2	MAE
$PM_{2.5}$	13.8	8.5	LMM	5.03	0.67	0.63	0.63	3.35	0.38	1.82	0.68	2.92
			GMRF	4.09	0.46	0.75	0.75	2.72	0.45	1.68	0.81	2.24
			RF	4.52	-0.14	0.63	0.70	3.18	0.47	1.69	0.74	2.68
			GAM	4.02	-0.01	0.76	0.76	2.72	0.49	1.63	0.81	2.23
PM_{10}	21.5	11.9	LMM	7.65	1.04	0.60	0.58	5.21	0.32	2.98	0.64	4.39
			GMRF	6.40	0.73	0.72	0.70	4.26	0.39	2.79	0.78	3.36
			RF	7.07	-0.19	0.57	0.64	5.00	0.41	2.80	0.70	4.11
			GAM	6.28	-0.02	0.72	0.71	4.26	0.43	2.71	0.78	3.34

a Standard deviation; b Mean error; [c] Slope of regression of CV predicted PM on observed PM

R^2 for all models increased in early years with the number of monitors and remained high from 2009 to 2019; MAE covaried with mean observed PM (Figure 2-8). The sharp increase in PM_{10} MAE in 2007 coincided with a change in monitor technology: in 2007, all PM_{10} monitors were modified to measure semi-volatile particles in addition to non-volatile particles, increasing observed PM_{10} concentrations. $PM_{2.5}$ monitors were modified in 2008 and 2009, corresponding to the increase in $PM_{2.5}$ MAE in 2008 and 2009. R^2 was highest in winter and spring and lowest in summer; MAE was highest in winter and lowest in summer, corresponding to typical seasonal trends in PM concentration (Figure A2-8). For the random forest model, the most important predictors were day of year, followed by some meteorological variables (boundary layer height, temperature, wind, and precipitation) and AOD (Figure A2-10).

The base learners and GAM ensemble captured day-to-day variation in PM concentration better than between-location differences in annual mean PM concentration (GAM ensemble spatial $R^2 \approx 0.46$, temporal $R^2 \approx 0.80$). This is in part because PM concentration varies more over time than space; spatial MAE was lower than temporal MAE ($PM_{2.5}$ GAM ensemble spatial MAE = 1.6, temporal MAE = 2.2), indicating that predicted annual mean PM concentrations were quite accurate. It may also reflect difficulty capturing spatial variation in urban areas. The lowest spatial R^2 and highest spatial MAE were in Île-de-France, the densely populated region that contains Paris, which also had the highest and most variable PM concentrations (Table A2-3, Figure A2-9). There may not have been enough monitors for the model to capture complex spatial variation in PM concentration over greater Paris.

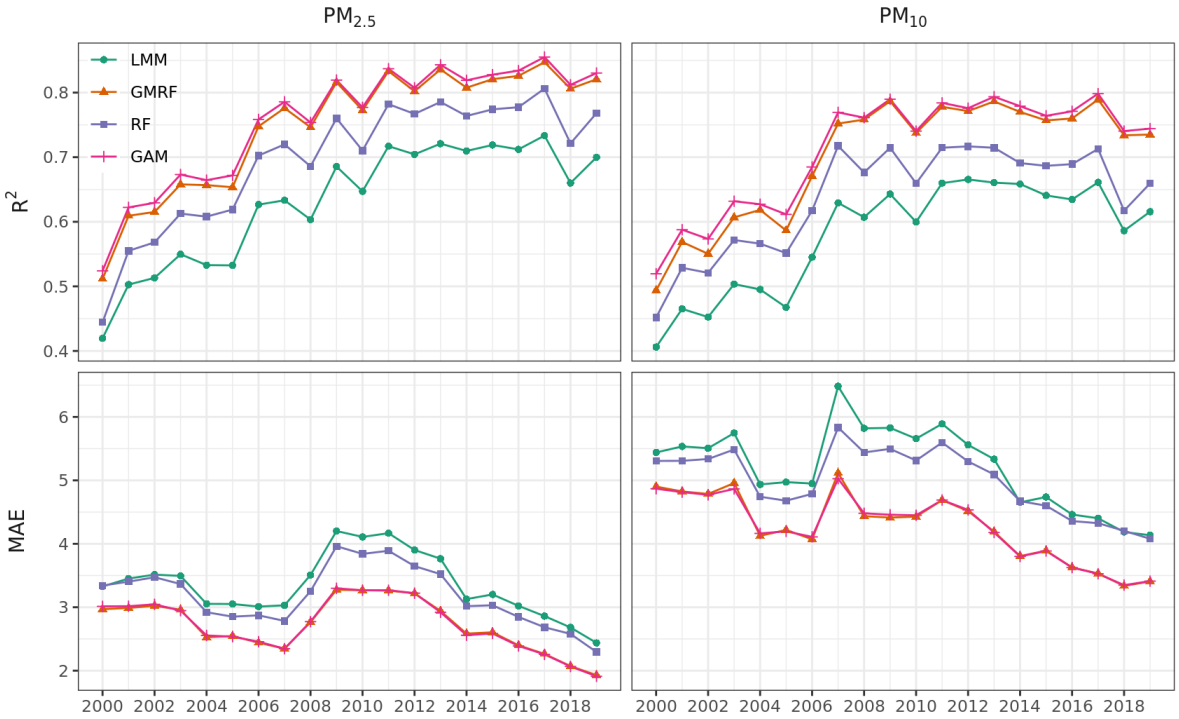

Figure 2-8. Annual cross-validated R^2 (top) and MAE (bottom; $\mu g/m^3$) of the stage 3 base learners (LMM, GMRF, RF) and stage 4 ensemble (GAM) predicting daily 1 km $PM_{2.5}$ (left) and PM_{10} (right).

Since the majority of our $PM_{2.5}$ data consisted of imputed $PM_{2.5}$ concentration at PM_{10} monitors from the stage 1 RF, we performed a sensitivity analysis comparing the cross-validated predictions of the GAM ensemble to only observed $PM_{2.5}$ concentration at $PM_{2.5}$ monitors. Apart from 2000 (when there were only 5 $PM_{2.5}$ monitors), performance was similar at $PM_{2.5}$ monitors (mean $R^2 = 0.77$, mean MAE = 2.98) and across all monitors (mean $R^2 = 0.77$, mean MAE = 2.71). We also constructed an alternate model by retraining the base learners and

ensemble using only $PM_{2.5}$ monitors for all years except 2000. This alternate model was less accurate (mean R^2 = 0.66, mean MAE = 3.62) than the stage 4 GAM ensemble evaluated only at $PM_{2.5}$ monitors (Figure 2-9), indicating that increasing the quantity of training data by imputing $PM_{2.5}$ at PM_{10} monitors resulted in more accurate final $PM_{2.5}$ predictions than if we had relied solely on $PM_{2.5}$ monitors.

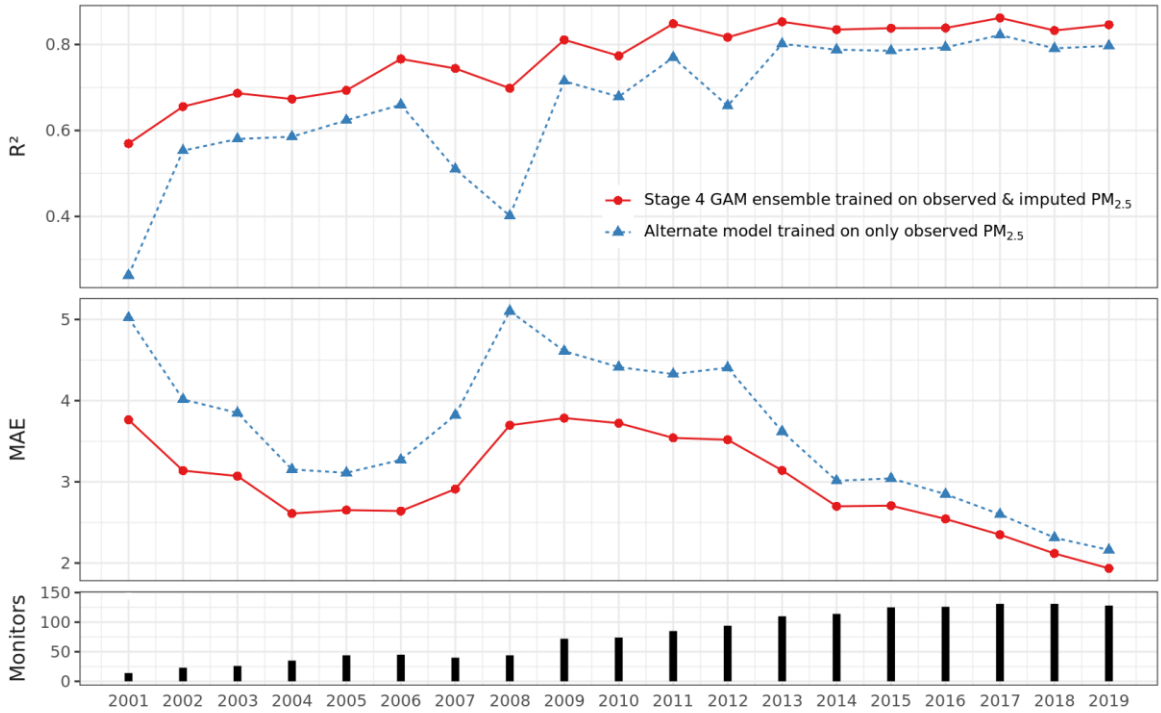

Figure 2-9. Annual cross-validated performance at $PM_{2.5}$ monitors of the stage 4 GAM ensemble (red circles) and an alternate model (blue triangles) trained on only observed $PM_{2.5}$. Top: R^2; middle: MAE (μg/m^3); bottom: number of $PM_{2.5}$ monitors.

Figure 2-10 shows the mean 2000 to 2019 $PM_{2.5}$ and PM_{10} concentration predicted by each base learner and GAM ensemble and Figure 2-11 shows varying spatial patterns of PM predicted by the GAM ensemble over greater Paris on three example days. The LMM and GMRF predictions are similar; the RF predictions are slightly higher in rural areas. The GAM ensemble predictions resemble those of the GMRF with some contribution from the RF in the southeast and southwest. PM concentrations are high in the north with a hotspot over greater Paris. In the southeast, high concentrations extend south down the Rhône river valley from the hotspot of greater Lyon, east into alpine valleys, and along the Mediterranean coast. The lowest concentrations are over the sparsely populated south centre, the Pyrenees in the southwest, and the Bretagne and Cotentin peninsulas in the northwest. $PM_{2.5}$ concentrations show less contrast between urban and rural areas than PM_{10}. PM concentrations over Paris show varying spatial patterns but are generally highest over built-up areas and roads.

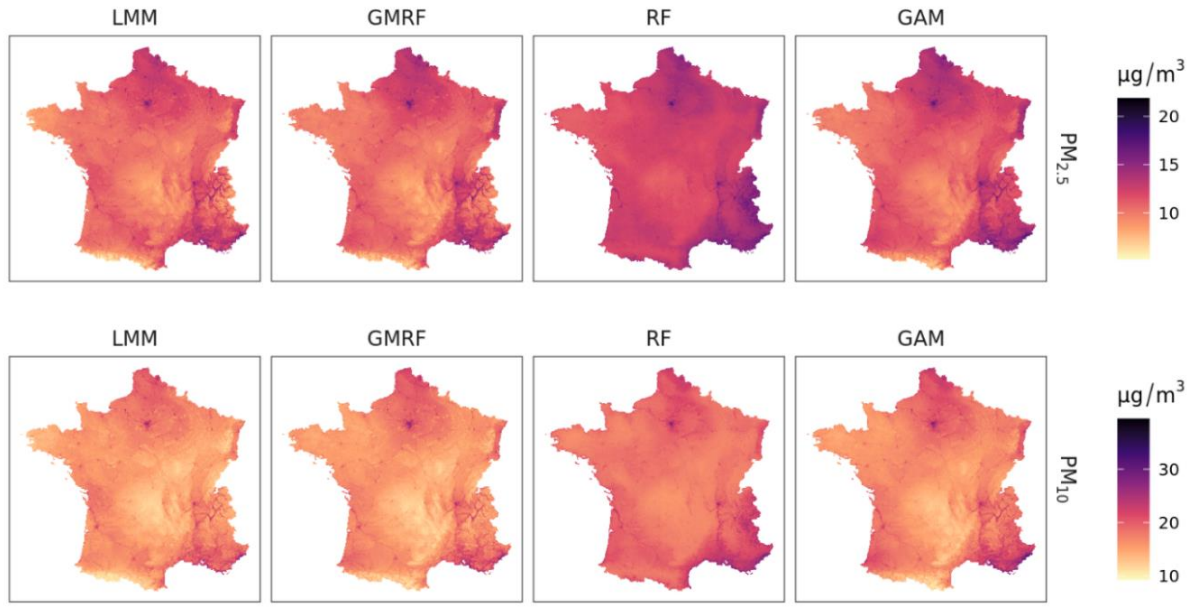

Figure 2-10. Mean $PM_{2.5}$ (top) and PM_{10} (bottom) concentration predicted by the stage 3 base learners (LMM, GMRF, RF) and stage 4 ensembles (GAM) for 2000 to 2019.

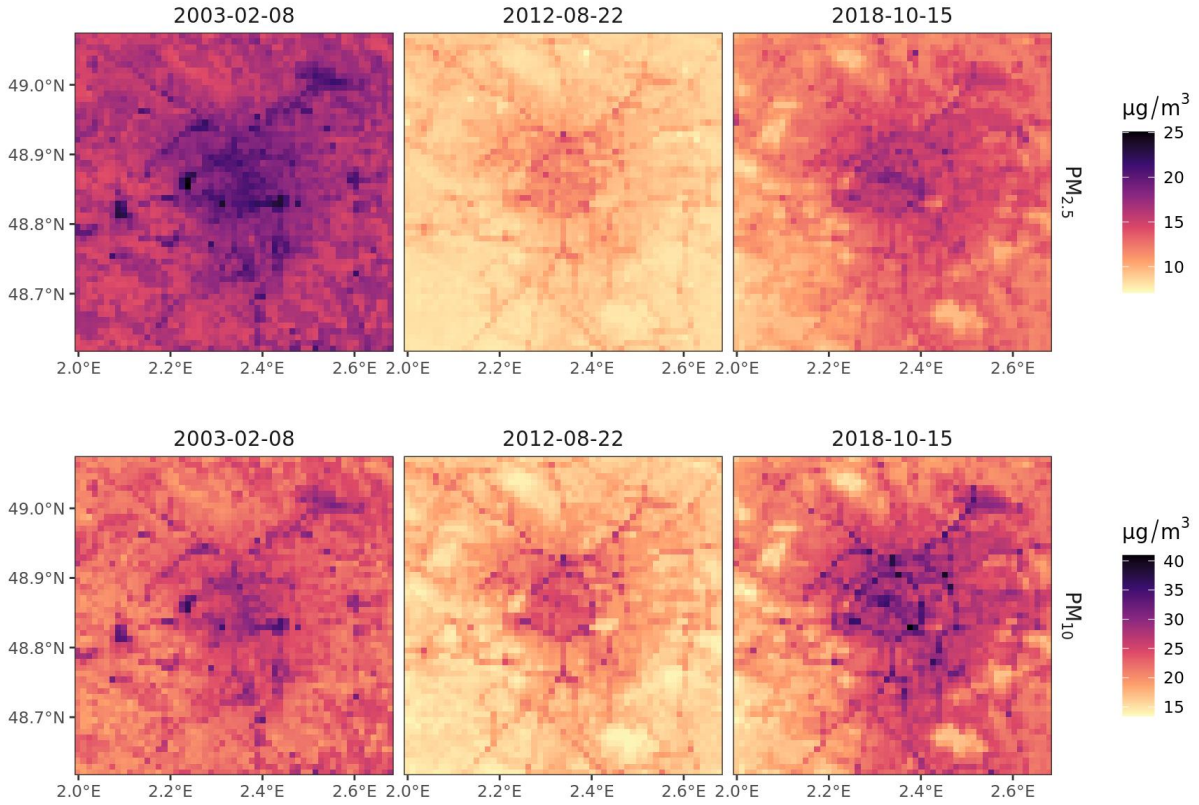

Figure 2-11. Mean 24-hour $PM_{2.5}$ (top) and PM_{10} (bottom) concentration over greater Paris predicted by the stage 4 GAM ensemble on three example days.

2.5 Discussion

Our finding that GMRFs predicted daily 1 km PM concentration more accurately than LMMs is consistent with results in the northeastern United States (Sarafian et al., 2019). Our finding that GMRFs were also more accurate than RFs is novel and of note, as RFs and other tree-based machine learning algorithms performed well in several recent studies (Di et al., 2019; Just et al., 2020; Schneider et al., 2020; Stafoggia et al., 2020, 2019). We emphasize the importance of careful performance evaluation when using flexible machine learning algorithms: GMRFs had the best cross-validated performance (corresponding to accuracy at unmonitored locations), but RFs performed better than GMRF on non-independent training data (corresponding to accuracy at monitors). Evaluation methods that do not ensure independence between training and testing data risk mistaking good performance at monitors for good performance everywhere.

Our GAM ensemble captured temporal variation in PM concentration better than spatial variation. We attempted to improve spatial performance in urban areas by downscaling the residuals of the GAM ensemble using RFs trained on the high spatial resolution predictors listed in Table 2-1, as was done in a few previous studies (Di et al., 2016; Kloog et al., 2014a; Stafoggia et al., 2019). Unlike previous studies, we used 5-fold cross-validation blocked by monitor to assess whether the downscaling improved accuracy: the downscaled predictions were less accurate than the 1 km GAM ensemble predictions (higher MAE, lower R^2). Downscaling over cities is an area for further research, as epidemiological studies would benefit from better estimates of differences in PM exposure within a city.

Despite good overall performance, our approach has some limitations. First, the sparsity of the monitoring network limited performance in early years. Even in later years, most monitors were clustered in cities, making it difficult to evaluate accuracy in smaller towns and rural areas and risking overreliance on predictors that work well in urban areas but may not work well elsewhere. The clustering of monitors near cities means our model is roughly weighted by population density, which may or may not be appropriate depending on the intended use for the predictions (Sarafian et al., 2020). New low-cost PM sensors might complement the existing monitoring network, particularly since our model's weaker spatial performance suggests that a few PM observations at new locations might be more useful than a long timeseries of measurements at a single location.

Second, MAIAC AOD was the only predictor with both a high spatial (1 km) and temporal (daily) resolution, but since it is based on a few daytime observations of the entire atmospheric column, it is both vertically and temporally misaligned with surface-level daily mean PM concentration. We included planetary boundary layer height to help distinguish between surface-level vs. high-altitude aerosols, but it had a much coarser spatial resolution than MAIAC AOD. Our model might have benefitted from AOD at coarser spatial but higher temporal resolution, such as from geostationary weather satellites, or from considering longer periods and giving greater weight to rare observations when filling gaps in MAIAC AOD.

Despite these limitations, our multi-stage ensemble approach was able to predict daily 1 km $PM_{2.5}$ and PM_{10} with low error across a large area over 20 years. To our knowledge, this is the first work conducted in France with such a high spatiotemporal resolution (1 km-daily), large spatial extent (national) and long temporal coverage (2000 to 2019). We increased accuracy by supplementing sparse $PM_{2.5}$ observations with imputed data and by ensembling the predictions of three base learners. We confirmed that Gaussian Markov random fields predict daily PM concentration better than linear mixed models and provide the first evidence that they may also outperform random forests. Our dataset of daily 1 km $PM_{2.5}$ and PM_{10} is available to health and ecosystems researchers in France and may inform policy makers on air quality issues.

Chapter 3

Ambient temperature and preterm birth risk in France

Ian Hough[a,b], Matthieu Roland[a], Ariane Guilbert[a], Emie Seyve[a], Barbara Heude[c], Rémy Slama[a], Sarah Lyon-Caen[a], Isabelle Pin[a,d], Cécile Chevrier[e], Itai Kloog[b], Johanna Lepeule[a]

[a] *Institute for Advanced Biosciences, Université Grenoble Alpes, INSERM, CNRS, La Tronche, France*

[b] *Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Be'er Sheva, Israel*

[c] *Centre for Research in Epidemiology and Statistics (CRESS), Université de Paris, INSERM, INRAE, Paris, France*

[d] *Department of Paediatric Pneumology, Grenoble Teaching Hospital, La Tronche, France*

[e] *Research Institute for Environmental and Occupational Health (IRSET), Université Rennes, INSERM, EHESP, Rennes, France*

Portions of this chapter are currently under review as:

Hough, I., Roland, M., Guilbert, A., Seyve, E., Heude, B., Slama, R., Lyon-Caen, S., Pin, I., Chevrier, C., Kloog, I., and Lepeule, J. (Submitted) Early delivery following in utero ambient temperature exposure in a temperate climate: a comprehensive survival approach.

3.1 Introduction

The final objective of this PhD was to study the association between temperature and preterm birth in France. The current evidence suggests that extreme temperatures may increase the risk of preterm birth, but results have been mixed and there is not yet a consensus on how the timing of exposure during pregnancy affects risk (Chersich et al., 2020; Ju et al., 2021). We aimed to comprehensively assess the association between ambient temperature and risk of preterm birth by addressing limitations of previous studies.

First, we used a survival modelling approach with Cox proportional hazards models to account for the increasing risk of birth as pregnancy progresses and possible seasonal trends in conception rates. We limited exposure assessment error by using daily residence-based exposure from the state-of-the-art spatiotemporal temperature model described in Chapter 1 and used distributed lags to account for the time-varying association between temperature and preterm birth risk. We examined exposure throughout pregnancy using narrow windows and adjusting for exposure during every window.

Second, we assessed multiple exposure indicators, as the temperature metric most relevant for health may vary depending on the location and outcome (Basu et al., 2008; Laaidi et al., 2012; Murage et al., 2017). We examined mean temperature (an indicator of overall exposure), maximum temperature (an indicator of daytime exposure), minimum temperature (an indicator of night-time exposure), temperature variability, which recent studies have linked to adverse birth outcomes (Jakpor et al., 2020; Li et al., 2018; Molina & Saldarriaga, 2017), and a heatwave indicator that accounts for acclimation to local and seasonal temperatures.

We found windows of susceptibility to cold (5^{th} vs 50^{th} percentile of mean temperature) during weeks 7-9 after conception and days 10-4 before delivery, and to night-time heat (95^{th} vs 50^{th} percentile of minimum temperature) during weeks 1-5 and 20-26 after conception (Figure 3-1). The relative risk of preterm birth ranged from 1.29 (95% CI 1.02 to 1.64) following cold (2°C vs 11.6°C) during weeks 7-9 after conception to 2.87 (95% CI 1.21 to 6.79) following night-time heat during weeks 20-26 after conception. Overall and daytime heat (high mean and maximum temperature) showed similar but weaker effects than night-time heat, and we found no clear associations with temperature variability or a heatwave indicator.

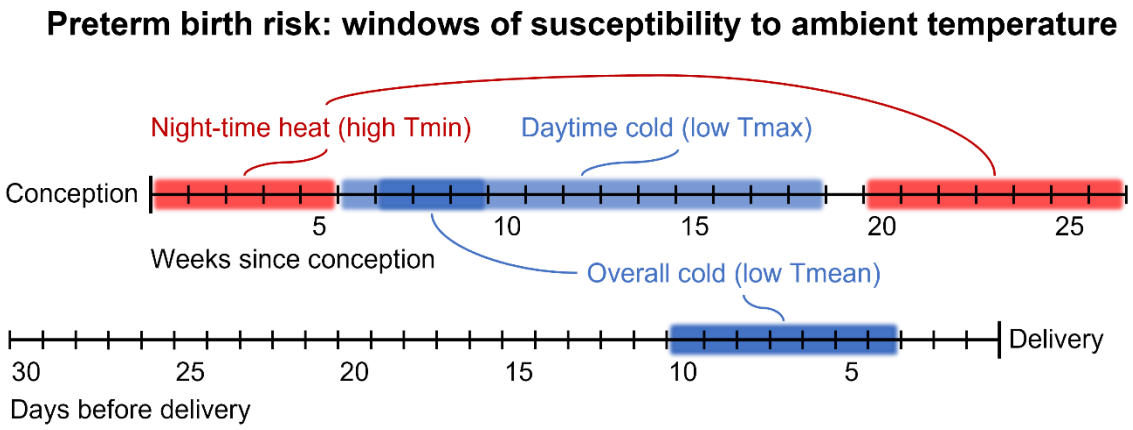

Figure 3-1. Windows of susceptibility to ambient temperature identified in this study.

Our robust findings suggest that, in a temperate climate, heat, particularly at night, and cold may be dangerous for pregnant women and their offspring. In the context of rising temperatures and more frequent weather hazards, these results should inform public health policies to reduce the growing burden of preterm birth.

3.2 Methods

3.2.1 Study population

We obtained data from three French prospective mother-child cohorts that were designed to study the effects of prenatal environmental exposures on child development and health: EDEN (Étude des Déterminants pré et post natals du développement et de la santé de l'Enfant), PELAGIE (Perturbateurs Endocriniens : étude Longitudinale sur les Anomalies de la Grossesse, l'Infertilité, et l'Enfance), and SEPAGES (Suivi de l'Exposition à la Pollution Atmosphérique durant la Grossesse et Effets sur la Santé). The cohorts' protocols are described in detail elsewhere (Heude et al., 2016; Lyon-Caen et al., 2019; Petit et al., 2010). Briefly, EDEN included 2002 women recruited between 2003 and 2006 at <24 weeks amenorrhea in the metropolitan areas of Poitiers and Nancy; PELAGIE included 3421 women recruited between 2002 and 2006 at <19 weeks amenorrhea in the Brittany region; SEPAGES included 484 women recruited between 2014 and 2017 at <19 weeks amenorrhea in the metropolitan area of Grenoble (Figure 3-2). Poitiers, Nancy, and Brittany have a temperate oceanic climate (Köppen classification Cfb) while Grenoble has a warm temperate climate (Köppen classification Cfa) with montane influences.

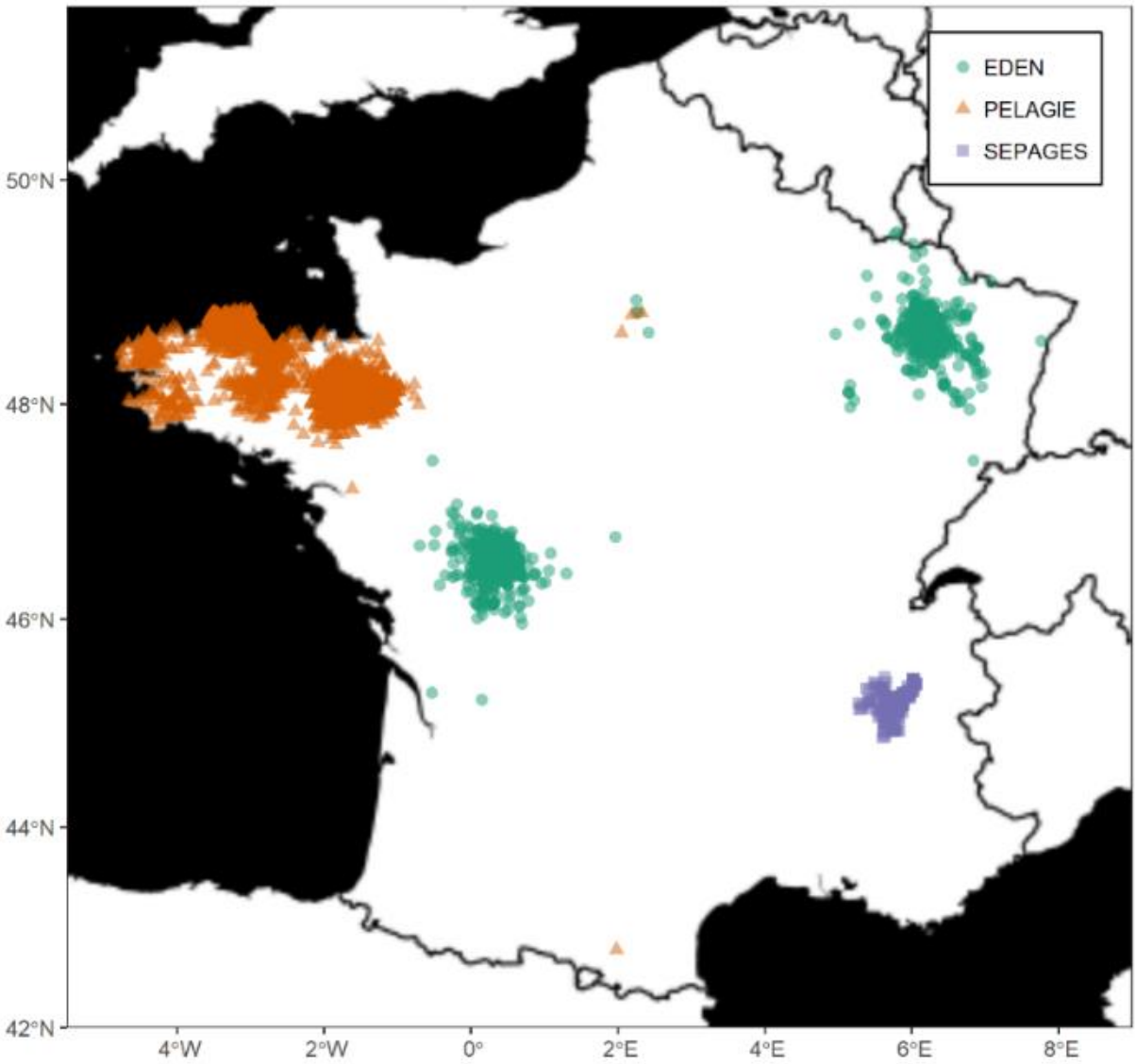

Figure 3-2. Locations of the cohort participants.

All three cohorts collected medical and sociodemographic information via clinical examinations and questionnaires during and after pregnancy. Home addresses (including any changes during pregnancy) were geocoded using the IGN Mon Géocodeur tool and Geocible. For 43% of PELAGIE participants, only the municipality or neighbourhood of residence at inclusion (on average 10.4 weeks after conception) was available; we assumed these women did not move during pregnancy. We excluded multiple gestation, non-livebirths, pre-existing diabetes or hypertension, and participants lost to follow-up before delivery. To ensure complete equal-length exposure histories, we further excluded participants missing covariates (described below) or missing exposure for more than one day in any of the 26 weeks following conception (among these were 5 extremely preterm births at <28 weeks amenorrhea). This left 5,347

mother-child pairs (Figure 3-3). All births were independent (no mother participated for more than one pregnancy).

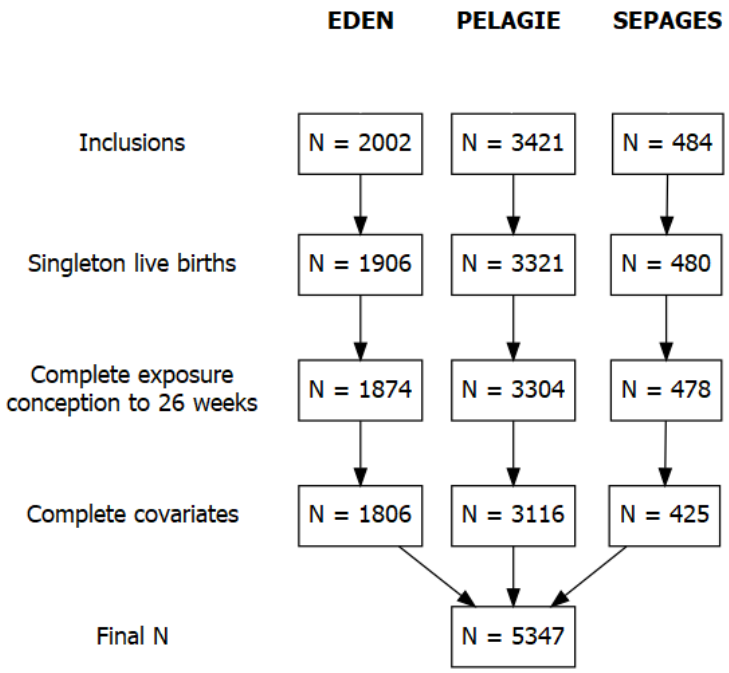

Figure 3-3. Flow chart of the study population.

3.2.2 Outcome definition

Duration of pregnancy (conception to birth) was assessed using both the reported date of the last menstrual period (LMP) and the estimate from the first trimester ultrasound (when LMP was not reported or when the two differed by more than 30 days). When neither of these were available (n = 3 in EDEN; n = 84 in PELAGIE), we used the obstetrician's estimate of gestational age at delivery. We defined preterm birth as delivery at <37 weeks amenorrhea (<35 weeks since conception).

3.2.3 Exposure assessment

We estimated daily ambient temperature at women's home address using the multi-resolution spatiotemporal exposure model described in Chapter 1. The model estimates daily minimum, maximum, and mean air temperature from 2000 to 2018 at a 1 km spatial resolution across France and at a 200 m spatial resolution over urban areas with >50,000 inhabitants. The estimates are very accurate, with cross-validated R^2 better than 0.9 and mean absolute error of about 1°C. We used 200 m temperature for women in urban areas covered by the model (n = 1,741; 33%) and 1 km temperature otherwise.

We calculated five indicators of exposure based on each woman's daily temperature profile from conception to delivery: 1) weekly mean temperature (T_{mean}), a marker of overall exposure; 2) weekly average of daily maximum temperature (T_{max}), a marker of daytime exposure because temperature is usually highest in the afternoon; 3) weekly average of daily minimum temperature (T_{min}), a marker of night-time exposure because temperature is typically lowest before sunrise; 4) weekly temperature variability (T_{SD}, the standard deviation of daily mean temperature), a marker of exposure to temperature swings; and 5) daily Excess Heat Factor (EHF), a marker of exposure to extreme heat that accounts for both spatial and seasonal acclimation(Nairn & Fawcett, 2014). EHF is calculated as:

$$EHF = \max\left(0, EHI_{sig}\right) \times \max(1, EHI_{accl}) \quad \text{Eq. (3-1)}$$

$$EHI_{sig} = \frac{\sum_{j=1}^{3} T_{-j}}{3} - P_{95} \quad \text{Eq. (3-2)}$$

$$EHI_{accl} = \frac{\sum_{j=1}^{3} T_{-j}}{3} - \frac{\sum_{j=1}^{30} T_{-j}}{30} \quad \text{Eq. (3-3)}$$

where EHI_{sig} is a significance index that reflects how the previous three days were relative to the location's typical temperature, EHI_{accl} is an acclimation index that reflects how hot the previous three days were relative to the current season, T_{-j} is the mean temperature on the j^{th} day before day T, and P_{95} is the 95th percentile of historical daily mean temperature. We calculated P_{95} over the 5 years preceding the first conception for each cohort area (Nancy and Poitiers for EDEN; Côtes-d'Armor, Finistère, and Ille-et-Vilaine for PELAGIE; Grenoble for SEPAGES) using daily mean temperature from Météo France meteorological stations, excluding stations missing >35 observations or at >900 m elevation and weighting each station by the number of participants within 20 km. Table A3-1 lists P_{95} of each cohort area.

3.2.4 Main analysis

We assessed the association between temperature indicators and preterm birth using Cox proportional hazards models with duration of pregnancy (weeks since conception) as the time variable and birth as the outcome (censored at 35 weeks after conception). We fit a separate model for each of mean temperature, night-time temperature (T_{min}), daytime temperature (T_{max}), temperature variability, and EHF. In each model, we accounted for the time-varying effects of exposure using a distributed lag nonlinear model (DLNM) (Gasparrini et al., 2010) with two exposure matrices: the 26 weeks following conception (weekly chronic exposure) and the 30

days ending at delivery (daily acute exposure). We censored chronic exposure at 26 weeks because the DLNM requires a complete equal-length exposure history for all subjects; examining chronic exposure later in pregnancy would have required excluding a substantial fraction of preterm births. For EHF, we used days 30-181 after conception (181 days = 26 weeks) as the chronic exposure matrix, because calculating EHF requires a 30-day history. For temperature variability, we used the 4 weeks preceding delivery as the (weekly) acute exposure matrix, because variability is only defined for multi-day windows. We modelled both the exposure-response and the lag-response relationship using natural cubic splines with equally spaced knots and 3 degrees of freedom (chosen by testing 3-6 degrees of freedom for the lowest value that minimized the Akaike information criterion).

We adjusted all models for possible confounders or predictors of the outcome selected *a priori* based on the literature and our reasoning: cohort area (6 levels: EDEN Nancy or Poitiers, PELAGIE Côtes-d'Armor, Finistère, or Ille-et-Vilaine, or SEPAGES Grenoble), season of conception, urbanicity (rural, small city-centre, or suburban), normalized difference vegetation index (NDVI), child sex, parity, and maternal characteristics (age at conception, height, pre-pregnancy BMI, education, and smoking during pregnancy). We did not adjust for gestational hypertension or preeclampsia as these may mediate the association between temperature and preterm birth (we excluded women with pre-existing hypertension). Urbanicity was based on data from the French National Institute of Statistics and Economics for the home address at birth. NDVI was the mean Landsat satellite NDVI (Robinson et al., 2017) in a 500 m buffer around the home address at birth during June-August of the year of birth. We considered NDVI missing if it was unavailable over more than 25% of the buffer.

We calculated the relative risk (RR) and 95% confidence interval (CI) of preterm birth associated with moderate (10^{th}, 90^{th} percentiles), severe (5^{th}, 95^{th} percentiles), and extreme (1^{st}, 99^{th} percentiles) exposure compared to the median exposure (50^{th} percentile) during the chronic and acute periods. We conducted all statistical analyses using R version 4.1.0 (R Core Team, 2021) with the packages survival v3.2-11 (Therneau, 2021) and dlnm v2.4.6 (Gasparrini, 2011).

3.2.5 Sensitivity analyses

To better understand the associations with temperature and evaluate the robustness of our findings, we repeated our analyses 1) including the 309 participants (of which 16 preterm births) that were missing covariates, which we imputed using the cohort area-specific median (for

NDVI, maternal age at conception, maternal height, and maternal pre-pregnancy BMI) or mode (for parity, maternal education and maternal smoking during pregnancy); and 2) using temperature estimated at a 1 km spatial resolution for all participants (rather than using 200 m temperature for the 33% of participants that lived in large urban areas).

3.3 Results

Over half (58%) of the women in this study lived in Brittany; 18% lived in Nancy, 16% in Poitiers, and 8% in Grenoble (Table 3-1). Almost half (45%) lived in a rural area. Most women (72%) were 25 to 34 years old at conception, had completed at least one year of post-secondary education (63%), and were multiparous (56%). A quarter (25%) of women smoked during pregnancy. Mean duration of pregnancy (conception to delivery) was 37.9 weeks and 4.3% of births were preterm. Mean temperature ± SD during the 26 weeks following conception was 11.8 ± 3.6°C; the SD of weekly temperature averaged 2.0 ± 0.3°C. Mean EHF was 0.7 ± 1.1 over days 30 to 181 since conception. Figure A3-1 shows temperature and EHF over time for each cohort and Table A3-2 summarizes their distributions. Table A3-3 compares the characteristics of the participants included and excluded from the main analyses.

3.3.1 Cold and preterm birth

Severe cold (5^{th} vs 50^{th} percentile of T_{mean}) 7-9 weeks after conception and 10-4 days before delivery increased the risk of preterm birth (Figure 3-4). Aa mean temperature of 2°C (vs 11.6°C) throughout weeks 7-9 after conception was associated with RR for preterm birth of 1.29 (95% CI 1.02 to 1.64) (Figure 3-5). A mean temperature of 1.2°C (vs 12.1°C) throughout days 10-4 before delivery had RR of 1.55 (95% CI 1.14-2.11) (Figure 3-5). Focusing on a single week or day during each critical window, the mean RR for preterm birth was 1.09 (95% CI 1.01 to 1.18) following severe cold on one of weeks 7-9 after conception; RR was 1.06 (95% CI 1.01 to 1.12) following severe cold on one of days 10-4 before delivery (Figure A3-2). For extreme cold (1^{st} percentile of T_{mean}), the association was significant 4-9 weeks after conception and 10-4 days before delivery. Moderate cold (10^{th} percentile of T_{mean}) was only significant 10-5 days before delivery. There was a longer critical window for chronic daytime cold (low T_{max}; weeks 6-18 after conception) (Figure 3-4), while night-time cold (low T_{min}) was only significant when it was extreme (1^{st} percentile) 7-6 days before delivery (Figure 3-5). Imputing missing covariates did not substantially alter associations between cold and preterm birth, but using only 1 km temperature made the critical windows shorter or non-significant. Maternal education was

the only variable that did not satisfy the proportional hazards assumption; stratifying the Cox models on education did not substantially alter the results.

Table 3-1. Population characteristics

	Consortium *n (%) or mean (SD)*	EDEN *n (%) or mean (SD)*	PELAGIE *n (%) or mean (SD)*	SEPAGES *n (%) or mean (SD)*
Participants	5347 (100%)	1806 (33.8%)	3116 (58.3%)	425 (7.8%)
Preterm births	232 (4.3%)	103 (5.7%)	110 (3.5%)	19 (4.5%)
Duration of pregnancy[a] (weeks)	37.9 (1.7)	37.7 (1.8)	38.1 (1.6)	37.7 (1.5)
Temperature[b] (°C)	11.8 (3.6)	11.3 (4.1)	12.1 (3.2)	12 (4.4)
Temperature variability[c] (°C)	2.0 (0.3)	2.2 (0.2)	1.9 (0.3)	2.1 (0.2)
Excess Heat Factor[d]	0.7 (1.1)	0.8 (1.3)	0.6 (0.9)	0.7 (1.0)
NDVI	0.50 (0.11)	0.46 (0.10)	0.52 (0.10)	0.49 (0.16)
Child sex				
Boy	2737 (51.2%)	940 (52%)	1572 (50.4%)	225 (52.9%)
Girl	2610 (48.8%)	866 (48%)	1544 (49.6%)	200 (47.1%)
Parity				
0	2374 (44.4%)	794 (44.0%)	1386 (44.5%)	194 (45.6%)
1	2021 (37.8%)	670 (37.1%)	1166 (37.4%)	185 (43.5%)
>=2	952 (17.8%)	342 (18.9%)	564 (18.1%)	46 (10.8%)
Maternal age at conception (years)	30 (4.5)	29.4 (4.9)	29.9 (4.3)	32.5 (3.9)
Maternal height				
135-175 cm	4612 (86.3%)	1561 (86.4%)	2719 (87.3%)	332 (78.1%)
170-190 cm	735 (13.7%)	245 (13.6%)	397 (12.7%)	93 (21.9%)
Maternal pre-pregnancy BMI				
<18.5 kg/m^2	412 (7.7%)	157 (8.7%)	229 (7.3%)	26 (6.1%)
18.5 – 25 kg/m^2	3844 (71.9%)	1177 (65.2%)	2346 (75.3%)	321 (75.5%)
>25 kg/m^2	1091 (20.4%)	472 (26.1%)	541 (17.4%)	78 (18.4%)
Maternal education				
Baccalaureate or less	1987 (37.2%)	840 (46.5%)	1125 (36.1%)	22 (5.2%)
Baccalaureate + 1 or 2 years	1373 (25.7%)	415 (23.0%)	906 (29.1%)	52 (12.2%)
>= Baccalaureate +3 years	1987 (37.2%)	551 (30.5%)	1085 (34.8%)	351 (82.6%)
Smoking status during pregnancy				
None	4006 (74.9%)	1347 (74.6%)	2257 (72.4%)	402 (94.6%)
Active smoker	1341 (25.1%)	459 (25.4%)	859 (27.6%)	23 (5.4%)
Urbanicity				
City-centre	896 (16.8%)	191 (10.6%)	547 (17.6%)	158 (37.2%)
Small city-centre or suburban	2021 (37.8%)	879 (48.7%)	918 (29.5%)	224 (52.7%)
Rural	2430 (45.4%)	736 (40.8%)	1651 (53%)	43 (10.1%)
Season of conception				
Winter	1321 (24.7%)	431 (23.9%)	762 (24.5%)	128 (30.1%)
Spring	1211 (22.6%)	371 (20.5%)	751 (24.1%)	89 (20.9%)
Summer	1477 (27.6%)	527 (29.2%)	860 (27.6%)	90 (21.2%)
Autumn	1338 (25.0%)	477 (26.4%)	743 (23.8%)	118 (27.8%)
Area				
Nancy	940 (17.6%)	940 (52%)	0 (0%)	0 (0%)
Poitiers	866 (16.2%)	866 (48%)	0 (0%)	0 (0%)
Brittany	3116 (58.3%)	0 (0%)	3116 (100%)	0 (0%)
Grenoble	425 (7.9%)	0 (0%)	0 (0%)	425 (100%)

[a] Conception to delivery
[b] Over the 26 weeks following conception
[c] Weekly standard deviation (SD) over the 26 weeks following conception
[d] Over days 30-181 after conception

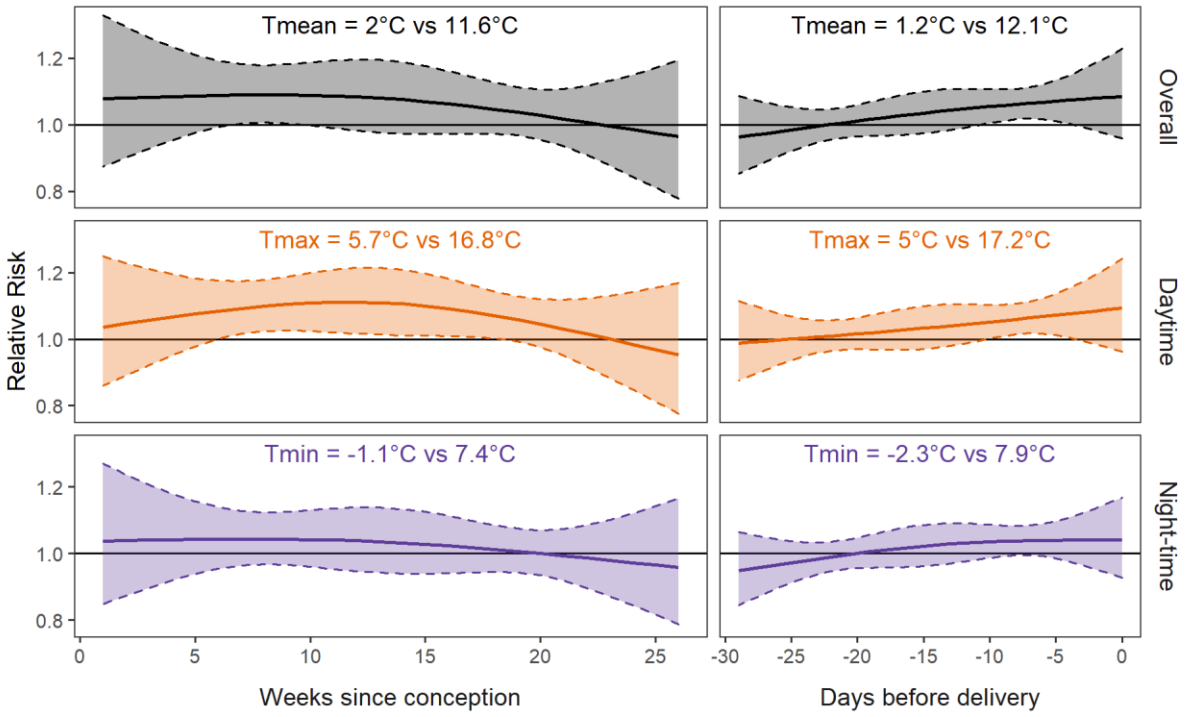

Figure 3-4. Adjusted relative risk (solid line) and 95% confidence interval (shaded area) for preterm birth associated with severe cold during each of the 26 weeks following conception (left) and the 30 days ending at delivery (right). Top: overall cold (5^{th} percentile of T_{mean}); middle: daytime cold (5^{th} percentile of T_{max}); bottom: night-time cold (5^{th} percentile of T_{min}). Reference is 50^{th} percentile of temperature.

3.3.2 Heat and preterm birth

Heat (high T_{mean}) during the first weeks following conception, the second half of trimester 2, and the last days before delivery seemed to correspond with increased risk of preterm birth, but the association was not statistically significant (Figure 3-6). However, we found critical windows for severe night-time heat (95^{th} vs 50^{th} percentile of T_{min}; 15.7°C vs 7.4°C) during weeks 1-5 after conception (RR: 2.00; 95% CI:1.05-3.84) and weeks 20-26 after conception (RR: 2.87; 95% CI 1.21 to 6.79) (Figure 3-6). Moderate (90^{th} percentile) night-time heat showed a smaller effect during the same windows; extreme (99^{th} percentile) night-time heat showed a larger effect but was only significant during weeks 21-26 after conception (Figure 3-5). Focusing on a single week during each critical window, an average minimum temperature of 15.7°C during any one of weeks 1-5 or weeks 21-26 after conception was associated with mean RR for preterm birth of about 1.16 (95% CI 1.02 to 1.34) (Figure A3-2). Sensitivity analyses did not substantially alter the results.

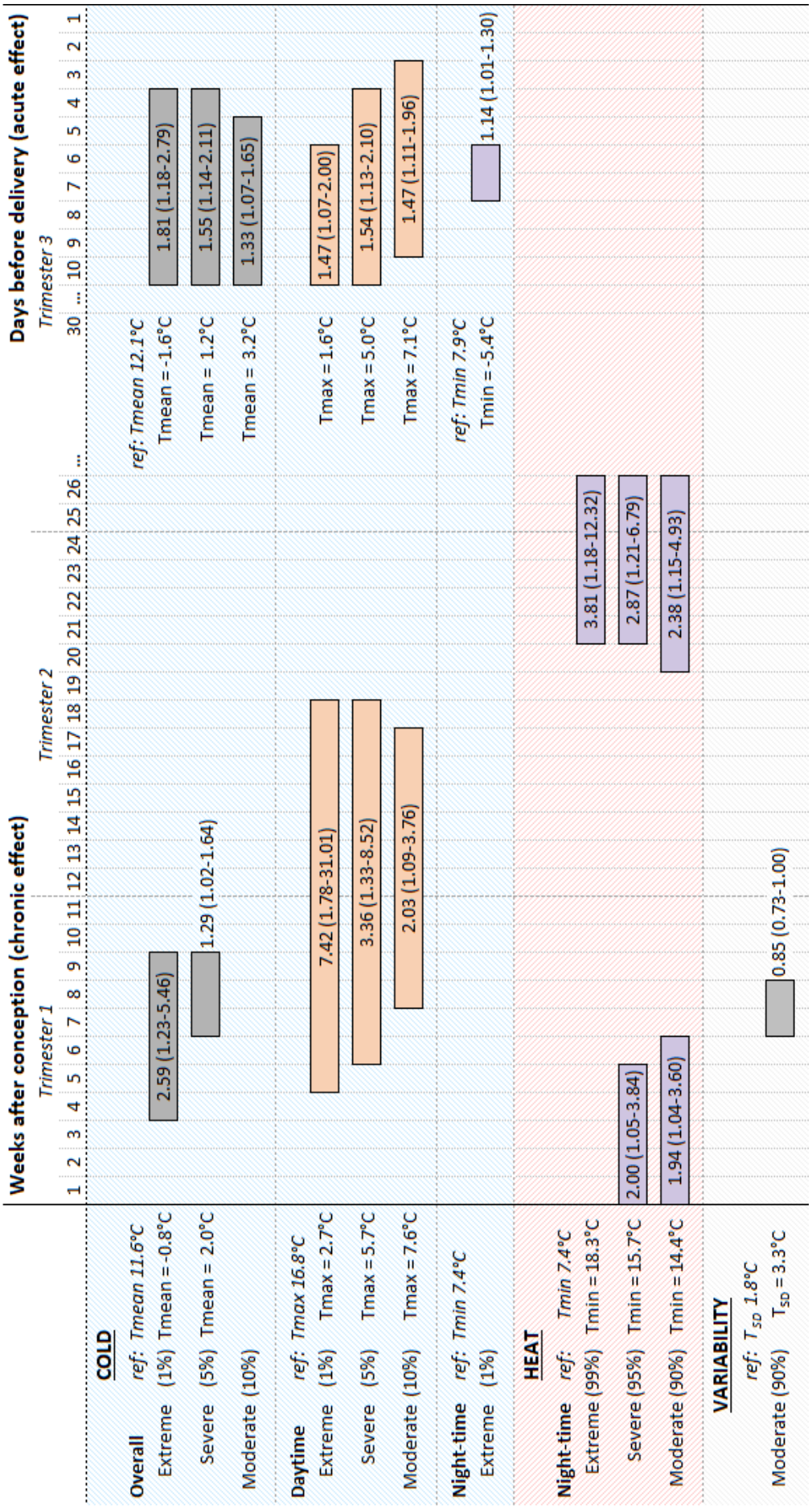

Figure 3-5. Statistically significant critical windows for exposure to indicators of ambient temperature and risk of preterm birth. Bars show timing of critical windows; bar labels show cumulative adjusted relative risk (95% CI) for preterm birth associated with exposure throughout the entire critical window.

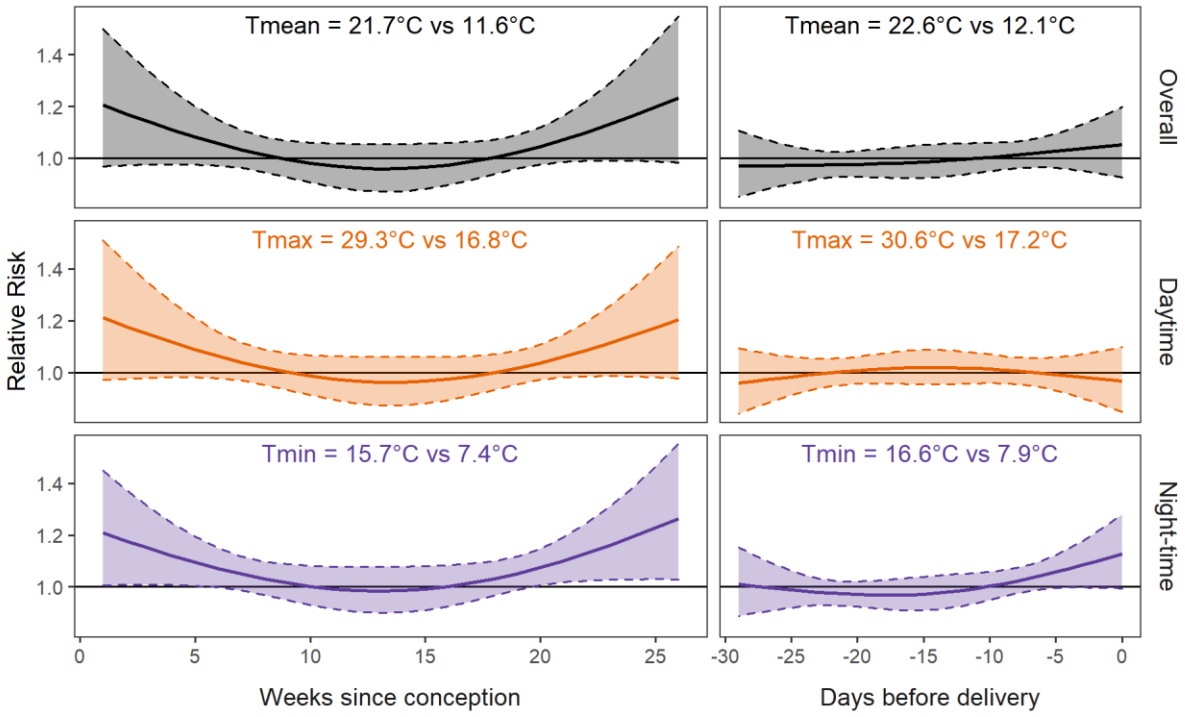

Figure 3-6. Adjusted relative risk (solid line) and 95% confidence interval (shaded area) for preterm birth associated with severe heat during each of the 26 weeks following conception (left) and the 30 days ending at delivery (right). Top: overall heat (95th percentile of T_{mean}); middle: daytime heat (95th percentile of T_{max}); bottom: night-time heat (95th percentile of T_{min}). Reference is 50th percentile of temperature.

3.3.3 Temperature variability and Excess Heat Factor

There was no clear association between temperature variability and risk of preterm birth. Moderately variable temperature (90th vs 50th percentile of T_{SD}; 3.3°C vs 1.8°C) throughout weeks 7 and 8 following conception (Figure A3-3) may have had a protective effect, but the association was barely significant (RR: 0.85; 95% CI 0.73 to 1.00). Other exposure levels were not significant, and sensitivity analyses showed similar results. There was no significant association between EHF and preterm birth (Figure A3-4).

3.4 Discussion

We found evidence of susceptibility to cold from the middle of the first to the middle of the second trimester and around a week before delivery. Women were susceptible to heat during the weeks immediately following conception and the end of the second trimester. Night-time heat (high T_{min}) seemed to be more harmful than daytime (high T_{max}) or overall (high T_{mean}) heat, while night-time cold seemed less harmful than daytime or overall cold. Our results are

based on survival analysis using a spatiotemporally resolved daily exposure model of ambient temperature coupled to home address, controlling for both chronic (weekly for the 26 weeks following conception) and acute (daily for the 30 days ending at delivery) exposure, and accounting for nonlinear lags in the effect of exposure.

3.4.1 Cold and preterm birth

We found evidence of a critical window for chronic cold (low T_{mean}) during weeks 4-9 after conception (in the second half of the first trimester). Women were more sensitive to chronic daytime cold (low T_{max}) than chronic overall or night-time cold (T_{mean} or T_{min}): the critical window for chronic daytime cold continued through week 18 after conception (the middle of the second trimester), while we found no significant critical window for chronic night-time cold. Women may have been less exposed to night-time cold as they were likely indoors and most homes in France are heated.

Relatively few studies have examined the effects of early pregnancy cold in temperate climates, with most finding windows of protective effect: cold before conception in the United States (Ha et al., 2017a), during weeks 6-19 after conception in the United States and China (Guo et al., 2018; Ha et al., 2017a), or during the first trimester in Europe and China (Giorgis-Allemand et al., 2017; Y.-Y. Wang et al., 2020) decreased the risk of preterm birth. However, the United States study also found that cold during the 2 weeks before and 5 weeks following conception increased the risk of early preterm birth (<34 weeks amenorrhea) by 20% (95% CI 11 to 30%) and of late preterm birth (34-36 weeks amenorrhea) by 9% (95% CI 4 to 15%). This compares to our finding of a 29% increase (95% CI 2 to 64%) in risk for preterm birth (<37 weeks amenorrhea) associated with exposure to severe cold (5^{th} percentile of T_{mean}) during weeks 7-9 after conception. The studies in China also reported increased risk from entire pregnancy cold in temperate (Y.-Y. Wang et al., 2020) and cold areas (Guo et al., 2018). Compared to previous studies, we examined narrower windows for chronic exposure (single weeks during the first two trimesters) and estimated temperature based on home address coupled to a fine-scale exposure model (rather than using city-wide or regional temperature), which may have improved our ability to detect critical windows.

We also found an increased risk of preterm birth associated with acute overall and daytime cold (low T_{mean} and T_{max}) 10-4 days before delivery. This effect was adjusted for temperatures earlier in pregnancy, as our models included both chronic (weekly from conception through week 26)

and acute (daily for the 30 days ending at delivery) exposure. Previous studies of late pregnancy cold in temperate climates have reported conflicting results of no association with preterm birth (Guo et al., 2018; Kloog et al., 2015a; Lee et al., 2008, 2018; Vicedo-Cabrera et al., 2015), a decreased risk (Ha et al., 2017a), or a decreased risk from cold on the day before or of delivery but an increased risk from cold 2 or 3 days before delivery (Cox et al., 2016; Sun et al., 2019). Most of these studies considered only the last 7 days (daily exposure) or last four weeks (weekly exposure) of pregnancy, which may have limited their ability to detect a critical window starting around 10 days before delivery, and many did not adjust for temperature earlier in pregnancy. Two recent studies in China found an increased risk of preterm birth associated with cold during the third trimester (He et al., 2016; Y.-Y. Wang et al., 2020). We did not examine third trimester exposure, but the association with daytime cold (low T_{max}) remained significant when cumulated over the 30 days ending at delivery.

3.4.2 Heat and preterm birth

We found critical windows for chronic heat at the start of pregnancy (weeks 1-5 after conception) and at the end of the second and beginning of the third trimester (weeks 20-26 after conception). The association was only significant for night-time heat (high T_{min}), but the shape of the lag-response curve was similar for overall and daytime heat (high T_{mean} or T_{max}). This may indicate a stronger link between night-time rather than daytime heat and preterm birth. Heatwave mortality studies have suggested that hot nights following hot days may be particularly dangerous because they limit the ability to recover from daytime heat (Basu et al., 2008; Laaidi et al., 2012; Murage et al., 2017), and recent studies in California and Belgium found a clearer association with preterm birth for high minimum temperature than high mean temperature (Avalos et al., 2017; Cox et al., 2016). The association with night-time heat may have been particularly clear in our study because we estimated temperature at each woman's home using a fine-scale exposure model that captures the higher night-time temperatures of urban heat islands (Chapter 1). Taken together, these results suggest that future studies should consider night-time heat indicators in order to clarify the effects of heat during pregnancy, particularly in countries such as France where only about 13% of homes have air conditioning (Randazzo et al., 2020). A recent multi-centre study in Europe also found no association between T_{mean} and preterm birth (Giorgis-Allemand et al., 2017), but did not consider night-time heat nor fine temporal windows for chronic exposure. In contrast, studies in the U.S., Europe, China, and Australia have found an increased risk of preterm birth associated with varying windows for heat (T_{mean}) including preconception (Ha et al., 2017a), weeks 1-5, 13-19,

and 19-24 after conception (Ha et al., 2017a; Liu et al., 2020), every trimester (Y.-Y. Wang et al., 2020), and entire pregnancy (Ha et al., 2017a; Kloog et al., 2015a).

Many studies have reported that heat in the last days of pregnancy may trigger preterm delivery (Chersich et al., 2020; Zhang et al., 2017), although a few have reported no effect in cold or cool climates (Guo et al., 2018; Kloog et al., 2015a; Lee et al., 2008; Sun et al., 2019; Vicedo-Cabrera et al., 2015). In our study, heat during the last 5 days of pregnancy suggested an increased risk of preterm birth, but the association did not cross the threshold of significance. We adjusted for temperature earlier in pregnancy, which may have reduced the importance of end of pregnancy heat. Our exposure estimates may also be less accurate during the final days of pregnancy because we estimated exposure based on home address, but some women were likely admitted to maternity units before the day of delivery.

3.4.3 Temperature variability and acclimation

Temperature variability and acclimation to location and season have been shown to affect the risk of mortality (Guo et al., 2016; Lee et al., 2014; Nordio et al., 2015; Shi et al., 2015), but few studies have examined them in relation to birth outcomes. A recent study in the Andes associated more variable temperature with lower birth weight (Molina & Saldarriaga, 2017), and a study in France using a subset of our study population associated more variable temperature during weeks 4-18 after conception with lower term birth weight (Jakpor et al., 2020). A study in Brisbane found evidence of acclimation to higher temperatures between 1994 and 2013: the preterm risk curve for temperature exposure during the third trimester shifted to the right, resulting in greater susceptibility to cold and greater resistance to heat in 2013 compared to 1994 (Li et al., 2018).

We did not find an association between preterm birth and the variability of temperature. This could in part be because more than half of our study population lived in Brittany, a coastal region with an oceanic climate characterized by relatively stable temperatures. Nor did we find an association with EHF, a heatwave index that accounts for acclimation to both location and season. This could be related to the fact that we only found a significant association for night-time heat (high T_{min}) whereas EHF was based on T_{mean}. Our sample size may also have limited our ability to detect an effect associated with an infrequent acute exposure such as EHF. We also adjusted for EHF earlier in pregnancy, which may have reduced the effect of EHF shortly before delivery: our heat analyses similarly showed an increased but non-significant association

during the final days of pregnancy. Further research should seek to clarify the possible role of temperature variability and acclimation.

3.4.4 Biological pathways

The biological pathways linking temperature exposure to birth outcomes remain unclear. Pregnant women have increased fat deposition, decreased surface area to mass ratio, weight gain, and higher metabolic heat production (as the foetus contributes), which could make them more susceptible to heat. Heat may cause the release of cytokines involved in labour induction such as prostaglandin and oxytocin (Dadvand et al., 2011), and their concentration in the blood might be increased by heat-induced dehydration (Schifano et al., 2013). Dehydration and shifting of blood flow to the skin to dissipate heat could limit oxygen supply to the foetus (Sun et al., 2019), and heat shock proteins might cause inflammation (Basu et al., 2017). Heat may also be linked to preeclampsia (Beltran et al., 2013; Shashar et al., 2020). Fewer mechanisms have been proposed for cold, but thermoregulatory responses can cause peripheral vasoconstriction and increase blood pressure and viscosity, which might restrict blood flow to the placenta or contribute to gestational hypertension (Basu et al., 2017; Bruckner et al., 2014).

3.4.5 Strengths and limitations

We note that our study has some limitations. We had a relatively small sample size representing only oceanic and warm temperate climates, which may have limited our ability to detect effects and generalize our findings. We estimated exposure based on ambient (outdoor) temperature at women's home address, but women may have spent a substantial portion of time at other locations (particularly during the day and early in pregnancy) or indoors (particularly at night and late in pregnancy). We also lacked complete address history for a quarter of participants (all from the PELAGIE cohort). We estimated exposure for these women based on municipality or neighbourhood of residence as reported at inclusion (mean 10.4 weeks after conception), which may have increased exposure measurement error and biased our associations towards null.

Since the DLNM requires complete equal-length exposure histories, we only considered exposure during the 26 weeks following conception and the 30 days ending at delivery. This could have led us to miss critical windows early in the third trimester (e.g. weeks 27-30 after conception) because we only considered daily exposure during this period for the few preterm infants born at 32 weeks amenorrhea or less. A recent study in Rome and Barcelona found that

heat during the previous three days was more harmful earlier than later in the third trimester (Schifano et al., 2016), so future studies with sufficient sample size should investigate critical windows in the first half of the third trimester. We also did not account for humidity, which may modify the physiological effects of heat (Davis et al., 2016). However, the evidence for humidity is mixed: some studies found it did not substantially modify the association between temperature and mortality (Armstrong et al., 2019; Barnett et al., 2010) or birth outcomes (Huang et al., 2021) while others suggested a significant role (Jakpor et al., 2020) or that humidity's importance varies between locations (Bobb et al., 2011).

Despite these limitations, our study has several strengths. Pooling three cohorts from different regions of France allowed us to increase the study population and capture greater climatic variability while maintaining detailed health data and similar lifestyles across most participants. We estimated exposure based on daily ambient temperature at participants' home address from a spatiotemporally resolved exposure model. This likely reduced exposure error for the 45% of women that lived in rural areas, which often have few weather monitors, and the 33% of women that lived in large urban areas, where we were able to use 200 m temperature estimates that better capture urban heat islands and other fine-scale spatial patterns. One of the few previous studies to use residence-based temperature exposure found a significant association with birth weight that became close to null when exposure was estimated based on the closest weather monitor (Kloog et al., 2015a), and our sensitivity analyses showed that using 1 km rather than 200 m temperature for urban women weakened the association with preterm birth and shortened the critical windows for exposure.

We used Cox proportional hazards models with pregnancy duration as the time variable to avoid possible confounding by temporal trends in conception rates and the fact that the risk of preterm birth increases exponentially later in pregnancy (Darrow et al., 2009; Strand et al., 2011b). To avoid underestimating gestational age in the case that temperature affects foetal growth during the first trimester (Olsen & Basso, 2005), we preferred gestational duration calculated from the last menstrual period rather than from measurements performed at the first ultrasound. We accounted for nonlinear lags in the effect of exposure and examined narrow windows (each of the 26 weeks following conception and the 30 days ending at delivery) that did not necessarily correspond to clinical trimesters. We adjusted our estimates of chronic effects (weekly for the 26 weeks following conception) for acute effects (daily for the 30 days ending at delivery) and vice versa. We further adjusted for potential confounders such as season of conception, maternal age, education, and smoking, but did not adjust for air pollution because it may be on

the causal pathway from temperature to preterm birth (Buckley et al., 2014). Future studies could investigate the possible synergistic effects of air pollution and temperature (Sun et al., 2020a; Q. Wang et al., 2020).

In this study, we used highly resolved spatiotemporal exposure estimates to assess the association between preterm birth and heat and cold and explored the effects of temperature variability and acclimation. Our results indicate that, in a temperate climate, cold between the middle of the first and second trimesters increases the risk of preterm birth and cold late in pregnancy may trigger preterm birth with a lag time of about one week. We also found that night-time heat may be harmful during the five weeks following conception and the 6th month of pregnancy. We found inconclusive evidence for heat as a short-term trigger of preterm birth. In the context of climate change, the already visible effects of rising temperatures and more frequent weather hazards are expected to multiply over the course of the century, adding to the burden of preterm birth. Health professionals and policy makers could use these findings to increase awareness of the risks of extreme temperature for pregnant women and improve the health of their infants both at birth and throughout their lifespan, as preterm birth is associated with poorer health in childhood and adulthood.

Discussion

Contributions

This PhD made several contributions to the fields of environmental exposure assessment and environmental epidemiology. First, we developed a spatial downscaling technique that allows geostatistical temperature models to increase the spatial resolution of their predictions by incorporating high spatial but low temporal resolution thermal data from the Landsat satellites. We used the technique to estimate daily minimum, mean, and maximum air temperature at a 200 m spatial resolution over 103 urban areas in France from 2000 to 2018. Higher-resolution temperature estimates are particularly relevant for urban areas since they are home to much of the world's population, are often warmer than the surrounding countryside, and can have substantial temperature contrasts over short spatial scales. Improved temperature estimates for urban areas may be useful for epidemiological studies seeking to understand temperature's health effects, for public health officials seeking to identify at-risk neighbourhoods for extreme weather events, and for public planners interested in understanding and mitigating urban heat islands. Our temperature model also estimates daily minimum, mean, and maximum air temperature at a 1 km spatial resolution over continental France from 2000 to 2018.

Second, we developed the first satellite-based geostatistical PM model covering France at daily 1 km resolution from 2000 to 2019. Accurate PM prediction is challenging because monitors are often sparse while AOD is often missing and is not specific to ground-level aerosols. We found that Gaussian Markov random fields (GMRF) captured spatial and particularly temporal variation in PM levels better than widely used mixed models and random forests. Ensembling the predictions of the three algorithms allowed us to correct bias and take advantage of the fact that GMRF was occasionally outperformed by random forest in some areas. To the best of our knowledge, our model's PM estimates represent the most complete record (2000 to 2019) of daily $PM_{2.5}$ and PM_{10} covering continental France at a 1 km resolution.

Finally, we comprehensively assessed the association between ambient temperature and the risk of preterm birth in France. We simultaneously addressed several methodological limitations of previous studies by using spatiotemporally resolved residence-based exposure estimates, examining narrow exposure windows throughout pregnancy, and conducting a survival analysis that accounted for changes over time in both the risk of delivery and the effects of temperature. In contrast to many studies that highlighted heat as a possible short-term trigger of preterm birth, we found susceptibility to cold during trimesters 1 and 2 and around 7 days before delivery. We also found evidence that night-time heat may be particularly harmful, with

susceptibility during the 5 weeks following conception, the second half of trimester 2, and the beginning of trimester 3. Although we examined temperature variability and a heatwave indicator that accounts for acclimation to location and season, we found no clear associations with preterm birth. This suggests that minimum, mean, and maximum temperature are relevant indicators to study in association with preterm birth, while temperature variability and heatwave-acclimation indicators may be less appropriate.

Our findings provide a detailed view of ambient temperature's role during pregnancy with regards to the duration of gestation, which is particularly important in the context of climate change and increasing preterm birth rates. More generally, our ambient temperature and PM models provide a much-improved combination of high spatial and temporal resolution covering a large area (France) and long time period (two decades) with a consistent methodology. This will advance our understanding of the health effects of temperature and PM by allowing pooling across some epidemiological studies, investigating both acute and chronic exposure windows, studying populations that were previously excluded because no exposure assessment tool was available for their area, and exploring acclimation to warming temperatures over the past 20 years.

Methodology

We focused on satellite-based geostatistical models because they can accurately estimate exposure over large areas and long time periods using a consistent methodology. This approach is limited by the availability of satellite data and the number and spatial distribution of temperature and PM monitors. MODIS data is not available prior to 2000, which limits our ability to extend models backwards in time. $PM_{2.5}$ monitors were very rare in France in the early 2000s, so our estimates during that period may be less reliable due to greater average distance from a monitor and lower diversity of monitored areas. Differential weighting of monitors and including training data from more recent years might help alleviate this, provided that the factors that drive PM distribution did not change too much over time. Improving predictions in one area or under certain conditions (e.g. high PM levels) may worsen performance in other areas due to the bias-variance trade-off, and it may be important to consider how explicit or implicit tuning of an exposure model affects subsequent health studies (Sarafian et al., 2020).

Our ambient exposure models only reconstruct outdoor exposure, yet many people spend much of the day and night indoors. The relationship between indoor and outdoor temperature and PM

depends on many factors including building construction (e.g. infiltration and air exchange rates), heating and cooling, behaviour (e.g. opening windows), and the presence of indoor PM sources (tobacco smoke, fuel combustion for heating or cooking) (Hondula et al., 2021; Kuras et al., 2017; Morawska et al., 2013; Nguyen & Dockery, 2016; Ouidir et al., 2015; Quinn et al., 2014). Incorporating indoor exposure could refine effect estimates and improve our understanding of how temperature and PM affect health, complementing the larger body of research on ambient exposure.

Incorporating space-time activity could also refine exposure estimates and improve our understanding of different activities, travel modes, and trajectory choices affect exposure. This could inform public health recommendations and give individuals greater control over their personal exposure. Accounting for individuals' movements becomes increasingly important as the resolution of exposure models increases. For example, it makes little sense to estimate air temperature a 10 m spatial resolution if individuals cannot be located more precisely than their home address. The spatiotemporal variability of exposure is also important: higher resolutions are particularly useful in urban areas, where exposure can vary over fine spatial scales, and for exposures such as nitrous oxides that show large spatial contrasts. For some exposures accounting for space-time activity may be less important than estimating indoor exposure (Ouidir et al., 2015).

The most direct way to assess total personal exposure is with a portable monitor, but this imposes major financial and logistical costs. Emerging low-cost sensors could make personal monitoring more feasible, but data quality is a concern. For example, an evaluation of low-cost temperature monitors used in the SEPAGES study found that 25% were unreliable (Gajardo Alarcón, 2019). Carrying a personal monitor also imposes a substantial burden on study participants, making it challenging to measure exposure over long time periods. This could introduce bias if exposure during the monitored period is not representative of exposure at other times or if willingness to carry a portable monitor and adherence to the study protocol are correlated with factors such as socioeconomic status that may also influence health outcomes.

Regardless of the method used, any measurement of exposure will have some error. If the error is differential, meaning it is related to the health outcome of interest, then it will bias any estimate of the association between the exposure and the health outcome (Armstrong, 1998). If the error is non-differential, then whether it biases the estimate or not depends on whether the error is classical or Berkson (Armstrong, 1998; Zeger et al., 2000). Classical error occurs when

a true exposure x is measured as z with random error that is uncorrelated with the measurement such that $E(z|x) = x$. For example, ambient temperature measured with an unbiased thermometer has classical error because the average of many thermometer readings is equal to the true temperature; higher readings are "cancelled out" by lower readings. Berkson error, first proposed by Joseph Berkson (1950), occurs when a true exposure x is measured as z with random error that is uncorrelated with the true exposure such that $E(x|z) = z$. For example, PM exposure estimated for a group of people based on a central air quality monitor has Berkson error if the true average exposure of the group is equal to the monitor reading; individuals with higher exposure are "cancelled out" by individuals with lower exposure.

Classical error tends to attenuate any association between exposure and health effects (i.e. bias the association towards null). If exposure estimates have substantial classical error, the regression coefficient will tend towards zero even if the true value is non-zero. The degree of attenuation depends on the relative variance of the error and the true exposure: attenuation increases with more variable errors but is mitigated by more variable exposures (Armstrong, 1998; Zeger et al., 2000). In contrast, Berkson error does not bias any association, but it decreases power by increasing the variance of the regression coefficient (Armstrong, 1998; Berkson, 1950; Zeger et al., 2000). If exposure estimates have substantial Berkson error, the regression coefficient will tend towards the true value but will have a wide confidence interval that may include null.

Zeger et al. (2000) analysed the sources, types, and effects of exposure error in time-series mortality studies of air pollution. They concluded that most exposure estimates include both classical and Berkson error and identified differences between personal exposure and ambient measurements as the main source of bias when estimating health effects with single-pollutant models. Our temperature and PM exposure models should reduce bias by providing better estimates of true personal exposure than central monitoring stations, but differences between indoor and ambient exposure remain a possible source of bias. For multi-pollutant models, correlations between error in the different pollutants may also be a substantial source of bias (Zeger et al., 2000).

Our study of the association between temperature and preterm birth has a few potential sources of bias in addition to exposure error. Using prospective cohorts allowed us to estimate exposure at participant's home address, account for changes of residence during pregnancy, and control for several possible confounders and predictors of preterm birth. But it also introduced the

potential for selection bias if willingness to participate was in some way associated with preterm birth risk. The study protocols were designed to limit this risk, but they could not eliminate it or guarantee that the participants are representative of the general population. For example, the SEPAGES participants were more educated, had fewer children, and were less likely to smoke than the EDEN and PELAGIE participants or the average French mother.

Pooling cohort participants from three geographic regions and two time periods increased the variability of participants characteristics and exposures and increased our power, which may have improved our ability to detect effects and decreased the risk that our findings are specific to the study population. Future studies could explore using the larger ELFE national longitudinal cohort (Charles et al., 2021) or the total population of births in France based on data from obligatory 8-day infant health certificates (CS8). On the other hand, using a larger, more diverse study population increases the risk of confounding due to factors such as smoking, education, and unmeasured characteristics, another example of the bias-variance trade-off. Datasets covering larger populations also tend to have less detailed information about address history, maternal characteristics, and behaviour, which limits the accuracy of exposure assessment and the ability to adjust for potential confounders.

We estimated gestational length based on the date of the last menstrual period (LMP) to avoid the risk that gestational age may be underestimated at the first trimester ultrasound if temperature exposure restricts intrauterine growth early in pregnancy. However, using LMP may introduce recall bias if participants did not remember the exact date of their last menses (Weinberg et al., 2015). Figure D-1 shows that participants of our study were substantially more likely to report their LMP as the 1st, 10th, 15th, or 20th day of the month, suggesting that they did not recall the true date. This may have resulted in nondifferential outcome misclassification, which would not bias our effect estimates but would reduce statistical power. It could also have biased our effect estimates if the likelihood of not recalling the exact LMP date was correlated with temperature exposure or factors that influence the risk of preterm birth (e.g. smoking).

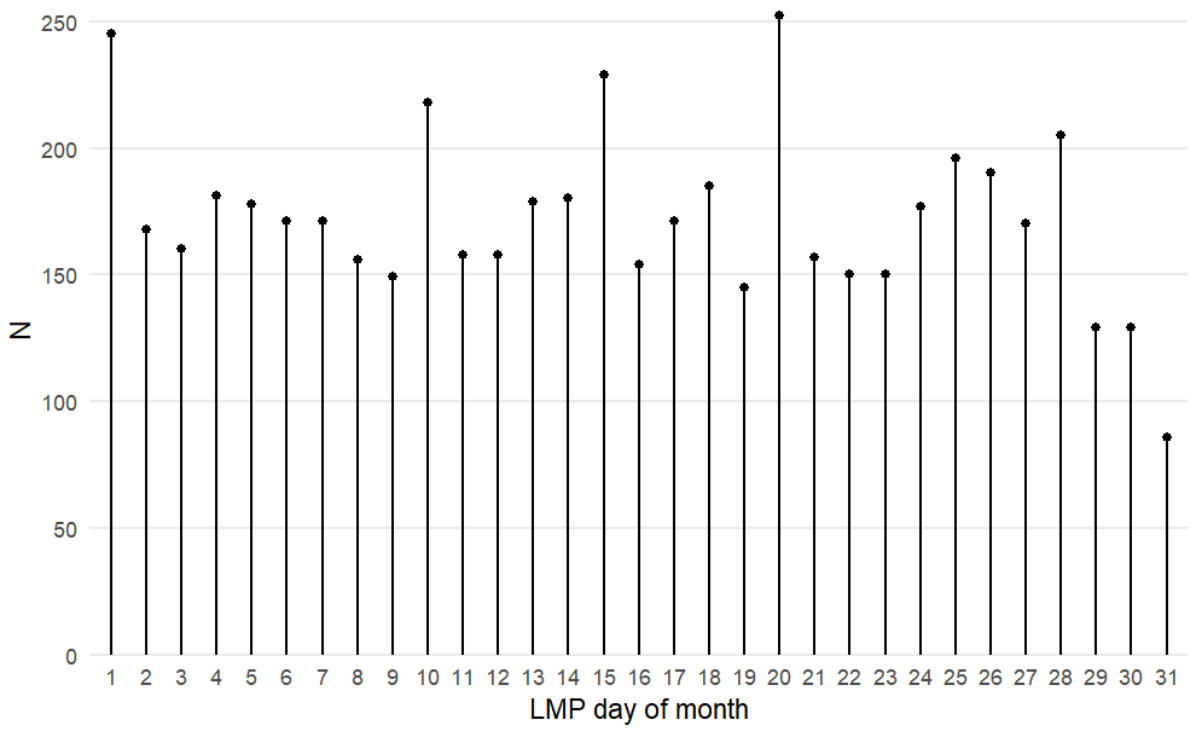

Figure D-1. Digit preference when reporting date of last menstrual period (LMP) in the pooled EDEN, PELAGIE, and SEPAGES data used in Chapter 3. The high frequency of LMP reported on the 1st, 10th, 15th, or 20th day of the month suggests some recall bias.

Perspectives

Taken together, the parts of this PhD suggest several directions for further research. The success of the spatial downscaling method we developed to estimate daily 200 m air temperature suggests that incorporating multiple satellite datasets might allow models to achieve very high spatiotemporal resolutions such as hourly at 200 m. A recent study in Israel combined daily 1 km MODIS LST and 15-minute 4 km SEVIRI LST to reconstruct hourly 1 km air temperature (Zhou et al., 2020), but to our knowledge no study has yet attempted to estimate hourly air temperature at higher than 1 km resolution. Hourly temperatures at high spatial resolution could help clarify which temperature metrics are most relevant to human health. They could also help understand the drivers of urban heat islands, identify vulnerable neighbourhoods, and evaluate heat island mitigation strategies.

Spatial and temporal downscaling could also be useful for PM models. For example, a recent study in China combined daily 1 km MAIAC AOD and hourly 5 km Himawari AOD to estimate hourly 1 km $PM_{2.5}$ over Beijing (Sun et al., 2021). But PM models are currently limited by a lack of AOD at greater than 1 km resolution. We attempted to downscale our daily PM model's

predictions in urban areas without incorporating additional satellite data, but the resulting estimates were less accurate than the original 1 km estimates. Future studies could investigate Sentinel 3's 2-day 300 m AOD product, which is available since late 2018. AOD retrieval at 30 m from Landsat 8 (launched in 2013) and at 10 m from Sentinel 2 (launched in 2015) is an area of active research that might also benefit future studies.

The gap-filling approach we applied to AOD in our PM model could also be applied to LST to facilitate air temperature modelling, as was done by the study that estimated hourly air temperature over Israel. LST may be particularly amenable to gap-filling as it may show repeating seasonal patterns. A study in the United States developed a spatiotemporal gap-filling method that exploited the temporal and seasonal autocorrelation of MODIS LST (X. Li et al., 2018a) and used the resulting gap-free LST to model daily 1 km air temperature (X. Li et al., 2018b). And a recent study developed a method to generate gap-free daily mean, daytime, and night-time 1 km LST worldwide based on MODIS LST and a climate forecasting model (Shiff et al., 2021). If gap-filling techniques are effective for Landsat data, they might increase the accuracy of high spatial resolution temperature estimates, and improved gap-filling methods might translate to better model performance at any spatial resolution. However, few studies have thoroughly investigated how the quality of gap-filling affects the accuracy of final temperature or PM estimates (Pu & Yoo, 2021). Many gap-filling approaches assume that the distribution of LST and AOD is similar on clear-sky and cloudy days, but this may not be the case (Zeng et al., 2018; Zhu et al., 2017). Gap-filling methods that rely on weather forecasting or chemical transport models may also introduce circular dependencies if those models incorporate data from the same PM or temperature monitors that are later used to calibrate the gap-filled AOD or LST.

Our PM model generally captured temporal variation better than spatial variation, possibly due to the sparsity of PM monitors. This suggests that even short timeseries of PM measures at new locations might substantially improve performance. If this is the case, then mobile measurement campaigns or crowdsourcing of data from low-cost home air quality monitors provide a good compliment to longer timeseries of high-quality measurements from official monitoring networks. A better understanding of the relationship between monitor sparsity and model accuracy could focus modelling efforts, support targeting measurement campaigns, or help improve official monitoring networks. Improving model interpretability, particularly for complex machine learning algorithms, could also contribute to modelling and exposure mitigation efforts by providing insight into the factors that drive temperature and PM. Exposure

models could also be extended to estimate PM composition (Chen et al., 2020), which is increasingly recognized as having important health effects (Achilleos et al., 2017).

Epidemiological studies that use temperature and PM models could benefit from more granular information about the uncertainty of the exposure estimates. Models usually report aggregate metrics such as cross-validated annual or regional R^2 and RMSE but do not report uncertainty for individual estimates. Bayesian approaches could provide spatiotemporally continuous uncertainty estimates to accompany predictions (Murray et al., 2019). Epidemiological studies might also benefit from exposure models that are tuned for specific populations or health outcomes (Sarafian et al., 2020).

To complement our results on ambient temperature and preterm birth in France, we are currently using similar methods to study ambient temperature and term birth weight. This could help clarify temperature's role during pregnancy because different pathways may affect gestational duration and intrauterine growth. Our group is also studying prenatal ambient temperature exposure in relation to newborn lung function and pre- and post-natal ambient PM exposure in relation to child cognitive function. Further studies could incorporate exposure estimates from our PM model to explore interactions between temperature and PM, as recent work suggests they may synergistically increase the risk of adverse birth outcomes (Qiu et al., 2020; Sun et al., 2020a; Q. Wang et al., 2020). Further studies could also explore suggested interactions or mediation relationships between temperature, PM, and green space (Asta et al., 2019; Jarvis et al., 2021; Sun et al., 2020b; Yitshak-Sade et al., 2020).

Conclusion

In this PhD, we developed novel approaches to increase the accuracy and spatial resolution of satellite-based exposure models and applied them to reconstruct daily 1 km temperature and PM in France over two decades starting in 2000. Our modelling approach could be applied to improve exposure estimates in other areas, and our high-resolution multi-decadal dataset of temperature and PM in France will be of interest to epidemiologists, climatologists, planners, policymakers, and the public. Future work might extend our approach to estimate hourly temperature, PM, or other exposures at 100 m resolution or finer, evaluate the impact of gap-filling accuracy, and provide uncertainty metrics for the exposure estimates that could be incorporated in epidemiological studies.

We used the reconstructed temperatures to identify periods during pregnancy when exposure to cold or heat increased the risk of preterm birth. We are in the process of examining the effects of heat and cold on term birth weight, and further studies will examine temperature and PM in relation to respiratory function and neurodevelopment. Our datasets could also be used to evaluate interactions between temperature and PM and study urban heat islands in relation to health. In the context of rising temperatures and more frequent weather hazards, our findings on the risks that heat and cold pose to pregnant women and their infants should inform public health policies to reduce the growing burden of preterm birth.

References

Abraham, E., Rousseaux, S., Agier, L., Giorgis-Allemand, L., Tost, J., Galineau, J., Hulin, A., Siroux, V., Vaiman, D., Charles, M.A., Heude, B., Forhan, A., Schwartz, J., Chuffart, F., Bourova-Flin, E., Khochbin, S., Slama, R., & Lepeule, J. (2018). Pregnancy exposure to atmospheric pollution and meteorological conditions and placental DNA methylation. *Environment International*, *118*, 334–347. https://doi.org/10.1016/j.envint.2018.05.007

Achilleos, S., Kioumourtzoglou, M.A., Wu, C. da, Schwartz, J.D., Koutrakis, P., & Papatheodorou, S.I. (2017). Acute effects of fine particulate matter constituents on mortality: A systematic review and meta-regression analysis. *Environment International*, *109*, 89–100. https://doi.org/10.1016/j.envint.2017.09.010

Armstrong, B., Sera, F., Vicedo-Cabrera, A.M., Abrutzky, R., Åström, D.O., Bell, M.L., Chen, B.-Y., Coelho, M. de S.Z.S., Correa, P.M., Dang, T.N., Diaz, M.H., Dung, D. Van, Forsberg, B., Goodman, P., Guo, Y.L., Guo, Y., Hashizume, M., Honda, Y., Indermitte, E., Íñiguez, C., Kan, H., Kim, H., Kyselý, J., Lavigne, E., Michelozzi, P., Orru, H., Ortega, N.V., Pascal, M., Ragettli, M.S., Saldiva, P.H.N., Schwartz, J., Scortichini, M., Seposo, X., Tobias, A., Tong, S., Urban, A., Valencia, C.D. la C., Zanobetti, A., Zeka, A., & Gasparrini, A. (2019). The Role of Humidity in Associations of High Temperature with Mortality: A Multicountry, Multicity Study. *Environmental Health Perspectives*, *127*, 1–8. https://doi.org/10.1289/EHP5430

Armstrong, B.G. (1998). Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occupational and Environmental Medicine*, *55*, 651–656. https://doi.org/10.1136/oem.55.10.651

Arnfield, A.J. (2003). Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology*, *23*, 1–26. https://doi.org/10.1002/joc.859

Arroyo, V., Díaz, J., Carmona, R., Ortiz, C., & Linares, C. (2016). Impact of air pollution and temperature on adverse birth outcomes: Madrid, 2001–2009. *Environmental Pollution*, *218*, 1154–1161. https://doi.org/10.1016/j.envpol.2016.08.069

Asta, F., Michelozzi, P., Cesaroni, G., de Sario, M., Badaloni, C., Davoli, M., & Schifano, P. (2019). The modifying role of socioeconomic position and greenness on the short-term effect of heat and air pollution on preterm births in Rome, 2001–2013. *International Journal of Environmental Research and Public Health*, *16*. https://doi.org/10.3390/ijerph16142497

Auger, N., Naimi, A.I., Smargiassi, A., Lo, E., & Kosatsky, T. (2014). Extreme heat and risk of early delivery among preterm and term pregnancies. *Epidemiology*, *25*, 344–350. https://doi.org/10.1097/EDE.0000000000000074

Avalos, L.A., Chen, H., Li, D.K., & Basu, R. (2017). The impact of high apparent temperature on spontaneous preterm delivery: a case-crossover study. *Environmental Health*, *16*, 1–13. https://doi.org/10.1186/s12940-017-0209-5

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., & Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, *10*, e1443. https://doi.org/10.1002/wics.1443

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D.P., Dutra, E., Munõz-Sabater, J., Pappenberger, F., De Rosnay, P., Stockdale, T., & Vitart, F. (2015). ERA-Interim/Land: a global land surface reanalysis data set. *Hydrology and Earth System Sciences*, *19*, 389–407. https://doi.org/10.5194/hess-19-389-2015

Barker, D.J.P. (2004). The Developmental Origins of Adult Disease. *Journal of the American College of Nutrition*, *23*, 588S-595S. https://doi.org/10.1080/07315724.2004.10719428

Barnett, A.G., Tong, S., & Clements, A.C.A. (2010). What measure of temperature is the best predictor of mortality?. *Environmental Research*, *110*, 604–611. https://doi.org/10.1016/j.envres.2010.05.006

Basu, R., Chen, H., Li, D.K., & Avalos, L.A. (2017). The impact of maternal factors on the association between temperature and preterm delivery. *Environmental Research*, *154*, 109–114. https://doi.org/10.1016/j.envres.2016.12.017

Basu, R., Feng, W.-Y., & Ostro, B.D. (2008). Characterizing temperature and mortality in nine California counties.. *Epidemiology*, *19*, 138–45. https://doi.org/10.1097/EDE.0b013e31815c1da7

Basu, R., Malig, B., & Ostro, B.D. (2010). High ambient temperature and the risk of preterm delivery. *American Journal of Epidemiology*, *172*, 1108–1117. https://doi.org/10.1093/aje/kwq170

Basu, R., Rau, R., Pearson, D., & Malig, B. (2018). Temperature and Term Low Birth Weight in California. *American Journal of Epidemiology*, *187*, 2306–2314. https://doi.org/10.1093/aje/kwy116

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 48. https://doi.org/10.18637/jss.v067.i01

Bechtel, B., Zakšek, K., Oßenbrügge, J., Kaveckis, G., & Böhner, J. (2017). Towards a satellite based monitoring of urban air temperatures. *Sustainable Cities and Society*, *34*, 22–31. https://doi.org/10.1016/j.scs.2017.05.018

Beckerman, B.S., Jerrett, M., Serre, M.L., Martin, R. V, Lee, S.-J., van Donkelaar, A., Ross, Z., Su, J., & Burnett, R.T. (2013). A Hybrid Approach to Estimating National Scale Spatiotemporal Variability of PM2.5 in the Contiguous United States. *Environmental Science & Technology*, *47*, 7233–7241. https://doi.org/10.1021/es400039u

Bekkar, B., Pacheco, S., Basu, R., Basu, R., & Denicola, N. (2020). Association of Air Pollution and Heat Exposure with Preterm Birth, Low Birth Weight, and Stillbirth in the US: A Systematic Review. *JAMA Network Open*, *3*, 1–13. https://doi.org/10.1001/jamanetworkopen.2020.8243

Belbasis, L., Savvidou, M.D., Kanu, C., Evangelou, E., & Tzoulaki, I. (2016). Birth weight in relation to health and disease in later life: An umbrella review of systematic reviews and meta-analyses. *BMC Medicine*, *14*. https://doi.org/10.1186/s12916-016-0692-5

Beloconi, A., Chrysoulakis, N., Lyapustin, A., Utzinger, J., & Vounatsou, P. (2018). Bayesian geostatistical modelling of PM10 and PM2.5 surface level concentrations in Europe using high-resolution satellite-derived products. *Environment International*, *121*, 57–70. https://doi.org/10.1016/j.envint.2018.08.041

Beltran, A.J., Wu, J., & Laurent, O. (2013). Associations of meteorology with adverse pregnancy outcomes: A systematic review of preeclampsia, preterm birth and birth weight. *International Journal of Environmental Research and Public Health*, *11*, 91–172. https://doi.org/10.3390/ijerph110100091

Berkson, J. (1950). Are There Two Regressions?. *Journal of the American Statistical Association*, *45*, 164–180. https://doi.org/10.2307/2280676

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., & Jones, Z.M. (2016). Mlr: Machine learning in R. *Journal of Machine Learning Research*, *17*, 1–5.

Bischl, B., Richter, J., Bossek, J., Horn, D., Thomas, J., & Lang, M. (2017). mlrMBO: A Modular Framework for Model-Based Optimization of Expensive Black-Box Functions. *arXiv*,. https://doi.org/10.13140/RG.2.2.11865.31849

Blondel, B., Coulm, B., Bonnet, C., Goffinet, F., & Le Ray, C. (2017). Trends in perinatal health in metropolitan France from 1995 to 2016: Results from the French National Perinatal Surveys. *Journal of Gynecology Obstetrics and Human Reproduction*, *46*, 701–713. https://doi.org/10.1016/j.jogoh.2017.09.002

Bobb, J.F., Dominici, F., & Peng, R.D. (2011). A Bayesian model averaging approach for estimating the relative risk of mortality associated with heat waves in 105 U.S. cities. *Biometrics*, *67*, 1605–1616. https://doi.org/10.1111/j.1541-0420.2011.01583.x

Bruckner, T.A., Modin, B., & Vågerö, D. (2014). Cold ambient temperature in utero and birth outcomes in Uppsala, Sweden, 1915-1929. *Annals of Epidemiology*, *24*, 116–121. https://doi.org/10.1016/j.annepidem.2013.11.005

Buckley, J.P., Samet, J.M., & Richardson, D.B. (2014). Does air pollution confound studies of temperature?. *Epidemiology*, *25*, 242–245. https://doi.org/10.1097/EDE.0000000000000051

Cameletti, M., Lindgren, F., Simpson, D., & Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, *97*, 109–131. https://doi.org/10.1007/s10182-012-0196-3

Charles, M.-A., Thierry, X., Lanoe, J.L., Bois, C., Dufourg, M.N., Popa, R., Cheminat, M., Zaros, C., & Geay, B. (2021). Cohort Profile: The French national cohort of children (ELFE): Birth to 5 years. *International Journal of Epidemiology*, *49*, 368–369. https://doi.org/10.1093/IJE/DYZ227

Chawanpaiboon, S., Vogel, J.P., Moller, A.B., Lumbiganon, P., Petzold, M., Hogan, D., Landoulsi, S., Jampathong, N., Kongwattanakul, K., Laopaiboon, M., Lewis, C., Rattanakanokchai, S., Teng, D.N., Thinkhamrop, J., Watananirun, K., Zhang, J., Zhou, W., & Gülmezoglu, A.M. (2019). Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *The Lancet Global Health*, *7*, e37–e46. https://doi.org/10.1016/S2214-109X(18)30451-0

Chen, F., Yang, S., Yin, K., & Chan, P. (2017). Challenges to quantitative applications of Landsat observations for the urban thermal environment. *Journal of Environmental Sciences*, *59*, 80–88. https://doi.org/10.1016/j.jes.2017.02.009

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Janssen, N.A.H., Martin, R. V., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Vermeulen, R., Brunekreef, B., & Hoek, G. (2019). A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environment International*, *130*. https://doi.org/10.1016/j.envint.2019.104934

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Weinmayr, G., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Atkinson, R.W., Janssen, N.A.H., Martin, R. v., Samoli, E., Andersen, Z.J., Oftedal, B.M., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Brunekreef, B., & Hoek, G. (2020). Development of Europe-Wide Models for Particle Elemental Composition Using

Supervised Linear Regression and Random Forest. *Environmental Science & Technology*, *54*, 15698–15709. https://doi.org/10.1021/acs.est.0c06595

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System, in: KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Cheng, J., Xu, Z., Zhu, R., Wang, X., Jin, L., Song, J., & Su, H. (2014). Impact of diurnal temperature range on human health: a systematic review. *International Journal of Biometeorology*, *58*, 2011–2024. https://doi.org/10.1007/s00484-014-0797-5

Chersich, M.F., Pham, M.D., Area, A., Haghighi, M.M., Manyuchi, A., Swift, C.P., Wernecke, B., Robinson, M., Hetem, R., Boeckmann, M., & Hajat, S. (2020). Associations between high temperatures in pregnancy and risk of preterm birth, low birth weight, and stillbirths: Systematic review and meta-analysis. *The BMJ*, *371*, 1–13. https://doi.org/10.1136/bmj.m3811

Chudnovsky, A.A., Koutrakis, P., Kloog, I., Melly, S.J., Nordio, F., Lyapustin, A., Wang, Y., & Schwartz, J. (2014). Fine particulate matter predictions using high resolution Aerosol Optical Depth (AOD) retrievals. *Atmospheric Environment*, *89*, 189–198. https://doi.org/10.1016/j.atmosenv.2014.02.019

Cox, B., Vicedo-Cabrera, A.M., Gasparrini, A., Roels, H.A., Martens, E., Vangronsveld, J., Forsberg, B., & Nawrot, T.S. (2016). Ambient temperature as a trigger of preterm delivery in a temperate climate. *Journal of Epidemiology and Community Health*, *70*, 1191–1199. https://doi.org/10.1136/jech-2015-206384

Dadvand, P., Basagaña, X., Sartini, C., Figueras, F., Vrijheid, M., de Nazelle, A., Sunyer, J., & Nieuwenhuijsen, M.J. (2011). Climate extremes and the length of gestation. *Environmental Health Perspectives*, *119*, 1449–1453. https://doi.org/10.1289/ehp.1003241

Dadvand, P., Parker, J., Bell, M.L., Bonzini, M., Brauer, M., Darrow, L.A., Gehring, U., Glinianaia, S. v., Gouveia, N., Ha, E.-H., Leem, J.-H., van den Hooven, E.H., Jalaludin, B., Jesdale, B.M., Lepeule, J., Morello-Frosch, R., Morgan, G.G., Pesatori, A.C., Pierik, F.H., Pless-Mulloli, T., Rich, D.Q., Sathyanarayana, S., Seo, J., Slama, R., Strickland, M.J., Tamburic, L., Wartenberg, D., Nieuwenhuijsen, M.J., & Woodruff, T.J. (2013). Maternal Exposure to Particulate Air Pollution and Term Birth Weight: A Multi-Country Evaluation of Effect and Heterogeneity. *Environmental Health Perspectives*, *121*, 367–373. https://doi.org/10.1289/ehp.1205575

Darrow, L.A., Strickland, M.J., Klein, M., Waller, L.A., Flanders, W.D., Correa, A., Marcus, M., & Tolbert, P.E. (2009). Seasonality of Birth and Implications for Temporal Studies of Preterm Birth. *Epidemiology*, *20*, 699–706. https://doi.org/10.1097/EDE.0b013e3181a66e96

Davis, R.E., McGregor, G.R., & Enfield, K.B. (2016). Humidity: A review and primer on atmospheric moisture and human health. *Environmental Research*, *144*, 106–116. https://doi.org/10.1016/j.envres.2015.10.014

de Hoogh, K., Héritier, H., Stafoggia, M., Künzli, N., & Kloog, I. (2018). Modelling daily PM2.5 concentrations at high spatio-temporal resolution across Switzerland. *Environmental Pollution*, *233*, 1147–1154. https://doi.org/10.1016/j.envpol.2017.10.025

De'Donato, F.K., Stafoggia, M., Rognoni, M., Poncino, S., Caranci, N., Bisanti, L., Demaria, M., Forastiere, F., Michelozzi, P., Pelosini, R., & Perucci, C.A. (2008). Airport and city-centre temperatures in the evaluation of the association between heat and mortality. *International Journal of Biometeorology*,. https://doi.org/10.1007/s00484-007-0124-5

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N.N., & Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597. https://doi.org/10.1002/qj.828

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., & Schwartz, J. (2019). An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. *Environment International*, *130*, 104909. https://doi.org/10.1016/j.envint.2019.104909

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., & Schwartz, J. (2016). Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States. *Environmental Science & Technology*, *50*, 4712–4721. https://doi.org/10.1021/acs.est.5b06121

Didan, K., Barreto Munoz, A., Solano, R., & Huete, A. (2015). MODIS Vegetation Index User's Guide (MOD13 Series).

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dedele, A., Dons, E., De Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Gražulevičiene, R., Heinrich, J., Hoffmann, B., Jerrett, M., Keidel, D., Korek, M., Lanki, T., Lindley, S., Madsen, C., Mölter, A., Nádor, G., Nieuwenhuijsen, M.J., Nonnemacher, M., Pedeli, X., Raaschou-Nielsen, O., Patelarou, E., Quass, U., Ranzi, A., Schindler, C., Stempfelet, M., Stephanou, E.,

Sugiri, D., Tsai, M.Y., Yli-Tuomi, T., Varró, M.J., Vienneau, D., Klot, S. Von, Wolf, K., Brunekreef, B., & Hoek, G. (2012). Development of land use regression models for PM2.5, PM2.5 absorbance, PM10 and PMcoarse in 20 European study areas; Results of the ESCAPE project. *Environmental Science & Technology*, *46*, 11195–11205. https://doi.org/10.1021/es301948k

Florio, E.N., Lele, S.R., Chang, Y.C., Sterner, R., & Glass, G.E. (2004). Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: a statistical approach. *International Journal of Remote Sensing*, *25*, 2979–2994. https://doi.org/10.1080/01431160310001624593

Freitas, S.C., Trigo, I.F., Macedo, J., Barroso, C., Silva, R., & Perdigão, R. (2013). Land surface temperature from multiple geostationary satellites. *International Journal of Remote Sensing*, *34*, 3051–3068. https://doi.org/10.1080/01431161.2012.716925

Fu, G., Shen, Z., Zhang, X., Shi, P., Zhang, Y., & Wu, J. (2011). Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature. *Acta Ecologica Sinica*, *31*, 8–13. https://doi.org/10.1016/j.chnaes.2010.11.002

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2019). Constructing Priors that Penalize the Complexity of Gaussian Random Fields. *Journal of the American Statistical Association*, *114*, 445–452. https://doi.org/10.1080/01621459.2017.1415907

Gajardo Alarcón, N. (2019). Comparison of temperature exposure using different exposure models for women from three mother-child cohorts.

Gasparrini, A. (2011). Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. *Journal of statistical software*, *43*, 1–20. https://doi.org/10.18637/jss.v043.i08

Gasparrini, A., Armstrong, B.G., & Kenward, M.G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, *29*, 2224–2234. https://doi.org/10.1002/sim.3940

Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S., Rocklöv, J., Forsberg, B., Leone, M., De Sario, M., Bell, M.L., Guo, Y.L., Wu, C.-F., Kan, H., Yi, S.M., De Sousa Zanotti Stagliorio Coelho, M., Saldiva, P.H.N., Honda, Y., Kim, H., & Armstrong, B.G. (2015). Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *The Lancet*, *386*, 369–375. https://doi.org/10.1016/S0140-6736(14)62114-0

Gasparrini, A., Guo, Y., Sera, F., Vicedo-Cabrera, A.M., Huber, V., Tong, S., de Sousa Zanotti Stagliorio Coelho, M., Nascimento Saldiva, P.H., Lavigne, E., Matus Correa, P., Valdes Ortega, N., Kan, H., Osorio, S., Kyselý, J., Urban, A., Jaakkola, J.J.K., Ryti, N.R.I., Pascal, M., Goodman, P.G., Zeka, A., Michelozzi, P., Scortichini, M., Hashizume, M., Honda, Y., Hurtado-Diaz, M., Cesar Cruz, J., Seposo, X., Kim, H.,

Tobias, A., Iñiguez, C., Forsberg, B., Åström, D.O., Ragettli, M.S., Guo, Y.L., Wu, C.-F., Zanobetti, A., Schwartz, J., Bell, M.L., Dang, T.N., Van, D. Do, Heaviside, C., Vardoulakis, S., Hajat, S., Haines, A., & Armstrong, B.G. (2017). Projections of temperature-related excess mortality under climate change scenarios. *The Lancet Planetary Health*, 360–367. https://doi.org/10.1016/S2542-5196(17)30156-0

Giorgis-Allemand, L., Pedersen, M., Bernard, C., Aguilera, I., Beelen, R.M.J., Chatzi, L., Cirach, M., Danileviciute, A., Dedele, A., Van Eijsden, M., Estarlich, M., Fernández-Somoano, A., Fernández, M.F., Forastiere, F., Gehring, U., Grazuleviciene, R., Gruzieva, O., Heude, B., Hoek, G., de Hoogh, K., Van Den Hooven, E.H., Haberg, S.E., Iñiguez, C., Jaddoe, V.W.V., Korek, M., Lertxundi, A., Lepeule, J., Nafstad, P., Nystad, W., Patelarou, E., Porta, D., Postma, D., Raaschou-Nielsen, O., Rudnai, P., Siroux, V., Sunyer, J., Stephanou, E., Sørensen, M., Thorup Eriksen, K., Tuffnell, D., Varró, M.J., Vrijkotte, T.G.M., Wijga, A., Wright, J., Nieuwenhuijsen, M.J., Pershagen, G., Brunekreef, B., Kogevinas, M., & Slama, R. (2017). The influence of meteorological factors and atmospheric pollutants on the risk of preterm birth. *American Journal of Epidemiology*, *185*, 247–258. https://doi.org/10.1093/aje/kww141

Goldberg, D.L., Gupta, P., Wang, K., Jena, C., Zhang, Y., Lu, Z., & Streets, D.G. (2019). Using gap-filled MAIAC AOD and WRF-Chem to estimate daily PM2.5 concentrations at 1 km resolution in the Eastern United States. *Atmospheric Environment*, *199*, 443–452. https://doi.org/10.1016/j.atmosenv.2018.11.049

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gui, K., Che, H., Zeng, Z., Wang, Y., Zhai, S., Wang, Z., Luo, M., Zhang, L., Liao, T., Zhao, H., Li, L., Zheng, Y., & Zhang, X. (2020). Construction of a virtual PM2.5 observation network in China based on high-density surface meteorological observations using the Extreme Gradient Boosting model. *Environment International*, *141*, 105801. https://doi.org/10.1016/j.envint.2020.105801

Guo, T., Wang, Y.-Y., Zhang, H., Zhang, Y., Zhao, J., Wang, Y., Xie, X., Wang, L., Zhang, Q., Liu, D., He, Y., Yang, Y., Xu, J., Peng, Z., & Ma, X. (2018). The association between ambient temperature and the risk of preterm birth in China. *Science of the Total Environment*, *613–614*, 439–446. https://doi.org/10.1016/j.scitotenv.2017.09.104

Guo, Y., Barnett, A.G., & Tong, S. (2013). Spatiotemporal model or time series model for assessing city-wide temperature effects on mortality?. *Environmental Research*, *120*, 55–62. https://doi.org/10.1016/j.envres.2012.09.001

Guo, Y., Gasparrini, A., Armstrong, B.G., Li, S., Tawatsupa, B., Tobias, A., Lavigne, E., De Sousa Zanotti Stagliorio Coelho, M., Leone, M., Pan, X., Tong, S., Tian, L., Kim, H., Hashizume, M., Honda, Y., Guo, Y.L., Wu, C.-F., Punnasiri, K., Yi, S.M., Michelozzi, P., Saldiva, P.H.N., & Williams, G.M. (2014). Global variation in the effects of ambient temperature on mortality: A systematic evaluation. *Epidemiology*, *25*, 781–789. https://doi.org/10.1097/EDE.0000000000000165

Guo, Yuming, Gasparrini, A., Armstrong, B.G., Tawatsupa, B., Tobias, A., Lavigne, E., De Sousa Zanotti Stagliorio Coelho, M., Pan, X., Kim, H., Hashizume, M., Honda, Y., Leon Guo, Y.L., Wu, C.-F., Zanobetti, A., Schwartz, J.D., Bell, M.L., Scortichini, M., Michelozzi, P., Punnasiri, K., Li, S., Tian, L., Garcia, S.D.O., Seposo, X., Overcenco, A., Zeka, A., Goodman, P.G., Dang, T.N., Van Dung, D., Mayvaneh, F., Saldiva, P.H.N., Williams, G., & Tong, S. (2017). Heat wave and mortality: A multicountry, multicommunity study. *Environmental Health Perspectives*, *125*, 1–11. https://doi.org/10.1289/EHP1026

Guo, Y., Gasparrini, A., Armstrong, B.G., Tawatsupa, B., Tobias, A., Lavigne, E., Zanotti Stagliorio Coelho, M. de S., Pan, X., Kim, H., Hashizume, M., Honda, Y., Guo, Y.L., Wu, C.-F., Zanobetti, A., Schwartz, J., Bell, M.L., Overcenco, A., Punnasiri, K., Li, S., Tian, L., Saldiva, P., Williams, G., & Tong, S. (2016). Temperature variability and mortality: A multi-country study. *Environmental Health Perspectives*, *124*, 1554–1559. https://doi.org/10.1289/EHP149

Guo, Yuanxi, Tang, Q., Gong, D.Y., & Zhang, Z. (2017). Estimating ground-level PM2.5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sensing of Environment*, *198*, 140–149. https://doi.org/10.1016/j.rse.2017.06.001

Ha, S., Hu, H., Roussos-Ross, D., Haidong, K., Roth, J., & Xu, X. (2014). The effects of air pollution on adverse birth outcomes. *Environmental Research*, *134*, 198–204. https://doi.org/10.1016/j.envres.2014.08.002

Ha, S., Liu, D., Zhu, Y., Kim, S.S., Sherman, S., & Mendola, P. (2017a). Ambient temperature and early delivery of Singleton Pregnancies. *Environmental Health Perspectives*, *125*, 453–459. https://doi.org/10.1289/EHP97

Ha, S., Zhu, Y., Liu, D., Sherman, S., & Mendola, P. (2017b). Ambient temperature and air quality in relation to small for gestational age and term low birthweight. *Environmental Research*, *155*, 394–400. https://doi.org/10.1016/j.envres.2017.02.021

Habets, F., Boone, A., Champeaux, J.L., Etchevers, P., Franchistéguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., Noilhan, J., Quintana-Seguí, P., Rousset-Regimbeau, F., & Viennot, P. (2008). The SAFRAN-ISBA-MODCOU

hydrometeorological model applied over France. *Journal of Geophysical Research: Atmospheres*, *113*, 1–18. https://doi.org/10.1029/2007JD008548

Hair, J.H., Reuter, D.C., Tonn, S.L., McCorkel, J., Amy, A.S., Djam, M., Alexander, D., Ballou, K., Barclay, R., Coulter, P., Edick, M., Efremova, B., Finneran, P., Florez, J., Graham, S., Harbert, K., Hewitt, D., Hickey, M., Hicks, S., Hoge, W., Jhabvala, M., Lilly, C., Lunsford, A., Mann, L., Masters, C., Montanaro, M., Muench, T., Otero, V., Parong, F., Pearlman, A., Penn, J., Vigneau, D., & Wenny, B. (2018). Landsat 9 thermal infrared sensor 2 architecture and design. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2018-July*, 8841–8844. https://doi.org/10.1109/IGARSS.2018.8518269

Hanson, M., & Gluckman, P. (2011). Developmental origins of noncommunicable disease: Population and public health implications. *American Journal of Clinical Nutrition*, *94*, 1754–1758. https://doi.org/10.3945/ajcn.110.001206

Harrison, M.S., & Goldenberg, R.L. (2016). Global burden of prematurity. *Seminars in Fetal and Neonatal Medicine*, *21*, 74–79. https://doi.org/10.1016/j.siny.2015.12.007

He, J.-R., Liu, Y., Xia, X.-Y., Ma, W.-J., Lin, H.-L., Kan, H., Lu, J.-H., Feng, Q., Mo, W.-J., Wang, P., Xia, H.-M., Qiu, X., & Muglia, L.J. (2016). Ambient Temperature and the Risk of Preterm Birth in Guangzhou, China (2001–2011). *Environmental Health Perspectives*, *124*, 1100–1106. https://doi.org/10.1289/ehp.1509778

He, Q., Gao, K., Zhang, L., Song, Y., & Zhang, M. (2021a). Satellite-derived 1-km estimates and long-term trends of PM2.5 concentrations in China from 2000 to 2018. *Environment International*, *156*, 106726. https://doi.org/10.1016/j.envint.2021.106726

He, Q., Gu, Y., & Zhang, M. (2020). Spatiotemporal trends of PM2.5 concentrations in central China from 2003 to 2018 based on MAIAC-derived high-resolution data. *Environment International*, *137*, 105536. https://doi.org/10.1016/j.envint.2020.105536

He, Q., & Huang, B. (2018). Satellite-based mapping of daily high-resolution ground PM2.5 in China via space-time regression modeling. *Remote Sensing of Environment*, *206*, 72–83. https://doi.org/10.1016/j.rse.2017.12.018

He, Q., Zhang, M., Song, Y., & Huang, B. (2021b). Spatiotemporal assessment of PM2.5 concentrations and exposure in China from 2013 to 2017 using satellite-derived data. *Journal of Cleaner Production*, *286*, 124965. https://doi.org/10.1016/j.jclepro.2020.124965

Hengl, T., Heuvelink, G.B.M., Tadić, M.P., & Pebesma, E. (2012). Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images.

Theoretical and Applied Climatology, *107*, 265–277. https://doi.org/10.1007/s00704-011-0464-2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., & Thépaut, J.N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049. https://doi.org/10.1002/qj.3803

Heude, B., Forhan, A., Slama, R., Douhaud, L., Bedel, S., Saurel-Cubizolles, M.-J., Hankard, R., Thiebaugeorges, O., Agostini, M. De, Annesi-Maesano, I., Kaminski, M., & Charles, M.-A. (2016). Cohort Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants of child health and development. *International Journal of Epidemiology*, 353–363. https://doi.org/10.1093/ije/dyv151

Ho, H.C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., & Henderson, S.B. (2014). Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*, *154*, 38–45. https://doi.org/10.1016/j.rse.2014.08.012

Ho, H.C., Knudby, A., Walker, B.B., & Henderson, S.B. (2017). Delineation of spatial variability in the temperature-mortality relationship on extremely hot days in greater Vancouver, Canada. *Environmental Health Perspectives*, *125*, 66–75. https://doi.org/10.1289/EHP224

Ho, H.C., Knudby, A., Xu, Y., Hodul, M., & Aminipouri, M. (2016). A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area. *Science of the Total Environment*, *544*, 929–938. https://doi.org/10.1016/j.scitotenv.2015.12.021

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, *42*, 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057

Hondula, D.M., Kuras, E.R., Betzel, S., Drake, L., Eneboe, J., Kaml, M., Munoz, M., Sevig, M., Singh, M., Ruddell, B.L., & Harlan, S.L. (2021). Novel metrics for relating personal heat exposure to social risk factors and outdoor ambient temperature. *Environment International*, *146*, 106271. https://doi.org/10.1016/j.envint.2020.106271

Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L., Strickland, M., & Liu, Y. (2017). Estimating PM2.5 Concentrations in the Conterminous United States Using the

Random Forest Approach. *Environmental Science & Technology*, 1–29. https://doi.org/10.1021/acs.est.7b01210

Hu, X., Waller, L.A., Lyapustin, A., Wang, Y., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G., Estes, S.M., Quattrochi, D.A., Puttaswamy, S.J., & Liu, Y. (2014). Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sensing of Environment*, *140*, 220–232. https://doi.org/10.1016/j.rse.2013.08.032

Huang, M., Strickland, M.J., Richards, M., Holmes, H.A., Newman, A.J., Garn, J. V., Liu, Y., Warren, J.L., Chang, H.H., & Darrow, L.A. (2021). Acute associations between heatwaves and preterm and early-term birth in 50 US metropolitan areas: a matched case-control study. *Environmental Health*, *20*, 1–14. https://doi.org/10.1186/s12940-021-00733-y

Inness, A., Ades, M., Agustí-Panareda, A., Barr, J., Benedictow, A., Blechschmidt, A.M., Jose Dominguez, J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.H., Razinger, M., Remy, S., Schulz, M., & Suttie, M. (2019). The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, *19*, 3515–3556. https://doi.org/10.5194/acp-19-3515-2019

Insee (2016). Gridded 2010 population at 200 m [WWW Document]. URL https://www.insee.fr/fr/statistiques/2520034 (accessed 5.15.17).

Jacobs, M., Zhang, G., Chen, S., Mullins, B., Bell, M., Jin, L., Guo, Y., Huxley, R., & Pereira, G. (2017). The association between ambient air pollution and selected adverse pregnancy outcomes in China: A systematic review. *Science of the Total Environment*, *579*, 1179–1192. https://doi.org/10.1016/j.scitotenv.2016.11.100

Jakpor, O., Chevrier, C., Kloog, I., Benmerad, M., Giorgis-Allemand, L., Cordier, S., Seyve, E., Vicedo-Cabrera, A.M., Slama, R., Heude, B., Schwartz, J., & Lepeule, J. (2020). Term birthweight and critical windows of prenatal exposure to average meteorological conditions and meteorological variability. *Environment International*, *142*, 105847. https://doi.org/10.1016/j.envint.2020.105847

Jarvis, I., Davis, Z., Sbihi, H., Brauer, M., Czekajlo, A., Davies, H.W., Gergel, S.E., Guhn, M., Jerrett, M., Koehoorn, M., Oberlander, T.F., Su, J., & van den Bosch, M. (2021). Assessing the association between lifetime exposure to greenspace and early childhood development and the mediation effects of air pollution and noise in Canada: a population-based birth cohort study.. *The Lancet. Planetary health*, *5*, e709–e717. https://doi.org/10.1016/S2542-5196(21)00235-7

Jedlička, K., Hájek, P., Čada, V., Martolos, J., Šťastný, J., Beran, D., Kolovský, F., & Kozhukh, D. (2016). Open Transport Map - Routable OpenStreetMap, in: IST-Africa

2016 Conference Proceedings. pp. 1–11. https://doi.org/10.1109/ISTAFRICA.2016.7530657

Jiang, T., Chen, B., Nie, Z., Ren, Z., Xu, B., & Tang, S. (2021). Estimation of hourly full-coverage PM2.5 concentrations at 1-km resolution in China using a two-stage random forest model. *Atmospheric Research*, *248*, 105146. https://doi.org/10.1016/j.atmosres.2020.105146

Jiménez-Muñoz, J.C., Cristóbal, J., Sobrino, J.A., Sòria, G., Ninyerola, M., & Pons, X. (2009). Revision of the single-channel algorithm for land surface temperature retrieval from landsat thermal-infrared data. *IEEE Transactions on Geoscience and Remote Sensing*, *47*, 339–349. https://doi.org/10.1109/TGRS.2008.2007125

Jiménez-Muñoz, J.C., Sobrino, J.A., Skokovic, D., Mattar, C., & Cristóbal, J. (2014). Land surface temperature retrieval methods from Landsat-8 thermal infrared sensor data. *IEEE Geoscience and Remote Sensing Letters*, *11*, 1840–1843. https://doi.org/10.1109/LGRS.2014.2312032

Jin, M., & Dickinson, R.E. (2010). Land surface skin temperature climatology: Benefitting from the strengths of satellite observations. *Environmental Research Letters*, *5*. https://doi.org/10.1088/1748-9326/5/4/044004

Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., & Wavresky, P. (2010). Les types de climats en France, une construction spatiale. *Cybergeo*, *501*. https://doi.org/10.4000/cybergeo.23155

Ju, L., Li, C., Yang, M., Sun, S., Zhang, Q., Cao, J., & Ding, R. (2021). Maternal air pollution exposure increases the risk of preterm birth: Evidence from the meta-analysis of cohort studies. *Environmental Research*, *202*, 111654. https://doi.org/10.1016/j.envres.2021.111654

Just, A.C., Arfer, K.B., Rush, J., Dorman, M., Shtein, A., Lyapustin, A., & Kloog, I. (2020). Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM2.5) using satellite data over large regions. *Atmospheric Environment*, *239*, 117649. https://doi.org/10.1016/j.atmosenv.2020.117649

Just, A.C., Wright, R.O., Schwartz, J., Coull, B.A., Baccarelli, A.A., Tellez-Rojo, M.M., Moody, E., Wang, Y., Lyapustin, A., & Kloog, I. (2015). Using High-Resolution Satellite Aerosol Optical Depth To Estimate Daily PM2.5 Geographical Distribution in Mexico City. *Environmental Science & Technology*, *49*, 8576–8584. https://doi.org/10.1021/acs.est.5b00859

Kannan, S., Misra, D.P., Dvonch, J.T., & Krishnakumar, A. (2006). Exposures to airbone particulate matter and adverse perinatal outcomes: A biologically plausible mechanistic framework for exploring potential effect modification by nutrition.

Environmental Health Perspectives, *114*, 1636–1642. https://doi.org/10.1289/ehp.9081

Keramitsoglou, I., Kiranoudis, C.T., Sismanidis, P., & Zakšek, K. (2016). An online system for nowcasting satellite derived temperatures for urban areas. *Remote Sensing*, *8*, 1–17. https://doi.org/10.3390/rs8040306

Kestens, Y., Brand, A., Fournier, M., Goudreau, S., Kosatsky, T., Maloley, M., & Smargiassi, A. (2011). Modelling the variation of land surface temperature as determinant of risk of heat-related health events.. *International Journal of Health Geographics*, *10*, 7. https://doi.org/10.1186/1476-072X-10-7

Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Tadić, M.P., & Bajat, B. (2014). Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research: Atmospheres*, *119*, 2294–2313. https://doi.org/10.1002/2013JD020803

Kloog, I., Chudnovsky, A.A., Just, A.C., Nordio, F., Koutrakis, P., Coull, B.A., Lyapustin, A., Wang, Y., & Schwartz, J. (2014a). A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth data. *Atmospheric Environment*, *95*, 581–590. https://doi.org/10.1016/j.atmosenv.2014.07.014

Kloog, I., Chudnovsky, A.A., Koutrakis, P., & Schwartz, J. (2012a). Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts, USA. *Science of the Total Environment*, *432*, 85–92. https://doi.org/10.1016/j.scitotenv.2012.05.095

Kloog, I., Koutrakis, P., Coull, B.A., Lee, H.J., & Schwartz, J. (2011). Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment*, *45*, 6267–6275. https://doi.org/10.1016/j.atmosenv.2011.08.066

Kloog, I., Melly, S.J., Coull, B.A., Nordio, F., & Schwartz, J. (2015a). Using Satellite-Based Spatiotemporal Resolved Air Temperature Exposure to Study the Association between Ambient Air Temperature and Birth Outcomes in Massachusetts. *Environmental Health Perspectives*, *123*, 1053–1058. https://doi.org/10.1289/ehp.1308075

Kloog, I., Melly, S.J., Ridgway, W.L., Coull, B.A., & Schwartz, J. (2012b). Using new satellite based exposure methods to study the association between pregnancy PM2.5 exposure, premature birth and birth weight in Massachusetts. *Environmental Health*, *11*, 40. https://doi.org/10.1186/1476-069X-11-40

Kloog, I., Nordio, F., Coull, B.A., & Schwartz, J. (2014b). Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the

Northeastern USA. *Remote Sensing of Environment*, *150*, 132–139. https://doi.org/10.1016/j.rse.2014.04.024

Kloog, I., Nordio, F., Coull, B.A., & Schwartz, J. (2012c). Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the mid-atlantic states. *Environmental Science & Technology*, *46*, 11913–11921. https://doi.org/10.1021/es302673e

Kloog, I., Nordio, F., Lepeule, J., Padoan, A., Lee, M., Auffray, A., & Schwartz, J. (2017). Modelling spatio-temporally resolved air temperature across the complex geo-climate area of France using satellite-derived land surface temperature data. *International Journal of Climatology*, *37*, 296–304. https://doi.org/10.1002/joc.4705

Kloog, I., Novack, L., Erez, O., Just, A.C., & Raz, R. (2018). Associations between ambient air temperature, low birth weight and small for gestational age in term neonates in southern Israel. *Environmental Health*, *17*, 76. https://doi.org/10.1186/s12940-018-0420-z

Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B.A., Wang, Y., Just, A.C., Schwartz, J., & Broday, D.M. (2015b). Estimating daily PM2.5 and PM10 across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. *Atmospheric Environment*, *122*, 409–416. https://doi.org/10.1016/j.atmosenv.2015.10.004

Krähenmann, S., Walter, A., Brienen, S., Imbery, F., & Matzarakis, A. (2018). High-resolution grids of hourly meteorological variables for Germany. *Theoretical and Applied Climatology*, *131*, 899–926. https://doi.org/10.1007/s00704-016-2003-7

Kuras, E.R., Richardson, M.B., Calkins, M.M., Ebi, K.L., Hess, J.J., Kintziger, K.W., Jagger, M.A., Middel, A., Scott, A.A., Spector, J.T., Uejio, C.K., Vanos, J.K., Zaitchik, B.F., Gohlke, J.M., & Hondula, D.M. (2017). Opportunities and challenges for personal heat exposure research. *Environmental Health Perspectives*, *125*. https://doi.org/10.1289/EHP556

Laaidi, K., Zeghnoun, A., Dousset, B., Bretin, P., Vandentorren, S., Giraudet, E., & Beaudeau, P. (2012). The impact of heat islands on mortality in Paris during the August 2003 heat wave. *Environmental Health Perspectives*, *120*, 254–259. https://doi.org/10.1289/ehp.1103532

Laaidi, M., Laaidi, K., & Besancenot, J.P. (2006). Temperature-related mortality in France, a comparison between regions with different climates from the perspective of global warming. *International Journal of Biometeorology*, *51*, 145–153. https://doi.org/10.1007/s00484-006-0045-8

Landrigan, P.J., Fuller, R., Acosta, N.J.R., Adeyi, O., Arnold, R., Basu, N., Baldé, A.B., Bertollini, R., Bose-O'Reilly, S., Boufford, J.I., Breysse, P.N., Chiles, T., Mahidol, C., Coll-Seck, A.M., Cropper, M.L., Fobil, J., Fuster, V., Greenstone, M., Haines, A.,

Hanrahan, D., Hunter, D., Khare, M., Krupnick, A., Lanphear, B., Lohani, B., Martin, K., Mathiasen, K. V., McTeer, M.A., Murray, C.J.L., Ndahimananjara, J.D., Perera, F., Potočnik, J., Preker, A.S., Ramesh, J., Rockström, J., Salinas, C., Samson, L.D., Sandilya, K., Sly, P.D., Smith, K.R., Steiner, A., Stewart, R.B., Suk, W.A., van Schayck, O.C.P., Yadama, G.N., Yumkella, K., & Zhong, M. (2017). The Lancet Commission on pollution and health. *The Lancet*, *391*. https://doi.org/10.1016/S0140-6736(17)32345-0

Le Tertre, A., Lefranc, A., Eilstein, D., Declercq, C., Medina, S., Blanchard, M., Chardon, B., Fabre, P., Filleul, L., Jusot, J.F., Pascal, L., Prouvost, H., Cassadou, S., & Ledrans, M. (2006). Impact of the 2003 heatwave on all-cause mortality in 9 French cities. *Epidemiology*, *17*, 75–79. https://doi.org/10.1097/01.ede.0000187650.36636.1f

Leaf, J.S., & Erell, E. (2018). A model of the ground surface temperature for micrometeorological analysis. *Theoretical and Applied Climatology*, *133*, 697–710. https://doi.org/10.1007/s00704-017-2207-5

Lee, H.J., Coull, B.A., Bell, M.L., & Koutrakis, P. (2012). Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM2.5 concentrations. *Environmental Research*, *118*, 8–15. https://doi.org/10.1016/j.envres.2012.06.011

Lee, H.J., Liu, Y., Coull, B.A., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict PM2.5 concentrations. *Atmospheric Chemistry and Physics*, *11*, 7991–8002. https://doi.org/10.5194/acp-11-7991-2011

Lee, M., Kloog, I., Chudnovsky, A.A., Lyapustin, A., Wang, Y., Melly, S., Coull, B.A., Koutrakis, P., & Schwartz, J. (2016). Spatiotemporal prediction of fine particulate matter using high-resolution satellite images in the Southeastern US 2003-2011. *Journal of Exposure Science and Environmental Epidemiology*, *26*, 377–384. https://doi.org/10.1038/jes.2015.41

Lee, M., Nordio, F., Zanobetti, A., Kinney, P.L., Vautard, R., & Schwartz, J. (2014). Acclimatization across space and time in the effects of temperature on mortality: a time-series analysis.. *Environmental Health*, *13*, 89. https://doi.org/10.1186/1476-069X-13-89

Lee, S.J., Hajat, S., Steer, P.J., & Filippi, V. (2008). A time-series analysis of any short-term effects of meteorological and air pollution factors on preterm births in London, UK. *Environmental Research*, *106*, 185–194. https://doi.org/10.1016/j.envres.2007.10.003

Lee, W., Bell, M.L., Gasparrini, A., Armstrong, B.G., Sera, F., Hwang, S., Lavigne, E., Zanobetti, A., Coelho, M. de S.Z.S., Saldiva, P.H.N., Osorio, S., Tobias, A., Zeka, A., Goodman, P.G., Forsberg, B., Rocklöv, J., Hashizume, M., Honda, Y., Guo, Y.L., Seposo, X., Van Dung, D., Dang, T.N., Tong, S., Guo, Y., & Kim, H. (2018). Mortality burden of diurnal temperature range and its temporal changes: A multi-

country study. *Environment International*, *110*, 123–130. https://doi.org/10.1016/j.envint.2017.10.018

Li, C., Yang, M., Zhu, Z., Sun, S., Zhang, Q., Cao, J., & Ding, R. (2020). Maternal exposure to air pollution and the risk of low birth weight: A meta-analysis of cohort studies. *Environmental Research*, *190*, 109970. https://doi.org/10.1016/j.envres.2020.109970

Li, J., Woodward, A., Hou, X.Y., Zhu, T., Zhang, J., Brown, H., Yang, J., Qin, R., Gao, J., Gu, S., Li, J., Xu, L., Liu, X., & Liu, Q. (2017). Modification of the effects of air pollutants on mortality by temperature: A systematic review and meta-analysis. *Science of the Total Environment*, *575*, 1556–1570. https://doi.org/10.1016/j.scitotenv.2016.10.070

Li, L., Franklin, M., Girguis, M., Lurmann, F., Wu, J., Pavlovic, N., Breton, C., Gilliland, F., & Habre, R. (2020). Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling. *Remote Sensing of Environment*, *237*, 111584. https://doi.org/10.1016/j.rse.2019.111584

Li, L., Zhang, J., Meng, X., Fang, Y., Ge, Y., Wang, J., Wang, C., Wu, J., & Kan, H. (2018). Estimation of PM2.5 concentrations at a high spatiotemporal resolution using constrained mixed-effect bagging models with MAIAC aerosol optical depth. *Remote Sensing of Environment*, *217*, 573–586. https://doi.org/10.1016/j.rse.2018.09.001

Li, L., Zhang, J., Qiu, W., Wang, J., & Fang, Y. (2017). An ensemble spatiotemporal model for predicting PM2.5 concentrations. *International Journal of Environmental Research and Public Health*, *14*. https://doi.org/10.3390/ijerph14050549

Li, S., Chen, G., Jaakkola, J.J.K., Williams, G., & Guo, Y. (2018). Temporal change in the impacts of ambient temperature on preterm birth and stillbirth: Brisbane, 1994–2013. *Science of the Total Environment*, *634*, 579–585. https://doi.org/10.1016/j.scitotenv.2018.03.385

Li, X., Huang, S., Jiao, A., Yang, X., Yun, J., Wang, Y., Xue, X., Chu, Y., Liu, F., Liu, Y., Ren, M., Chen, X., Li, N., Lu, Y., Mao, Z., Tian, L., & Xiang, H. (2017). Association between ambient fine particulate matter and preterm birth or term low birth weight: An updated systematic review and meta-analysis. *Environmental Pollution*, *227*, 596–605. https://doi.org/10.1016/j.envpol.2017.03.055

Li, X., Zhou, Y., Asrar, G.R., & Zhu, Z. (2018a). Creating a seamless 1 km resolution daily land surface temperature dataset for urban and surrounding areas in the conterminous United States. *Remote Sensing of Environment*, *206*, 84–97. https://doi.org/10.1016/j.rse.2017.12.010

Li, X., Zhou, Y., Asrar, G.R., & Zhu, Z. (2018b). Developing a 1 km resolution daily air temperature dataset for urban and surrounding areas in the conterminous United

States. *Remote Sensing of Environment*, *215*, 74–84. https://doi.org/10.1016/j.rse.2018.05.034

Li, Z.-L., Tang, B.H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., & Sobrino, J.A. (2013a). Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*, *131*, 14–37. https://doi.org/10.1016/j.rse.2012.12.008

Li, Z.-L., Wu, H., Wang, N., Qiu, S., Sobrino, J.A., Wan, Z., Tang, B.-H., & Yan, G. (2013b). Land surface emissivity retrieval from satellite data. *International Journal of Remote Sensing*, *34*, 3084–3127. https://doi.org/10.1080/01431161.2012.716540

Liang, F., Xiao, Q., Wang, Y., Lyapustin, A., Li, G., Gu, D., Pan, X., & Liu, Y. (2018). MAIAC-based long-term spatiotemporal trends of PM2.5 in Beijing, China. *Science of the Total Environment*, *616–617*, 1589–1598. https://doi.org/10.1016/j.scitotenv.2017.10.155

Liang, Z., Lin, Y., Ma, Y., Zhang, L., Zhang, X., Li, L., Zhang, S., Cheng, Y., Zhou, X., Lin, H., Miao, H., & Zhao, Q. (2016). The association between ambient temperature and preterm birth in Shenzhen, China: A distributed lag non-linear time series analysis. *Environmental Health*, *15*, 1–11. https://doi.org/10.1186/s12940-016-0166-4

Lin, V.W., Baccarelli, A.A., & Burris, H.H. (2016). Epigenetics—a potential mediator between air pollution and preterm birth. *Environmental Epigenetics*, *2*, dvv008. https://doi.org/10.1093/eep/dvv008

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, *73*, 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Liu, C., Chen, R., Sera, F., Vicedo-Cabrera, A.M., Guo, Y., Tong, S., Coelho, M.S.Z.S., Saldiva, P.H.N., Lavigne, E., Matus, P., Valdes Ortega, N., Osorio Garcia, S., Pascal, M., Stafoggia, M., Scortichini, M., Hashizume, M., Honda, Y., Hurtado-Díaz, M., Cruz, J., Nunes, B., Teixeira, J.P., Kim, H., Tobias, A., Íñiguez, C., Forsberg, B., Åström, C., Ragettli, M.S., Guo, Y.L., Chen, B.-Y., Bell, M.L., Wright, C.Y., Scovronick, N., Garland, R.M., Milojevic, A., Kyselý, J., Urban, A., Orru, H., Indermitte, E., Jaakkola, J.J.K., Ryti, N.R.I., Katsouyanni, K., Analitis, A., Zanobetti, A., Schwartz, J., Chen, J., Wu, T., Cohen, A., Gasparrini, A., & Kan, H. (2019). Ambient Particulate Air Pollution and Daily Mortality in 652 Cities. *New England Journal of Medicine*, *381*, 705–715. https://doi.org/10.1056/nejmoa1817364

Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., Lawn, J.E., Cousens, S., Mathers, C., & Black, R.E. (2016). Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable

Development Goals. *The Lancet*, *388*, 3027–3035. https://doi.org/10.1016/S0140-6736(16)31593-8

Liu, X., Xiao, J., Sun, X., Chen, Q., Yao, Z., Feng, B., Cao, G., Guo, L., He, G., Hu, J., Zeng, W., Rong, Z., Wang, Q., Zhang, B., Dong, M., Wang, J., Chen, D., Ma, W., & Liu, T. (2020). Associations of maternal ambient temperature exposures during pregnancy with the risk of preterm birth and the effect modification of birth order during the new baby boom: A birth cohort study in Guangzhou, China. *International Journal of Hygiene and Environmental Health*, *225*, 113481. https://doi.org/10.1016/j.ijheh.2020.113481

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, *11*, 5741–5765. https://doi.org/10.5194/amt-11-5741-2018

Lyon-Caen, S., Siroux, V., Lepeule, J., Lorimier, P., Hainaut, P., Mossuz, P., Quentin, J., Supernant, K., Meary, D., Chaperot, L., Bayat, S., Cassee, F., Valentino, S., Couturier-Tarrade, A., Rousseau-Ralliard, D., Chavatte-Palmer, P., Philippat, C., Pin, I., & Slama, R. (2019). Deciphering the impact of early-life exposures to highly variable environmental factors on foetal and child health: Design of SEPAGES couple-child cohort. *International Journal of Environmental Research and Public Health*, *16*. https://doi.org/10.3390/ijerph16203888

Ma, Z., Hu, X., Huang, L., Bi, J., & Liu, Y. (2014). Estimating Ground-Level PM2.5 in China Using Satellite Remote Sensing. *Environmental Science & Technology*, *48*, 7436–7444. https://doi.org/10.1021/es5009399

Malakar, N.K., Hulley, G.C., Hook, S.J., Laraby, K.G., Cook, M., & Schott, J.R. (2018). An Operational Land Surface Temperature Product for Landsat Thermal Data: Methodology and Validation. *IEEE Transactions on Geoscience and Remote Sensing*, 1–19. https://doi.org/10.1109/TGRS.2018.2824828

McCormick, M.C., Litt, J.S., Smith, V.C., & Zupancic, J.A.F. (2011). Prematurity: An Overview and Public Health Implications. *Annual Review of Public Health*, *32*, 367–379. https://doi.org/10.1146/annurev-publhealth-090810-182459

Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., Xue, W., Wang, J., Wang, D., Kan, H., & Liu, Y. (2016). Estimating ground-level PM10 in a Chinese city by combining satellite data, meteorological information and a land use regression model. *Environmental Pollution*, *208*, 177–184. https://doi.org/10.1016/j.envpol.2015.09.042

Meng, X., Liu, C., Zhang, L., Wang, W., Stowell, J., Kan, H., & Liu, Y. (2021). Estimating PM2.5 concentrations in Northeastern China with full spatiotemporal coverage, 2005–2016. *Remote Sensing of Environment*, *253*, 112203. https://doi.org/10.1016/j.rse.2020.112203

Molina, O., & Saldarriaga, V. (2017). The perils of climate change: In utero exposure to temperature variability and birth outcomes in the Andean region. *Economics and Human Biology*, *24*, 111–124. https://doi.org/10.1016/j.ehb.2016.11.009

Morawska, L., Afshari, A., Bae, G.N., Buonanno, G., Chao, C.Y.H., Hänninen, O., Hofmann, W., Isaxon, C., Jayaratne, E.R., Pasanen, P., Salthammer, T., Waring, M., & Wierzbicka, A. (2013). Indoor aerosols: From personal exposure to risk assessment. *Indoor Air*, *23*, 462–487. https://doi.org/10.1111/ina.12044

Murage, P., Hajat, S., & Kovats, R.S. (2017). Effect of night-time temperatures on cause and age-specific mortality in London. *Environmental Epidemiology*, 1. https://doi.org/10.1097/ee9.0000000000000005

Murray, C.J.L., Aravkin, A.Y., Zheng, P., Abbafati, C., Abbas, K.M., Abbasi-Kangevari, M., Abd-Allah, F., & et al. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *396*, 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2

Murray, N.L., Holmes, H.A., Liu, Y., & Chang, H.H. (2019). A Bayesian ensemble approach to combine PM2.5 estimates from statistical models using satellite imagery and numerical model simulation. *Environmental Research*, *178*, 108601. https://doi.org/10.1016/j.envres.2019.108601

Nairn, J.R., & Fawcett, R.J.B. (2014). The excess heat factor: A metric for heatwave intensity and its use in classifying heatwave severity. *International Journal of Environmental Research and Public Health*, *12*, 227–253. https://doi.org/10.3390/ijerph120100227

Ngo, N.S., & Horton, R.M. (2016). Climate change and fetal health: The impacts of exposure to extreme temperatures in New York City. *Environmental Research*, *144*, 158–164. https://doi.org/10.1016/j.envres.2015.11.016

Nguyen, J.L., & Dockery, D.W. (2016). Daily indoor-to-outdoor temperature and humidity relationships: a sample across seasons and diverse climatic regions. *International Journal of Biometeorology*, *60*, 221–229. https://doi.org/10.1007/s00484-015-1019-5

Noi, P.T., Degener, J., & Kappas, M. (2017). Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. *Remote Sensing*, *9*. https://doi.org/10.3390/rs9050398

Nordio, F., Kloog, I., Coull, B.A., Chudnovsky, A.A., Grillo, P., Bertazzi, P.A., Baccarelli, A.A., & Schwartz, J. (2013). Estimating spatio-temporal resolved PM10 aerosol mass concentrations using MODIS satellite data and land use regression over Lombardy, Italy. *Atmospheric Environment*, *74*, 227–236. https://doi.org/10.1016/j.atmosenv.2013.03.043

Nordio, F., Zanobetti, A., Colicino, E., Kloog, I., & Schwartz, J. (2015). Changing patterns of the temperature-mortality association by time and location in the US, and implications for climate change. *Environment International*, *81*, 80–86. https://doi.org/10.1016/j.envint.2015.04.009

Olsen, J., & Basso, O. (2005). Reproductive epidemiology, in: Ahrens, W., Pigeot, I. (Eds.), Handbook of Epidemiology. Springer, New York, NY, pp. 1705–1777. https://doi.org/10.1007/978-0-387-09834-0_27

Ouidir, M., Giorgis-Allemand, L., Lyon-Caen, S., Morelli, X., Cracowski, C., Pontet, S., Pin, I., Lepeule, J., Siroux, V., & Slama, R. (2015). Estimation of exposure to atmospheric pollutants during pregnancy integrating space-time activity and indoor air levels: Does it make a difference?. *Environment International*, *84*, 161–173. https://doi.org/10.1016/j.envint.2015.07.021

Oyler, J.W., Ballantyne, A., Jencso, K., Sweet, M., & Running, S.W. (2015). Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *International Journal of Climatology*, *35*, 2258–2279. https://doi.org/10.1002/joc.4127

Oyler, J.W., Dobrowski, S.Z., Holden, Z.A., & Running, S.W. (2016). Remotely sensed land skin temperature as a spatial predictor of air temperature across the conterminous United States. *Journal of Applied Meteorology and Climatology*, *55*, 1441–1457. https://doi.org/10.1175/JAMC-D-15-0276.1

Parastatidis, D., Mitraka, Z., Chrysoulakis, N., & Abrams, M. (2017). Online global land surface temperature estimation from landsat. *Remote Sensing*, *9*, 1–16. https://doi.org/10.3390/rs9121208

Park, Y., Kwon, B., Heo, J., Hu, X., Liu, Y., & Moon, T. (2020). Estimating PM2.5 concentration of the conterminous United States via interpretable convolutional neural networks. *Environmental Pollution*, *256*, 113395. https://doi.org/10.1016/j.envpol.2019.113395

Parmentier, B., McGill, B.J., Wilson, A.M., Regetz, J., Jetz, W., Guralnick, R.P., Tuanmu, M.N., & Schildhauer, M. (2015). Using multi-timescale methods and satellite-derived land surface temperature for the interpolation of daily maximum air temperature in Oregon. *International Journal of Climatology*, *35*, 3862–3878. https://doi.org/10.1002/joc.4251

Pedersen, M., Giorgis-Allemand, L., Bernard, C., Aguilera, I., Andersen, A.M.N., Ballester, F., Beelen, R.M.J., Chatzi, L., Cirach, M., Danileviciute, A., Dedele, A., Eijsden, M. van, Estarlich, M., Fernández-Somoano, A., Fernández, M.F., Forastiere, F., Gehring, U., Grazuleviciene, R., Gruzieva, O., Heude, B., Hoek, G., de Hoogh, K., van den Hooven, E.H., Håberg, S.E., Jaddoe, V.W.V., Klümper, C., Korek, M., Krämer, U., Lertxundi, A., Lepeule, J., Nafstad, P., Nystad, W., Patelarou, E., Porta, D., Postma, D., Raaschou-Nielsen, O., Rudnai, P., Sunyer, J., Stephanou, E., Sørensen, M., Thiering, E., Tuffnell, D., Varró, M.J., Vrijkotte, T.G.M., Wijga, A., Wilhelm, M., Wright, J., Nieuwenhuijsen, M.J., Pershagen, G., Brunekreef, B., Kogevinas, M., & Slama, R. (2013). Ambient air pollution and low birthweight: A European cohort study (ESCAPE). *The Lancet Respiratory Medicine*, *1*, 695–704. https://doi.org/10.1016/S2213-2600(13)70192-9

Pedersen, M., Stayner, L., Slama, R., Sørensen, M., Figueras, F., Nieuwenhuijsen, M.J., Raaschou-Nielsen, O., & Dadvand, P. (2014). Ambient air pollution and pregnancy-induced hypertensive disorders: A systematic review and meta-analysis. *Hypertension*, *64*, 494–500. https://doi.org/10.1161/HYPERTENSIONAHA.114.03545

Pelta, R., & Chudnovsky, A.A. (2017). Spatiotemporal estimation of air temperature patterns at the street level using high resolution satellite imagery. *Science of the Total Environment*, *579*, 675–684. https://doi.org/10.1016/j.scitotenv.2016.11.042

Petit, C., Chevrier, C., Durand, G., Monfort, C., Rouget, F., Garlantezec, R., & Cordier, S. (2010). Impact on fetal growth of prenatal exposure to pesticides due to agricultural activities: a prospective cohort study in Brittany, France. *Environmental Health*, *9*, 1–12. https://doi.org/10.1186/1476-069X-9-71

Poli, P., Hersbach, H., Dee, D.P., Berrisford, P., Simmons, A.J., Vitart, F., Laloyaux, P., Tan, D.G.H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V, Bonavita, M., Isaksen, L., & Fisher, M. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, *29*, 4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1

Prihodko, L., & Goward, S.N. (1997). Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of Environment*, *60*, 335–346. https://doi.org/10.1016/S0034-4257(96)00216-7

Pu, Q., & Yoo, E.H. (2021). Ground PM2.5 prediction using imputed MAIAC AOD with uncertainty quantification. *Environmental Pollution*, *274*, 116574. https://doi.org/10.1016/j.envpol.2021.116574

Qiu, X., Fong, K.C., Shi, L., Papatheodorou, S., Di, Q., Just, A.C., Kosheleva, A., Messerlian, C., & Schwartz, J.D. (2020). Prenatal exposure to particulate air pollution and gestational age at delivery in Massachusetts neonates 2001–2015. *Environmental Epidemiology*, *4*, e113. https://doi.org/10.1097/ee9.0000000000000113

Quinn, A., Tamerius, J.D., Perzanowski, M., Jacobson, J.S., Goldstein, I., Acosta, L., & Shaman, J. (2014). Predicting indoor heat exposure risk during extreme heat events. *Science of the Total Environment*, *490*, 686–693. https://doi.org/10.1016/j.scitotenv.2014.05.039

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchistéguy, L., & Morel, S. (2008). Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France. *Journal of Applied Meteorology and Climatology*, *47*, 92–107. https://doi.org/10.1175/2007JAMC1636.1

R Core Team (2021). R: A language and environment for statistical computing.

R Core Team (2020). R: A language and environment for statistical computing.

R Core Team (2018). R: A language and environment for statistical computing.

Randazzo, T., De Cian, E., & Mistry, M.N. (2020). Air conditioning and electricity expenditure: The role of climate in temperate countries. *Economic Modelling*, *90*, 273–287. https://doi.org/10.1016/j.econmod.2020.05.001

Randles, C.A., da Silva, A.M., Buchard, V., Colarco, P.R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinozuka, Y., & Flynn, C.J. (2017). The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *Journal of Climate*, *30*, 6823–6850. https://doi.org/10.1175/JCLI-D-16-0609.1

Riviere, E., Bernard, J., Hulin, A., Virga, J., Dugay, F., Charles, M.A., Cheminat, M., Cortinovis, J., Ducroz, F., Laborie, A., Malherbe, L., Piga, D., Real, E., Robic, P.Y., Zaros, C., Seyve, E., & Lepeule, J. (2019). Air pollution modeling and exposure

assessment during pregnancy in the French Longitudinal Study of Children (ELFE). *Atmospheric Environment*, *205*, 103–114. https://doi.org/10.1016/j.atmosenv.2019.02.032

Robinson, N.P., Allred, B.W., Jones, M.O., Moreno, A., Kimball, J.S., Naugle, D.E., Erickson, T.A., & Richardson, A.D. (2017). A dynamic landsat derived normalized difference vegetation index (NDVI) product for the conterminous United States. *Remote Sensing*, *9*, 1–14. https://doi.org/10.3390/rs9080863

Rosenfeld, A., Dorman, M., Schwartz, J., Novack, V., Just, A.C., & Kloog, I. (2017). Estimating daily minimum, maximum, and mean near surface air temperature using hybrid satellite models across Israel. *Environmental Research*, *159*, 297–312. https://doi.org/10.1016/j.envres.2017.08.017

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *71*, 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Saigal, S., & Doyle, L.W. (2008). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*, *371*, 261–269. https://doi.org/10.1016/S0140-6736(08)60136-1

Sarafian, R., Kloog, I., Just, A.C., & Rosenblatt, J.D. (2019). Gaussian Markov Random Fields versus Linear Mixed Models for satellite-based PM2.5 assessment: Evidence from the Northeastern USA. *Atmospheric Environment*, *205*, 30–35. https://doi.org/10.1016/j.atmosenv.2019.02.025

Sarafian, R., Kloog, I., Sarafian, E., Hough, I., & Rosenblatt, J.D. (2020). A Domain Adaptation Approach for Performance Estimation of Spatial Predictions. *IEEE Transactions on Geoscience and Remote Sensing*, 1–9. https://doi.org/10.1109/tgrs.2020.3012575

Schifano, P., Asta, F., Dadvand, P., Davoli, M., Basagana, X., & Michelozzi, P. (2016). Heat and air pollution exposure as triggers of delivery: A survival analysis of population-based pregnancy cohorts in Rome and Barcelona. *Environment International*, *88*, 153–159. https://doi.org/10.1016/j.envint.2015.12.013

Schifano, P., Lallo, A., Asta, F., De Sario, M., Davoli, M., & Michelozzi, P. (2013). Effect of ambient temperature and air pollutants on the risk of preterm birth, Rome 2001-2010. *Environment International*, *61*, 77–87. https://doi.org/10.1016/j.envint.2013.09.005

Schneider, R., Vicedo-Cabrera, A.M., Sera, F., Masselot, P., Stafoggia, M., de Hoogh, K., Kloog, I., Reis, S., Vieno, M., & Gasparrini, A. (2020). A satellite-based spatio-temporal machine learning model to reconstruct daily PM2.5 concentrations across great britain. *Remote Sensing*, *12*, 1–19. https://doi.org/10.3390/rs12223803

Sellier, Y., Galineau, J., Hulin, A., Caini, F., Marquis, N., Navel, V., Bottagisi, S., Giorgis-Allemand, L., Jacquier, C., Slama, R., & Lepeule, J. (2014). Health effects of ambient air pollution: Do different methods for estimating exposure lead to different results?. *Environment International*, *66*, 165–173. https://doi.org/10.1016/j.envint.2014.02.001

Shao, Y., Ma, Z., Wang, J., & Bi, J. (2020). Estimating daily ground-level PM2.5 in China with random-forest-based spatiotemporal kriging. *Science of the Total Environment*, *740*, 139761. https://doi.org/10.1016/j.scitotenv.2020.139761

Shashar, S., Kloog, I., Erez, O., Shtein, A., Yitshak-Sade, M., Sarov, B., & Novack, L. (2020). Temperature and preeclampsia: Epidemiological evidence that perturbation in maternal heat homeostasis affects pregnancy outcome. *PLoS ONE*, *15*, 1–14. https://doi.org/10.1371/journal.pone.0232877

Shi, L., Kloog, I., Zanobetti, A., Liu, P., & Schwartz, J. (2015). Impacts of temperature and its variability on mortality in New England. *Nature Climate Change*, *5*, 988–991. https://doi.org/10.1038/nclimate2704

Shi, L., Liu, P., Kloog, I., Lee, M., Kosheleva, A., & Schwartz, J. (2016). Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: A statistical modeling study. *Environmental Research*, *146*, 51–58. https://doi.org/10.1016/j.envres.2015.12.006

Shiff, S., Helman, D., & Lensky, I.M. (2021). Worldwide continuous gap-filled MODIS land surface temperature dataset. *Scientific Data*, *8*, 1–10. https://doi.org/10.1038/s41597-021-00861-7

Shtein, A., Karnieli, A., Katra, I., Raz, R., Levy, I., Lyapustin, A., Dorman, M., Broday, D.M., & Kloog, I. (2018). Estimating daily and intra-daily PM10 and PM2.5 in Israel using a spatio-temporal hybrid modeling approach. *Atmospheric Environment*, *191*, 142–152. https://doi.org/10.1016/j.atmosenv.2018.08.002

Shtein, A., Kloog, I., Schwartz, J., Silibello, C., Michelozzi, P., Gariazzo, C., Viegi, G., Forastiere, F., Karnieli, A., Just, A.C., & Stafoggia, M. (2019). Estimating Daily PM2.5 and PM10 over Italy Using an Ensemble Model. *Environmental Science & Technology*, *54*, 120–128. https://doi.org/10.1021/acs.est.9b04279

Simpson, D., Rue, H., Riebler, A., Martins, T.G., & Sørbye, S.H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*, 1–28. https://doi.org/10.1214/16-STS576

Sinclair, K.D., Allegrucci, C., Singh, R., Gardner, D.S., Sebastian, S., Bispham, J., Thurston, A., Huntley, J.F., Rees, W.D., Maloney, C.A., Lea, R.G., Craigon, J., Mcevoy, T.G., & Young, L.E. (2007). DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status BIOLOGY. *Proceedings of the National Academy of Sciences*, *104*, 19351–19356. https://doi.org/10.1073/pnas.0707258104

Slama, R., Darrow, L.A., Parker, J., Woodruff, T.J., Strickland, M.J., Nieuwenhuijsen, M.J., Glinianaia, S., Hoggatt, K.J., Kannan, S., Hurley, F., Kalinka, J., Šrám, R., Brauer, M., Wilhelm, M., Henrich, J., & Ritz, B. (2008). Meeting report: Atmospheric pollution and human reproduction. *Environmental Health Perspectives*, *116*, 791–798. https://doi.org/10.1289/ehp.11074

Song, W., Jia, H., Huang, J., & Zhang, Y. (2014). A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China. *Remote Sensing of Environment*, *154*, 1–7. https://doi.org/10.1016/j.rse.2014.08.008

Song, X., Wang, S., Hu, Y., Yue, M., Zhang, T., Liu, Y., Tian, J., & Shang, K. (2017). Impact of ambient temperature on morbidity and mortality: An overview of reviews. *Science of the Total Environment*, *586*, 241–254. https://doi.org/10.1016/j.scitotenv.2017.01.212

Spolter, F., Kloog, I., Dorman, M., Novack, L., Erez, O., & Raz, R. (2020). Prenatal exposure to ambient air temperature and risk of early delivery. *Environment International*, *142*, 1–5. https://doi.org/10.1016/j.envint.2020.105824

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., Donato, F. De, Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., & Schwartz, J. (2019). Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environment International*, *124*, 170–179. https://doi.org/10.1016/j.envint.2019.01.016

Stafoggia, M., Johansson, C., Glantz, P., Renzi, M., Shtein, A., de Hoogh, K., Kloog, I., Davoli, M., Michelozzi, P., & Bellander, T. (2020). A random forest approach to estimate daily particulate matter, nitrogen dioxide, and ozone at fine spatial resolution in Sweden. *Atmosphere*, *11*. https://doi.org/10.3390/atmos11030239

Stafoggia, M., Schwartz, J., Badaloni, C., Bellander, T., Alessandrini, E., Cattani, G., De'Donato, F.K., Gaeta, A., Leone, G., Lyapustin, A., Sorek-Hamer, M., de Hoogh, K., Di, Q., Forastiere, F., & Kloog, I. (2017). Estimation of daily PM10 concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. *Environment International*, *99*, 234–244. https://doi.org/10.1016/j.envint.2016.11.024

Stan, C.M., Boulvain, M., Pfister, R., & Hirsbrunner-Almagbaly, P. (2013). Hydration for treatment of preterm labour. *Cochrane Database of Systematic Reviews*, CD003096. https://doi.org/10.1002/14651858.CD003096.pub2

Stieb, D.M., Chen, L., Beckerman, B.S., Jerrett, M., Crouse, D.L., Omariba, D.W.R., Peters, P.A., van Donkelaar, A., Martin, R. v, Burnett, R.T., Gilbert, N.L., Tjepkema, M., Liu, S., & Dugandzic, R.M. (2016). Associations of Pregnancy Outcomes and PM in a National Canadian Study. *Environmental Health Perspectives*, *124*, 243–249. https://doi.org/10.1289/ehp.1408995

Strand, L.B., Barnett, A.G., & Tong, S. (2011a). The influence of season and ambient temperature on birth outcomes: A review of the epidemiological literature. *Environmental Research*, *111*, 451–462. https://doi.org/10.1016/j.envres.2011.01.023

Strand, L.B., Barnett, A.G., & Tong, S. (2011b). Methodological challenges when estimating the effects of season and seasonal exposures on birth outcomes. *BMC Medical Research Methodology*, *11*. https://doi.org/10.1186/1471-2288-11-49

Strobl, C., Boulesteix, A.L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*. https://doi.org/10.1186/1471-2105-8-25

Sun, J., Gong, J., & Zhou, J. (2021). Estimating hourly PM2.5 concentrations in Beijing with satellite aerosol optical depth and a random forest approach. *Science of the Total Environment*, *762*, 144502. https://doi.org/10.1016/j.scitotenv.2020.144502

Sun, S., Weinberger, K.R., Spangler, K.R., Eliot, M.N., Braun, J.M., & Wellenius, G.A. (2019). Ambient temperature and preterm birth: A retrospective study of 32 million US singleton births. *Environment International*, *126*, 7–13. https://doi.org/10.1016/j.envint.2019.02.023

Sun, Y., Ilango, S.D., Schwarz, L., Wang, Q., Chen, J.C., Lawrence, J.M., Wu, J., & Benmarhnia, T. (2020a). Examining the joint effects of heatwaves, air pollution, and green space on the risk of preterm birth in California. *Environmental Research Letters*, *15*. https://doi.org/10.1088/1748-9326/abb8a3

Sun, Y., Sheridan, P., Laurent, O., Li, J., Sacks, D.A., Fischer, H., Qiu, Y., Jiang, Y., Yim, I.S., Jiang, L.H., Molitor, J., Chen, J.C., Benmarhnia, T., Lawrence, J.M., & Wu, J. (2020b). Associations between green space and preterm birth: Windows of susceptibility and interaction with air pollution. *Environment International*, *142*, 105804. https://doi.org/10.1016/j.envint.2020.105804

Therneau, T.M. (2021). A Package for Survival Analysis in R.

Tøttrup, C. (2014). EU-DEM Statistical Validation Report.

UN DESA Population Division (2018). World Urbanization Prospects: The 2018 Revision.

USGS (2018a). Landsat 4-7 Surface Reflectance (LEDAPS) Product Guide.

USGS (2018b). Landsat 8 Surface Reflectance Code (LASRC) Product Guide.

Uwak, I., Olson, N., Fuentes, A., Moriarty, M., Pulczinski, J., Lam, J., Xu, X., Taylor, B.D., Taiwo, S., Koehler, K., Foster, M., Chiu, W.A., & Johnson, N.M. (2021). Application of the navigation guide systematic review methodology to evaluate prenatal exposure to particulate matter air pollution and infant birth weight. *Environment International*, *148*, 106378. https://doi.org/10.1016/j.envint.2021.106378

Van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., Lyapustin, A., Sayer, A.M., & Winker, D.M. (2016). Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. *Environmental Science & Technology*, *50*, 3762–3772. https://doi.org/10.1021/acs.est.5b05833

Vancutsem, C., Ceccato, P., Dinku, T., & Connor, S.J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, *114*, 449–465. https://doi.org/10.1016/j.rse.2009.10.002

Vicedo-Cabrera, A.M., Iñíguez, C., Barona, C., & Ballester, F. (2014). Exposure to elevated temperatures and risk of preterm birth in Valencia, Spain. *Environmental Research*, *134*, 210–217. https://doi.org/10.1016/j.envres.2014.07.021

Vicedo-Cabrera, A.M., Olsson, D., & Forsberg, B. (2015). Exposure to seasonal temperatures during the last month of gestation and the risk of preterm birth in stockholm. *International Journal of Environmental Research and Public Health*, *12*, 3962–3978. https://doi.org/10.3390/ijerph120403962

Voogt, J.A., & Oke, T.R. (1997). Complete Urban Surface Temperatures. *Journal of Applied Meteorology*, *36*, 1117–1132. https://doi.org/10.1175/1520-0450(1997)036<1117:CUST>2.0.CO;2

Wan, Z. (2014). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, *140*, 36–45. https://doi.org/10.1016/j.rse.2013.08.027

Wang, F., Qin, Z., Song, C., Tu, L., Karnieli, A., & Zhao, S. (2015). An Improved Mono-Window Algorithm for Land Surface Temperature Retrieval from Landsat 8 Thermal Infrared Sensor Data. *Remote Sensing*, *7*, 4268–4289. https://doi.org/10.3390/rs70404268

Wang, Q., Li, B., Benmarhnia, T., Hajat, S., Ren, M., Liu, T., Knibbs, L.D., Zhang, H., Bao, J., Zhang, Y., Zhao, Q., & Huang, C. (2020). Independent and combined effects of heatwaves and PM2:5 on preterm birth in Guangzhou, China: A survival analysis. *Environmental Health Perspectives*, *128*, 1–10. https://doi.org/10.1289/EHP5117

Wang, Y., Nordio, F., Nairn, J., Zanobetti, A., & Schwartz, J. (2018). Accounting for adaptation and intensity in projecting heat wave-related mortality. *Environmental Research*, *161*, 464–471. https://doi.org/10.1016/j.envres.2017.11.049

Wang, Y.-Y., Li, Q., Guo, Y., Zhou, H., Wang, Q.M., Shen, H.P., Zhang, Y.P., Yan, D.H., Li, S., Chen, G., Zhou, S., He, Y., Yang, Y., Peng, Z.Q., Wang, H.J., & Ma, X. (2020). Ambient temperature and the risk of preterm birth: A national birth cohort study in the mainland China. *Environment International*, *142*. https://doi.org/10.1016/j.envint.2020.105851

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., & Cribb, M. (2019). Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. *Remote Sensing of Environment*, *231*, 111221. https://doi.org/10.1016/j.rse.2019.111221

Wei, J., Li, Z., Cribb, M., Huang, W., Xue, W., Sun, L., Guo, J., Peng, Y., Li, J., Lyapustin, A., Liu, L., Wu, H., & Song, Y. (2020). Improved 1 km resolution PM2.5 estimates across China using enhanced space–time extremely randomized trees. *Atmospheric Chemistry and Physics*, *20*, 3273–3289. https://doi.org/10.5194/acp-20-3273-2020

Weinberg, C.R., Shi, M., DeRoo, L.A., Basso, O., & Skjærven, R. (2015). Season and preterm birth in Norway: A cautionary tale. *International Journal of Epidemiology*, *44*, 1068–1078. https://doi.org/10.1093/ije/dyv100

Weiss, D.J., Bhatt, S., Mappin, B., Van Boeckel, T.P., Smith, D.L., Hay, S.I., & Gething, P.W. (2014). Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: A high-resolution spatiotemporal prediction. *Malaria Journal*, *13*, 1–11. https://doi.org/10.1186/1475-2875-13-171

Wicki, A., Parlow, E., & Feigenwinter, C. (2018). Evaluation and Modeling of Urban Heat Island Intensity in Basel, Switzerland. *Climate*, *6*, 1–25. https://doi.org/10.3390/cli6030055

Wilson, A., Chiu, Y.H.M., Hsu, H.H.L., Wright, R.O., Wright, R.J., & Coull, B.A. (2017). Potential for Bias When Estimating Critical Windows for Air Pollution in Children's Health. *American Journal of Epidemiology*, *186*, 1281–1289. https://doi.org/10.1093/aje/kwx184

Wood, S.N. (2017). Generalized Additive Models: An Introduction with R, 2nd ed. CRC Press.

World Health Organization (2021). WHO global air quality guidelines. Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva.

Wright, M.N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*. https://doi.org/10.18637/jss.v077.i01

Wu, J., Yao, F., Li, W., & Si, M. (2016). VIIRS-based remote sensing estimation of ground-level PM2.5 concentrations in Beijing–Tianjin–Hebei: A spatiotemporal statistical model. *Remote Sensing of Environment*, *184*, 316–328. https://doi.org/10.1016/j.rse.2016.07.015

Wu, M., Song, L., Zheng, X., Zhang, L., Liu, B., Wang, L., Li, H., Xiong, C., Cao, Z., Wang, Y., & Xu, S. (2019). Prenatal exposure of diurnal temperature range and preterm birth: Findings from a birth cohort study in China. *Science of the Total Environment*, *656*, 1102–1107. https://doi.org/10.1016/j.scitotenv.2018.11.305

Xiao, Q., Geng, G., Liang, F., Wang, X., Lv, Z., Lei, Y., Huang, X., Zhang, Q., Liu, Y., & He, K. (2020). Changes in spatial patterns of PM2.5 pollution in China 2000-2018:

Impact of clean air policies. *Environment international*, *141*, 105776. https://doi.org/10.1016/j.envint.2020.105776

Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., & Liu, Y. (2017). Full-coverage high-resolution daily PM2.5 estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sensing of Environment*, *199*, 437–446. https://doi.org/10.1016/j.rse.2017.07.023

Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B., & Bai, Y. (2015). Daily Estimation of Ground-Level PM2.5 Concentrations over Beijing Using 3 km Resolution MODIS AOD. *Environmental Science & Technology*, *49*, 12280–12288. https://doi.org/10.1021/acs.est.5b01413

Xue, W., Zhang, J., Zhong, C., Ji, D., & Huang, W. (2020). Satellite-derived spatiotemporal PM2.5 concentrations and variations from 2006 to 2017 in China. *Science of the Total Environment*, *712*, 134577. https://doi.org/10.1016/j.scitotenv.2019.134577

Yan, X., Zang, Z., Luo, N., Jiang, Y., & Li, Z. (2020). New interpretable deep learning model to monitor real-time PM2.5 concentrations from satellite data. *Environment International*, *144*, 106060. https://doi.org/10.1016/j.envint.2020.106060

Yitshak-Sade, M., Fabian, M.P., Lane, K.J., Hart, J.E., Schwartz, J.D., Laden, F., James, P., Fong, K.C., Kloog, I., & Zanobetti, A. (2020). Estimating the combined effects of natural and built environmental exposures on birthweight among urban residents in massachusetts. *International Journal of Environmental Research and Public Health*, *17*, 1–16. https://doi.org/10.3390/ijerph17238805

Yitshak-Sade, M., Kloog, I., Schwartz, J.D., Novack, V., Erez, O., & Just, A.C. (2021). The effect of prenatal temperature and PM2.5 exposure on birthweight: Weekly windows of exposure throughout the pregnancy. *Environment International*, *155*, 106588. https://doi.org/10.1016/j.envint.2021.106588

Yoo, C., Im, J., Park, S., & Quackenbush, L.J. (2018). Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *137*, 149–162. https://doi.org/10.1016/j.isprsjprs.2018.01.018

You, W., Zang, Z., Zhang, L., Li, Z., Chen, D., & Zhang, G. (2015). Estimating ground-level PM10 concentration in northwestern China using geographically weighted regression based on satellite AOD combined with CALIPSO and MODIS fire count. *Remote Sensing of Environment*, *168*, 276–285. https://doi.org/10.1016/j.rse.2015.07.020

Yu, X., Guo, X., & Wu, Z. (2014). Land surface temperature retrieval from landsat 8 TIRS-comparison between radiative transfer equation-based method, split window

algorithm and single channel method. *Remote Sensing*, *6*, 9829–9852. https://doi.org/10.3390/rs6109829

Zeger, S.L., Thomas, D., Dominici, F., Samet, J.M., Schwartz, J., Dockery, D., & Cohen, A.J. (2000). Exposure measurement error in time-series studies of air pollution: Concepts and consequences. *Environmental Health Perspectives*, *108*, 419–426. https://doi.org/10.1289/ehp.00108419

Zeng, C., Long, D., Shen, H., Wu, P., Cui, Y., & Hong, Y. (2018). A two-step framework for reconstructing remotely sensed land surface temperatures contaminated by cloud. *ISPRS Journal of Photogrammetry and Remote Sensing*, *141*, 30–45. https://doi.org/10.1016/j.isprsjprs.2018.04.005

Zhai, B., & Chen, J. (2018). Development of a stacked ensemble model for forecasting and analyzing daily average PM2.5 concentrations in Beijing, China. *Science of the Total Environment*, *635*, 644–658. https://doi.org/10.1016/j.scitotenv.2018.04.040

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012a). Real-time air quality forecasting, part I: History, techniques, and current status. *Atmospheric Environment*, *60*, 632–655. https://doi.org/10.1016/j.atmosenv.2012.06.031

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012b). Real-time air quality forecasting, Part II: State of the science, current research needs, and future prospects. *Atmospheric Environment*, *60*, 656–676. https://doi.org/10.1016/j.atmosenv.2012.02.041

Zhang, Y., Yu, C., & Wang, L. (2017). Temperature exposure during pregnancy and birth outcomes: An updated systematic review of epidemiological evidence. *Environmental Pollution*, *225*, 700–712. https://doi.org/10.1016/j.envpol.2017.02.066

Zhang, Z., Wang, J., Hart, J.E., Laden, F., Zhao, C., Li, T., Zheng, P., Li, D., Ye, Z., & Chen, K. (2018). National scale spatiotemporal land-use regression model for PM2.5, PM10 and NO2 concentration in China. *Atmospheric Environment*, *192*, 48–54. https://doi.org/10.1016/j.atmosenv.2018.08.046

Zheng, Y., Zhang, Q., Liu, Y., Geng, G., & He, K. (2016). Estimating ground-level PM2.5 concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. *Atmospheric Environment*, *124*, 232–242. https://doi.org/10.1016/j.atmosenv.2015.06.046

Zhong, Q., Lu, C., Zhang, W., Zheng, X., & Deng, Q. (2018). Preterm birth and ambient temperature: Strong association during night-time and warm seasons. *Journal of Thermal Biology*, *78*, 381–390. https://doi.org/10.1016/j.jtherbio.2018.11.002

Zhou, B., Erell, E., Hough, I., Shtein, A., Just, A.C., Novack, V., Rosenblatt, J., & Kloog, I. (2020). Estimation of Hourly near Surface Air Temperature Across Israel Using an Ensemble Model. *Remote Sensing*, *12*. https://doi.org/10.3390/rs12111741

Zhu, W., Lű, A., & Jia, S. (2013). Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of Environment*, *130*, 62–73. https://doi.org/10.1016/j.rse.2012.10.034

Zhu, W., Lű, A., Jia, S., Yan, J., & Mahmood, R. (2017). Retrievals of all-weather daytime air temperature from MODIS products. *Remote Sensing of Environment*, *189*, 152–163. https://doi.org/10.1016/j.rse.2016.11.011

Appendix 1: supplement to Chapter 1

Table A1-1. Daily Ta observations at weather stations during the 19-year study period.

	N	Min	Mean	Max	SD*
T_{min}	14 725 113	-31.2	6.8	30.3	6.5
T_{mean}	8 737 237	-28.2	11.3	34.4	7.1
T_{max}	14 725 428	-26.0	16.5	44.1	8.3

* SD = standard deviation

Table A1-2. Aggregations of Corine Land Cover (CLC) classes used in this study.

Aggregated category	CLC codes	CLC class descriptions
Artificial	1	Artificial areas
Vegetation	2	Agricultural areas
	3.1	Forests
	3.2	Shrubs and/or herbaceous vegetation associations
Bare	3.3	Open spaces with little or no vegetation
Water	4	Wetlands
	5	Water bodies

Table A1-3. Stage 1 T_{mean} model performance (predicting daily 1 km T_{mean} from LST): 10-fold cross-validated performance by year; overall, spatial, and temporal components.

T_{mean}			Overall			Spatial			Temporal		
Year	LST*	N†	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
2000	TN	153	0.96	1.20	0.87	0.94	1.14	0.80	0.96	1.00	0.73
2001	TN	173	0.97	1.24	0.90	0.96	1.13	0.79	0.98	1.06	0.77
2002	TN	171	0.96	1.25	0.90	0.94	1.17	0.80	0.96	1.08	0.78
2003	TN	204	0.98	1.27	0.94	0.96	1.19	0.84	0.98	1.10	0.81
2004	TN	196	0.97	1.27	0.92	0.95	1.16	0.81	0.97	1.12	0.80
2005	TN	222	0.97	1.26	0.92	0.96	1.13	0.79	0.98	1.11	0.80
2006	TN	205	0.97	1.29	0.93	0.96	1.17	0.82	0.98	1.13	0.81
2007	TN	225	0.96	1.28	0.93	0.94	1.20	0.82	0.97	1.11	0.80
2008	TN	215	0.96	1.27	0.92	0.94	1.17	0.83	0.97	1.09	0.79
2009	TN	232	0.97	1.28	0.93	0.96	1.19	0.85	0.98	1.08	0.79
2010	TN	209	0.97	1.25	0.90	0.96	1.19	0.84	0.98	1.04	0.76
2011	TN	239	0.96	1.35	0.99	0.94	1.19	0.84	0.96	1.16	0.85
2012	TN	224	0.97	1.35	0.98	0.96	1.22	0.86	0.97	1.16	0.85
2013	TN	203	0.97	1.37	0.98	0.95	1.22	0.86	0.97	1.18	0.84
2014	TN	201	0.96	1.24	0.90	0.94	1.13	0.80	0.96	1.05	0.76
2015	TN	215	0.96	1.36	0.99	0.95	1.22	0.87	0.97	1.18	0.86
2016	TN	205	0.96	1.38	1.00	0.94	1.26	0.90	0.97	1.19	0.86
2017	TN	199	0.97	1.30	0.96	0.96	1.14	0.82	0.97	1.14	0.84
2018	TN	194	0.97	1.25	0.92	0.96	1.09	0.79	0.98	1.09	0.80

* LST = source of LST; TN = Terra night

† N = Thousands of observations used to fit model

Table A1-4. Stage 1 T_{min} model performance (predicting daily 1 km T_{min} from LST): 10-fold cross-validated performance by year; overall, spatial, and temporal components.

T_{min}			Overall			Spatial			Temporal		
Year	LST*	N†	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
2000	TN	299	0.87	1.92	1.47	0.86	1.54	1.14	0.88	1.65	1.27
2001	TN	332	0.92	1.88	1.43	0.91	1.50	1.10	0.93	1.65	1.25
2002	TN	323	0.88	1.99	1.52	0.87	1.57	1.15	0.89	1.74	1.33
2003	AN	405	0.94	1.88	1.41	0.91	1.67	1.21	0.96	1.50	1.10
2004	AN	367	0.92	1.86	1.40	0.90	1.59	1.16	0.94	1.53	1.12
2005	AN	398	0.94	1.89	1.42	0.91	1.65	1.20	0.95	1.53	1.12
2006	AN	365	0.94	1.84	1.38	0.91	1.58	1.15	0.95	1.49	1.09
2007	AN	385	0.91	1.88	1.41	0.89	1.60	1.17	0.93	1.52	1.11
2008	AN	358	0.91	1.85	1.39	0.89	1.56	1.15	0.93	1.50	1.10
2009	AN	386	0.93	1.86	1.41	0.90	1.63	1.20	0.95	1.49	1.09
2010	AN	347	0.93	1.84	1.38	0.92	1.60	1.18	0.95	1.48	1.08
2011	AN	392	0.90	1.95	1.48	0.87	1.67	1.24	0.92	1.54	1.13
2012	AN	362	0.93	1.92	1.45	0.91	1.61	1.19	0.95	1.56	1.15
2013	AN	322	0.93	1.87	1.39	0.91	1.56	1.15	0.94	1.53	1.11
2014	AN	324	0.89	1.82	1.37	0.88	1.52	1.12	0.92	1.46	1.07
2015	AN	336	0.91	1.95	1.47	0.88	1.68	1.24	0.93	1.57	1.14
2016	AN	316	0.91	1.94	1.45	0.88	1.66	1.22	0.93	1.55	1.13
2017	AN	303	0.92	1.91	1.45	0.90	1.62	1.19	0.94	1.54	1.14
2018	AN	299	0.93	1.79	1.36	0.90	1.53	1.13	0.94	1.45	1.07

* LST = source of LST; TN = Terra night; AN = Aqua night

† N = Thousands of observations used to fit the model

Table A1-5. Stage 1 T_{max} model performance (predicting daily 1 km T_{max} from LST): 10-fold cross-validated performance by year; overall, spatial, and temporal components.

T_{max}			Overall			Spatial			Temporal		
Year	LST*	N†	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
2000	TD	265	0.94	1.78	1.33	0.90	1.56	1.13	0.95	1.38	1.02
2001	TD	319	0.96	1.80	1.34	0.92	1.54	1.12	0.97	1.46	1.07
2002	TD	314	0.94	1.83	1.37	0.90	1.56	1.14	0.95	1.46	1.08
2003	AD	379	0.97	1.84	1.37	0.94	1.59	1.16	0.97	1.49	1.10
2004	AD	334	0.95	1.79	1.33	0.93	1.51	1.09	0.97	1.45	1.06
2005	AD	358	0.96	1.77	1.32	0.94	1.52	1.09	0.97	1.44	1.06
2006	AD	337	0.96	1.86	1.38	0.92	1.59	1.16	0.97	1.52	1.11
2007	AD	353	0.95	1.79	1.34	0.91	1.55	1.11	0.96	1.45	1.07
2008	AD	318	0.95	1.77	1.32	0.91	1.52	1.11	0.96	1.41	1.04
2009	AD	341	0.96	1.83	1.37	0.92	1.58	1.15	0.97	1.44	1.06
2010	AD	308	0.96	1.77	1.32	0.93	1.56	1.13	0.97	1.38	1.01
2011	AD	358	0.94	1.82	1.37	0.91	1.62	1.18	0.96	1.45	1.07
2012	AD	332	0.96	1.83	1.37	0.92	1.61	1.18	0.97	1.46	1.08
2013	AD	291	0.96	1.86	1.38	0.92	1.62	1.17	0.97	1.49	1.09
2014	AD	300	0.94	1.73	1.29	0.91	1.51	1.11	0.95	1.36	1.01
2015	AD	315	0.95	1.86	1.39	0.91	1.61	1.18	0.96	1.51	1.12
2016	AD	290	0.95	1.82	1.36	0.91	1.60	1.17	0.96	1.47	1.08
2017	AD	274	0.96	1.81	1.36	0.93	1.53	1.12	0.97	1.45	1.08
2018	AD	271	0.96	1.73	1.31	0.93	1.50	1.12	0.97	1.37	1.02

* LST = source of LST; TD = Terra day

† N = thousands of observations used to fit model

Table A1-6. Stage 1 model performance (predicting daily 1 km T_a from LST): 10-fold cross-validated performance across all years (2000 to 2016), by climatic region and urban *vs.* rural locations.

	T_{min}			T_{mean}			T_{max}		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Mountain	0.90	2.21	1.70	0.95	1.67	1.24	0.94	2.24	1.72
Semi-continental	0.91	2.10	1.61	0.96	1.44	1.06	0.95	1.99	1.51
Modified oceanic	0.94	1.53	1.16	0.98	0.98	0.73	0.98	1.33	1.01
Transitional oceanic	0.92	1.81	1.37	0.97	1.20	0.88	0.95	1.74	1.31
Oceanic	0.90	1.78	1.32	0.96	1.20	0.88	0.94	1.83	1.36
Mod. Mediterranean	0.90	2.21	1.71	0.96	1.43	1.08	0.94	2.03	1.55
Southwest basin	0.94	1.60	1.23	0.98	1.04	0.76	0.97	1.41	1.04
Mediterranean	0.93	1.81	1.40	0.98	1.11	0.85	0.96	1.62	1.24
Urban	0.93	1.84	1.35	0.97	1.32	0.96	0.95	1.79	1.35
Peri-urban*	0.93	1.71	1.28	0.97	1.18	0.87	0.96	1.71	1.27
Rural	0.92	1.90	1.44	0.97	1.30	0.95	0.95	1.81	1.36

* Non-urban locations within 5 km of a large urban area

Table A1-7. Stage 4 model performance (predicting daily 200 m residuals with an ensemble): 10-fold cross-validated performance across all years (2000 to 2016), by climatic region and urban *vs.* rural locations (residual scale).

	R_{Tmin}			R_{Tmean}			R_{Tmax}		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Mountain	0.83	0.67	0.42	0.83	0.46	0.30	0.87	0.58	0.37
Semi-continental	0.80	0.67	0.42	0.79	0.44	0.28	0.86	0.55	0.34
Modified oceanic	0.74	0.55	0.34	0.76	0.33	0.21	0.80	0.40	0.23
Transitional oceanic	0.77	0.63	0.40	0.78	0.39	0.26	0.84	0.51	0.31
Oceanic	0.74	0.62	0.40	0.77	0.39	0.26	0.83	0.51	0.30
Mod. Mediterranean	0.82	0.73	0.48	0.78	0.48	0.31	0.84	0.62	0.42
Southwest basin	0.75	0.60	0.37	0.68	0.39	0.24	0.78	0.48	0.29
Mediterranean	0.77	0.67	0.44	0.72	0.43	0.28	0.80	0.58	0.39
Urban	0.78	0.54	0.33	0.81	0.37	0.23	0.84	0.47	0.27
Peri-urban*	0.76	0.58	0.37	0.78	0.37	0.24	0.83	0.47	0.28
Rural	0.79	0.64	0.40	0.79	0.41	0.27	0.84	0.52	0.32

* Non-urban locations within 5 km of a large urban area

Table A1-8. Stage 4 model performance (predicting daily 200 m residuals with an ensemble): 10-fold cross-validated performance by year.

	R_{Tmin}				R_{Tmean}				R_{Tmax}			
	N*	R^2	RMSE	MAE	N*	R^2	RMSE	MAE	N*	R^2	RMSE	MAE
2000	842	0.88	0.47	0.30	425	0.80	0.36	0.22	842	0.87	0.46	0.26
2001	834	0.85	0.49	0.32	427	0.75	0.41	0.25	834	0.83	0.52	0.31
2002	829	0.87	0.50	0.33	431	0.77	0.39	0.24	829	0.85	0.51	0.30
2003	824	0.78	0.66	0.43	431	0.79	0.42	0.28	824	0.84	0.54	0.34
2004	829	0.77	0.64	0.40	447	0.91	0.25	0.16	829	0.83	0.53	0.31
2005	825	0.75	0.70	0.44	467	0.75	0.44	0.28	825	0.85	0.50	0.31
2006	815	0.74	0.68	0.41	471	0.76	0.42	0.27	815	0.84	0.53	0.33
2007	817	0.79	0.64	0.41	480	0.77	0.42	0.28	817	0.83	0.53	0.32
2008	810	0.78	0.63	0.39	486	0.77	0.42	0.27	810	0.84	0.50	0.30
2009	803	0.78	0.66	0.42	488	0.79	0.42	0.28	803	0.85	0.52	0.32
2010	801	0.77	0.65	0.39	490	0.77	0.41	0.25	801	0.86	0.49	0.29
2011	793	0.80	0.67	0.43	487	0.79	0.43	0.29	793	0.84	0.54	0.34
2012	776	0.78	0.66	0.42	482	0.80	0.43	0.28	776	0.84	0.54	0.33
2013	748	0.76	0.65	0.40	476	0.85	0.35	0.22	748	0.87	0.48	0.29
2014	733	0.80	0.60	0.38	470	0.78	0.40	0.25	733	0.83	0.52	0.31
2015	709	0.79	0.66	0.43	461	0.77	0.45	0.29	709	0.86	0.52	0.33
2016	692	0.78	0.67	0.42	458	0.78	0.43	0.28	692	0.84	0.52	0.32
2017	617	0.75	0.71	0.45	408	0.76	0.44	0.29	618	0.82	0.56	0.34
2018	609	0.74	0.68	0.42	412	0.75	0.43	0.27	610	0.82	0.54	0.32

* N = thousands of observations used to fit model

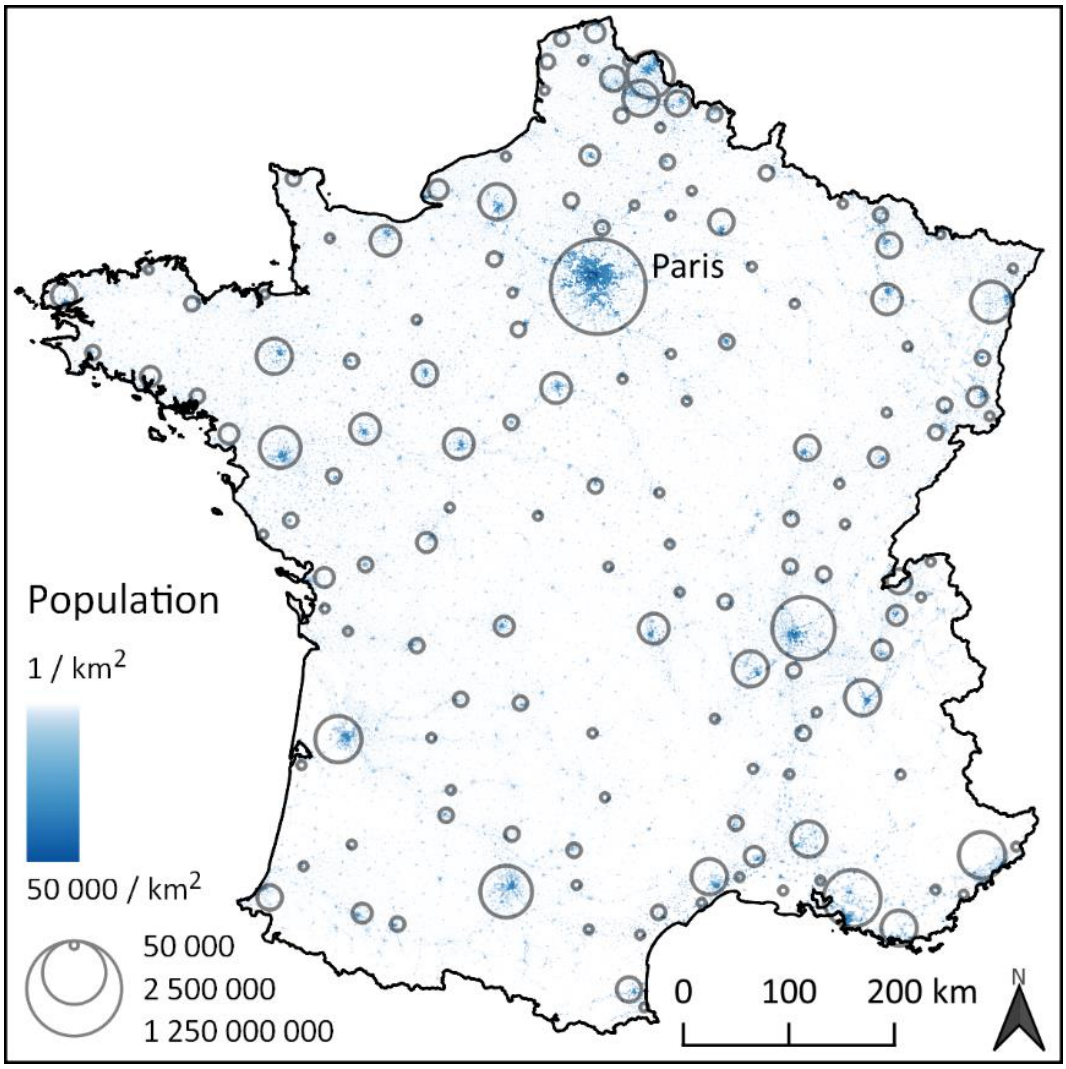

Figure A1-1. Population density of France and urban areas with at least 50 000 residents.

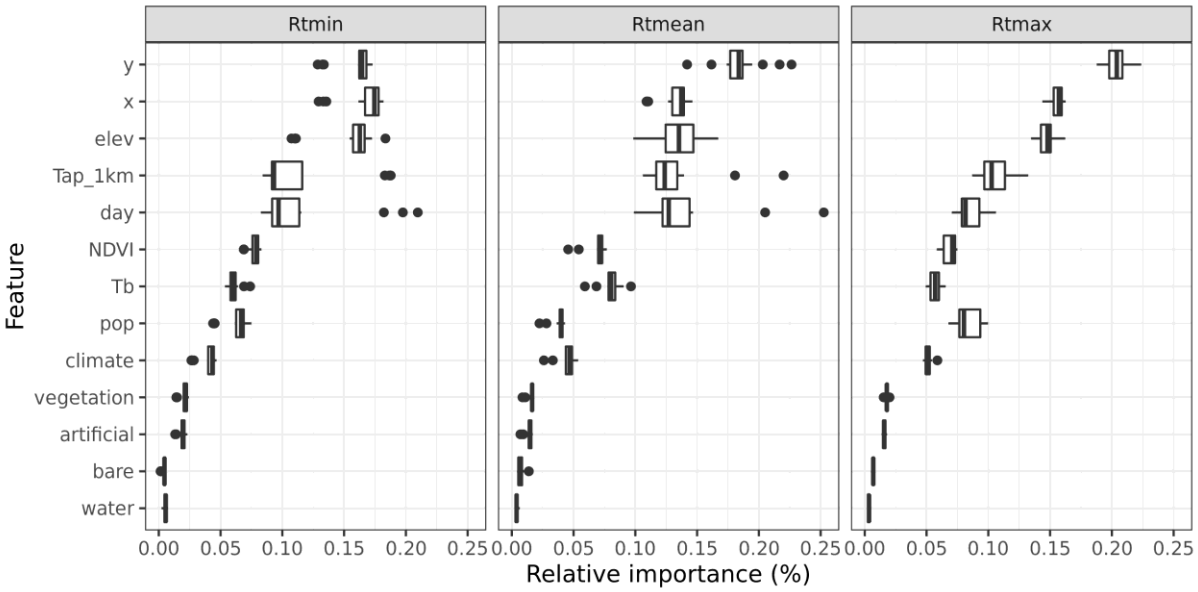

Figure A1-2. Relative importance (%) of the predictors in the stage 3 random forest model (predicting 200 m residual). Each box shows the distribution for the different model years (2000 to 2016).

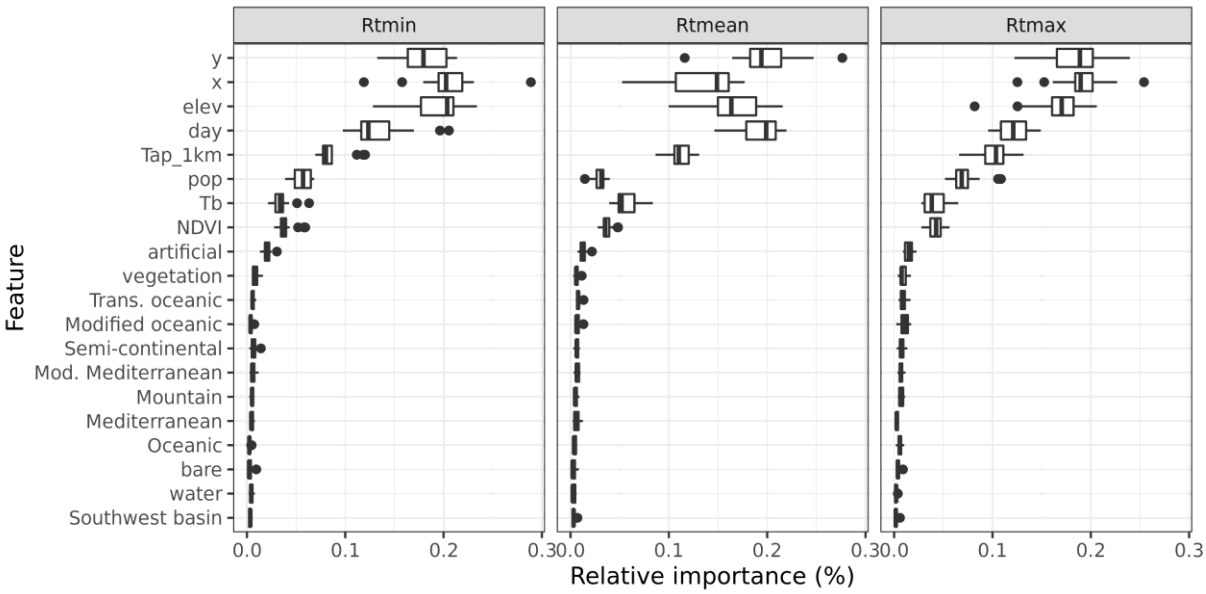

Figure A1-3. Relative importance (%) of the predictors in the stage 3 XGBoost model (predicting 200 m residual). Each box shows the distribution for the different model years (2000 to 2016).

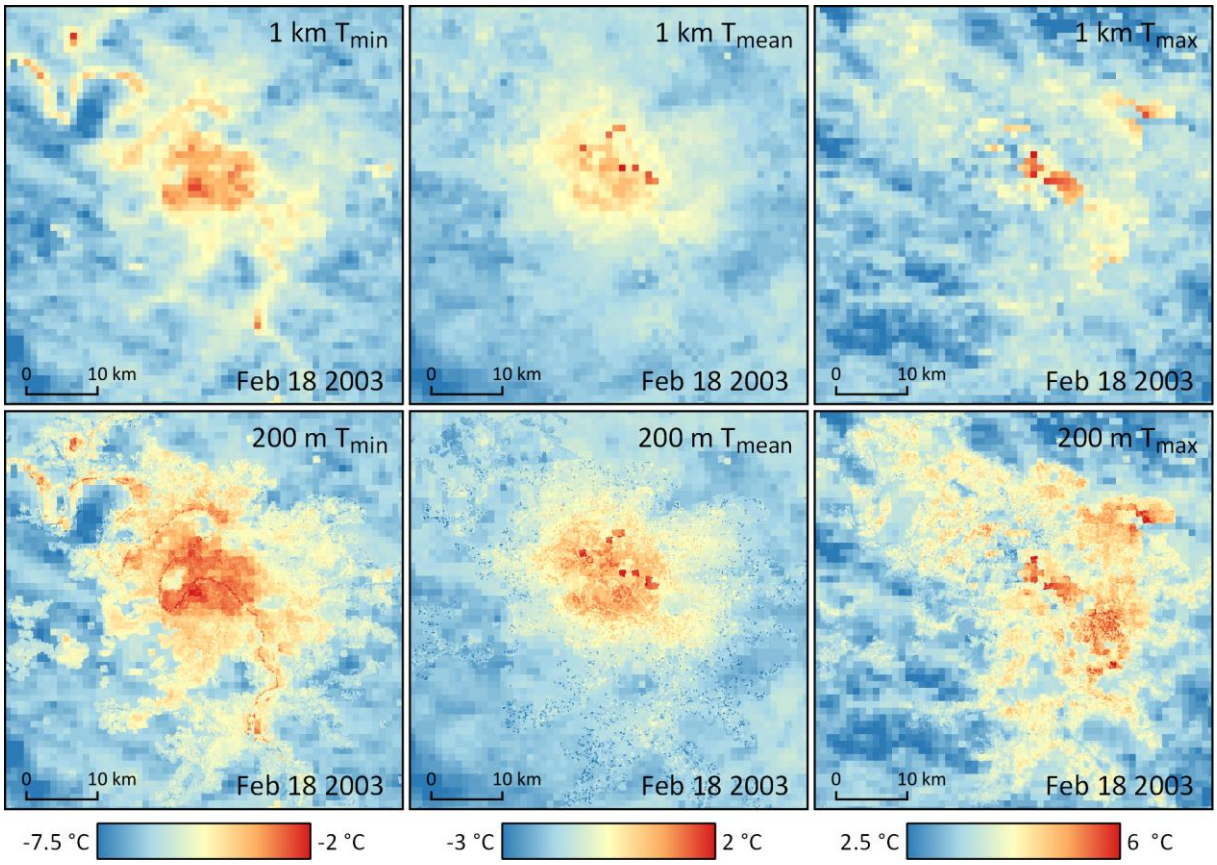

Figure A1-4. Predicted 1 km T_a from the stage 2 model alone (top row) and with predicted 200 m T_{min} from the stage 4 model overlaid (bottom row) on Feb 18, 2003 over the Paris metropolitan area.

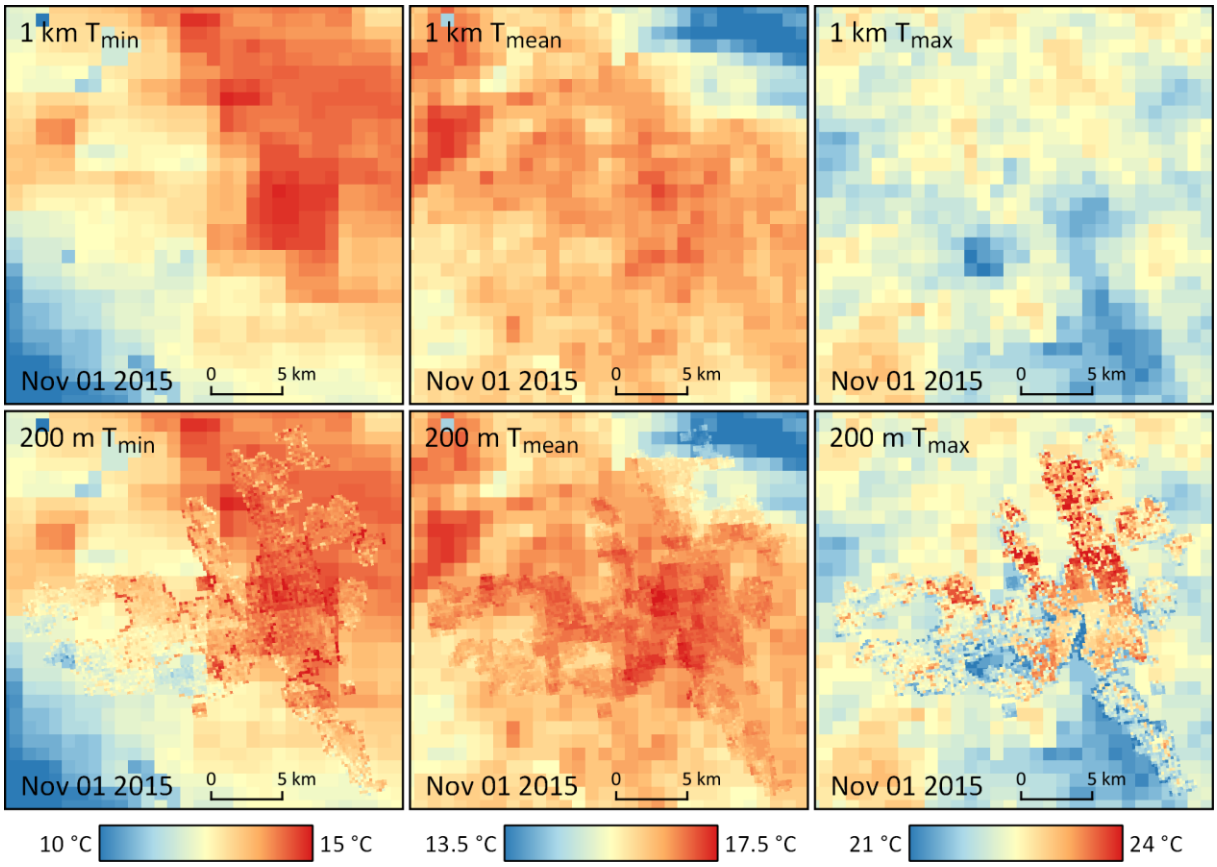

Figure A1-5. Predicted 1 km T_a from the stage 2 model alone (top row) and with predicted 200 m T_{min} from the stage 4 model overlaid (bottom row) on Nov 01, 2015 over the city of Toulouse.

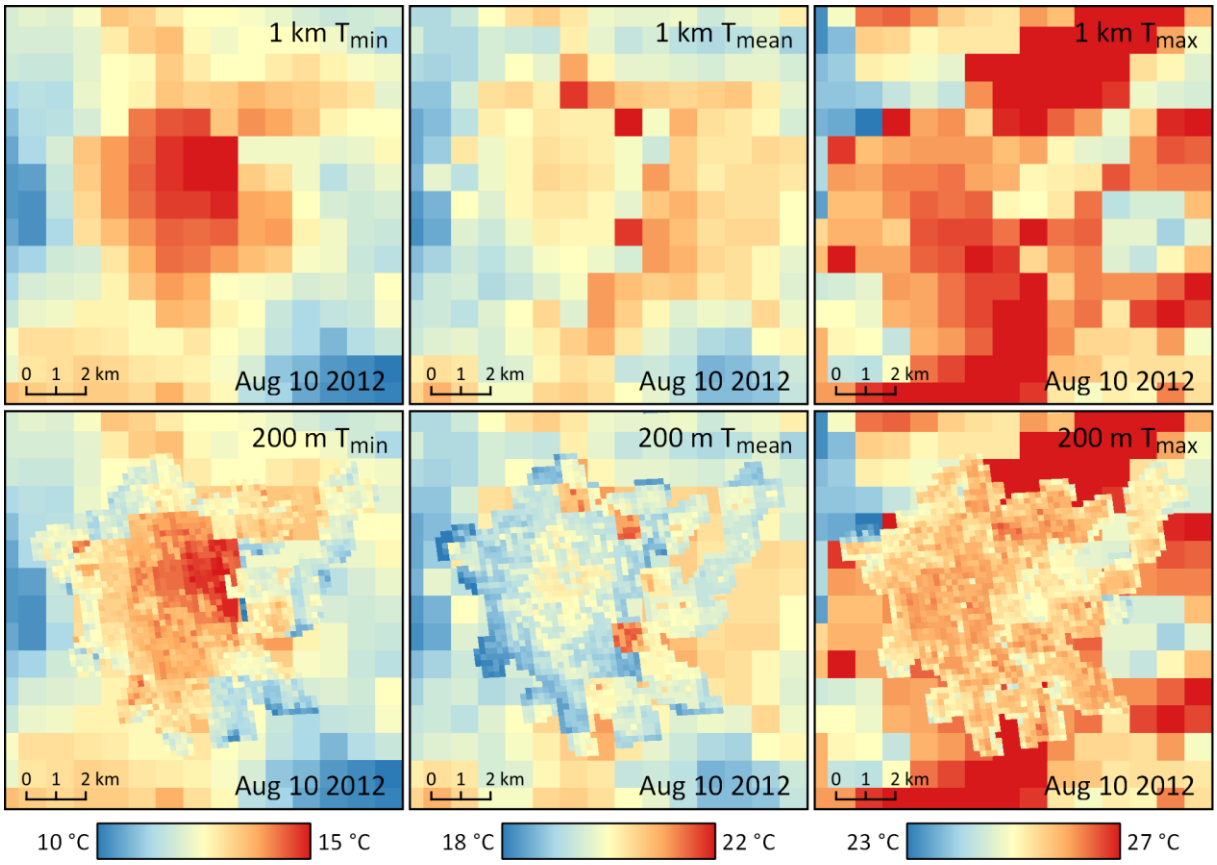

Figure A1-6. Predicted 1 km T_a from the stage 2 model alone (top row) and with predicted 200 m T_{min} from the stage 4 model overlaid (bottom row) on Aug 10, 2012 over the city of Nancy.

Appendix 2: supplement to Chapter 2

Table A2-1. Cross-validated (CV) performance of the stage 1 random forest (imputing $PM_{2.5}$ at PM_{10} monitors), stratified by year. The cross-validation scheme is described in Chapter 2, section 2.3.5 and performance metrics in section 2.3.6.

		Observed $PM_{2.5}$		5-Fold CV Performance								
				Total					Spatial		Temporal	
Year	Monitors[a]	Mean[b]	SD[c]	RMSE	Bias[d]	Slope[e]	R^2	MAE	R^2	MAE	R^2	MAE
2000	4	17.9	10.6	4.08	0.94	0.76	0.85	2.87	0.92	1.08	0.81	2.45
2001	13	16.2	8.7	3.00	0.23	0.84	0.88	2.13	0.91	1.01	0.87	1.92
2002	22	14.6	8.3	2.72	-0.03	0.86	0.89	1.94	0.91	0.89	0.88	1.82
2003	26	15.7	8.3	2.91	-0.18	0.87	0.88	2.09	0.90	0.87	0.87	1.98
2004	34	13.5	7.3	2.95	-0.19	0.83	0.84	2.00	0.77	0.93	0.83	1.87
2005	43	14.1	7.5	2.93	-0.06	0.84	0.85	1.99	0.86	0.83	0.84	1.86
2006	45	14.4	8.4	2.90	-0.07	0.86	0.88	2.02	0.83	1.03	0.89	1.82
2007	39	14.0	8.6	3.48	-0.91	0.85	0.83	2.46	0.72	1.64	0.87	1.98
2008	42	15.2	9.6	4.80	-0.16	0.76	0.75	3.35	0.66	2.54	0.80	2.70
2009	66	18.1	12.1	4.57	-0.07	0.85	0.86	3.27	0.81	2.22	0.88	2.69
2010	68	18.1	10.7	4.49	0.09	0.82	0.82	3.24	0.65	1.71	0.85	2.79
2011	76	18.2	12.6	4.55	0.00	0.87	0.87	3.22	0.69	1.77	0.89	2.70
2012	86	16.2	11.3	4.14	-0.21	0.89	0.87	3.03	0.61	1.87	0.90	2.44
2013	97	15.6	11.7	3.85	-0.03	0.88	0.89	2.82	0.70	1.58	0.91	2.42
2014	98	13.0	9.2	3.29	-0.18	0.90	0.87	2.43	0.63	1.43	0.90	2.01
2015	106	13.2	9.5	3.20	-0.12	0.89	0.89	2.31	0.80	1.15	0.90	2.03
2016	104	12.2	9.1	3.14	-0.05	0.89	0.88	2.23	0.77	1.15	0.90	1.93
2017	109	11.4	9.3	3.15	-0.30	0.90	0.89	2.21	0.62	1.27	0.92	1.82
2018	113	10.7	7.1	3.00	-0.24	0.85	0.82	2.09	0.52	1.10	0.86	1.76
2019	110	9.5	7.1	2.75	-0.32	0.88	0.85	1.87	0.49	1.04	0.89	1.58
Mean	65	14.6	9.4	3.50	-0.09	0.85	0.86	2.48	0.74	1.36	0.87	2.13

[a] Number of monitors measuring both $PM_{2.5}$ and PM_{10}
[b] $\mu g/m^3$
[c] Standard deviation
[d] Mean error
[e] Slope of regression of CV predicted $PM_{2.5}$ on observed $PM_{2.5}$

Table A2-2. Cross-validated (CV) performance of the stage 2 random forest (filling gaps in MAIAC AOD), stratified by year. The cross-validation scheme is described in Chapter 2, section 2.3.5 and performance metrics in section 2.3.6.

			5-fold CV Performance								
	MAIAC AOD		Total					Spatial		Temporal	
Year	Mean	SD[a]	RMSE	Bias[b]	Slope[c]	R^2	MAE	R^2	MAE	R^2	MAE
2000	0.12	0.087	0.045	-0.014	0.78	0.67	0.032	0.52	0.025	0.72	0.024
2001	0.13	0.092	0.047	-0.015	0.82	0.71	0.033	0.61	0.025	0.75	0.024
2002	0.13	0.091	0.044	-0.013	0.81	0.72	0.031	0.60	0.022	0.76	0.024
2003	0.16	0.101	0.044	-0.011	0.82	0.77	0.031	0.66	0.017	0.81	0.025
2004	0.13	0.088	0.044	-0.012	0.77	0.71	0.031	0.65	0.019	0.76	0.024
2005	0.13	0.084	0.043	-0.012	0.78	0.71	0.030	0.60	0.018	0.76	0.024
2006	0.13	0.090	0.045	-0.013	0.80	0.72	0.032	0.69	0.020	0.77	0.025
2007	0.15	0.101	0.046	-0.012	0.80	0.75	0.032	0.70	0.020	0.78	0.026
2008	0.12	0.077	0.042	-0.013	0.76	0.66	0.030	0.61	0.018	0.73	0.022
2009	0.14	0.088	0.043	-0.012	0.78	0.72	0.031	0.65	0.018	0.76	0.024
2010	0.12	0.082	0.042	-0.012	0.78	0.71	0.030	0.61	0.018	0.76	0.023
2011	0.13	0.082	0.040	-0.011	0.79	0.70	0.028	0.63	0.016	0.75	0.023
2012	0.13	0.089	0.042	-0.012	0.79	0.73	0.029	0.68	0.018	0.76	0.023
2013	0.13	0.086	0.044	-0.013	0.76	0.69	0.031	0.65	0.019	0.73	0.024
2014	0.11	0.073	0.041	-0.013	0.74	0.64	0.029	0.60	0.017	0.70	0.022
2015	0.12	0.077	0.039	-0.011	0.75	0.69	0.028	0.60	0.016	0.74	0.022
2016	0.11	0.072	0.040	-0.013	0.76	0.65	0.028	0.60	0.016	0.72	0.021
2017	0.11	0.078	0.040	-0.012	0.76	0.68	0.028	0.59	0.017	0.74	0.022
2018	0.12	0.075	0.042	-0.013	0.76	0.66	0.030	0.59	0.017	0.71	0.023
2019	0.12	0.074	0.039	-0.011	0.76	0.69	0.028	0.61	0.017	0.73	0.021
Mean	0.13	0.084	0.043	-0.012	0.78	0.70	0.030	0.62	0.019	0.75	0.023

[a] Standard deviation

[b] Mean error

[c] Slope of regression of CV predicted AOD on observed AOD

Table A2-3. Cross-validated (CV) performance of the stage 4 GAM ensemble, stratified by region. The cross-validation scheme is described in Chapter 2, section 2.3.5 and performance metrics in section 2.3.6.

					Multi-stage CV Performance						
		PM			Total			Spatial		Temporal	
Region	Area[a]	N[b]	Mean	SD[c]	RMSE	R^2	MAE	R^2	MAE	R^2	MAE
$PM_{2.5}$											
Auvergne-Rhône-Alpes	70.9	54	14.4	9.7	4.10	0.82	2.70	0.76	1.16	0.83	2.52
Bourgogne-Franche-Comté	48.1	20	12.4	8.6	3.93	0.79	2.65	0.63	0.99	0.80	2.50
Bretagne	27.5	7	11.8	7.2	4.01	0.69	2.76	0.33	1.19	0.71	2.55
Centre-Val de Loire	39.5	15	12.4	8.3	2.88	0.88	1.95	0.69	0.81	0.89	1.85
Grand Est	57.7	44	13.7	8.9	4.30	0.77	2.81	0.53	1.29	0.78	2.68
Hauts-de-France	31.9	38	14.8	10.0	3.78	0.86	2.53	0.52	0.92	0.87	2.39
Île-de-France	12.1	19	16.0	10.5	5.35	0.74	3.58	0.07	2.75	0.87	2.41
Normandie	30.1	20	13.3	9.3	3.69	0.84	2.54	0.43	1.24	0.86	2.32
Nouvelle-Aquitaine	85.2	35	12.3	7.6	3.79	0.75	2.58	0.31	1.17	0.78	2.34
Occitanie	73.4	21	12.9	7.2	3.59	0.75	2.52	0.38	1.32	0.77	2.31
Pays de la Loire	32.4	13	11.3	7.5	2.96	0.84	2.09	0.51	1.40	0.86	1.89
PACA	31.7	25	16.7	8.1	4.49	0.69	3.25	0.63	1.51	0.71	3.01
PM_{10}											
Auvergne-Rhône-Alpes	70.9	54	21.8	13.2	6.26	0.77	4.19	0.67	2.09	0.80	3.77
Bourgogne-Franche-Comté	48.1	20	18.6	11.5	5.80	0.75	4.03	0.55	1.58	0.76	3.75
Bretagne	27.5	7	19.1	10.2	5.76	0.68	4.08	0.31	1.63	0.70	3.86
Centre-Val de Loire	39.5	15	19.0	10.5	4.01	0.86	2.84	0.43	1.38	0.87	2.64
Grand Est	57.7	44	20.3	12.3	6.80	0.70	4.46	0.39	2.24	0.72	4.14
Hauts-de-France	31.9	38	23.1	13.1	5.27	0.84	3.63	0.51	1.49	0.85	3.37
Île-de-France	12.1	19	25.8	15.0	9.27	0.62	6.20	0.06	5.35	0.84	3.60
Normandie	30.1	20	20.6	11.7	5.18	0.80	3.73	0.28	2.29	0.83	3.24
Nouvelle-Aquitaine	85.2	35	19.7	10.5	5.89	0.68	4.05	0.30	1.90	0.72	3.60
Occitanie	73.4	21	20.0	10.3	5.50	0.71	3.91	0.22	2.20	0.74	3.53
Pays de la Loire	32.4	13	18.4	9.7	4.15	0.82	2.97	0.29	2.05	0.84	2.67
PACA	31.7	25	27.7	12.8	7.88	0.62	5.64	0.54	2.88	0.65	5.06

[a] 100 km^2

[b] Number of monitors

[c] Standard deviation

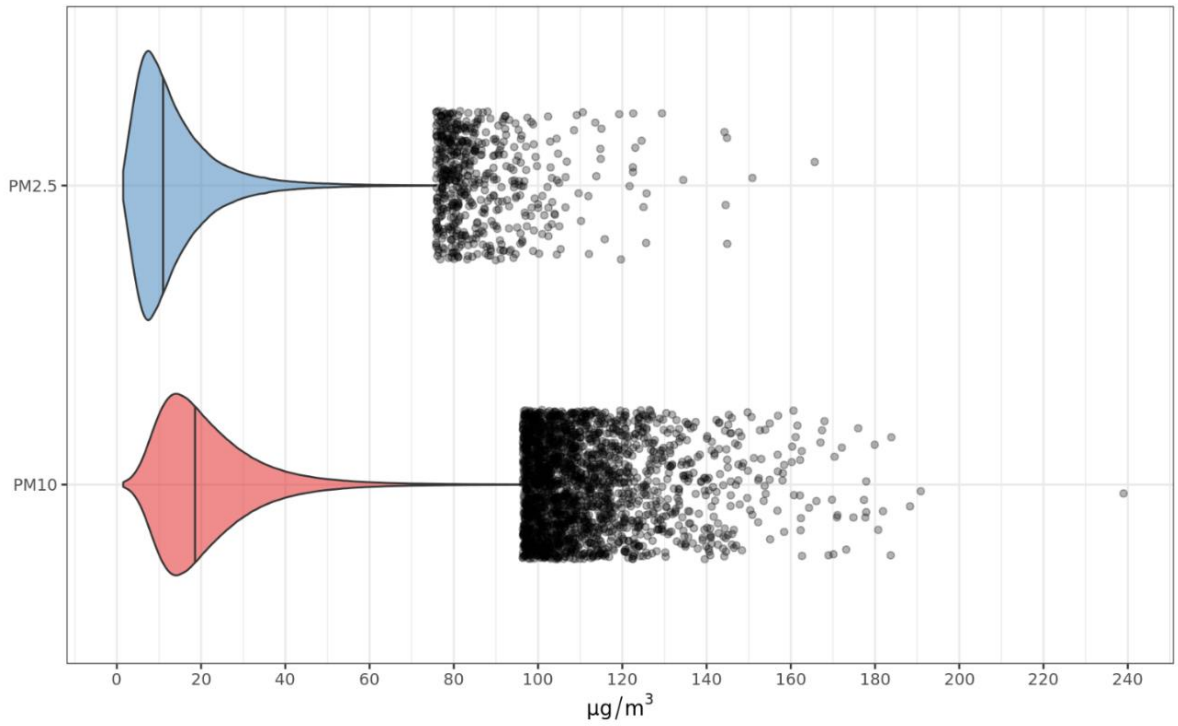

Figure A2-1. Distribution of daily $PM_{2.5}$ and PM_{10} concentration observed by all monitors from 2000 to 2019. Vertical lines show median and points show highest 0.1% (jittered vertically for clarity).

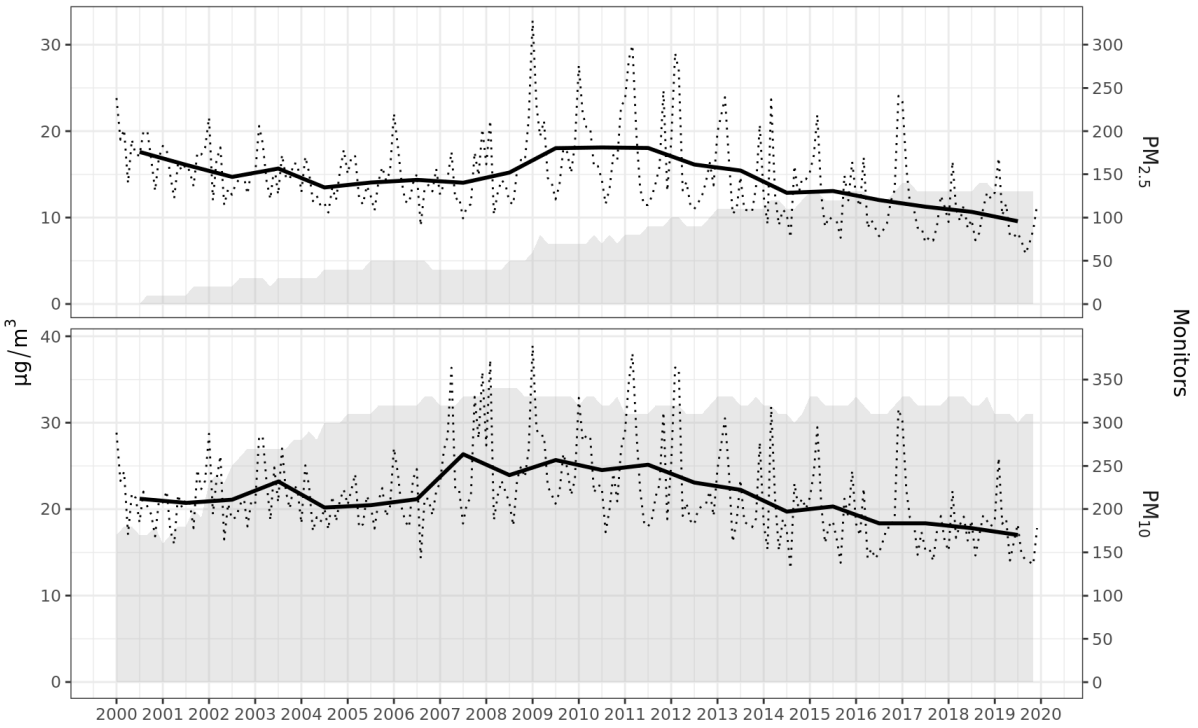

Figure A2-2. Mean $PM_{2.5}$ (top) and PM_{10} (bottom) concentration (left y-axis) observed by all monitors. Dotted line shows monthly mean; solid line shows annual mean. Grey area shows number of monitors (right y-axis). The increase in PM_{10} concentrations in 2007 and $PM_{2.5}$ concentrations in 2008 to 2009 is due to a change in monitor technology: monitors were modified to include semi-volatile PM as well as non-volatile PM in their measurements.

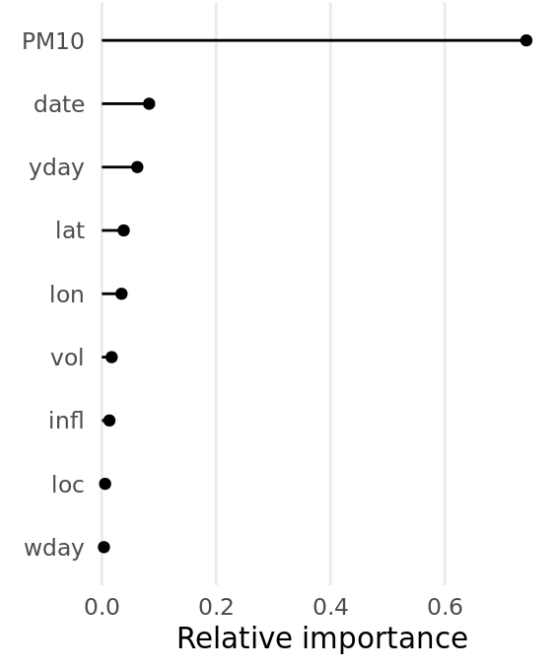

Figure A2-3. Relative importance (estimated by permutation) of the predictors used by the stage 1 random forest to impute $PM_{2.5}$ at PM_{10} stations.

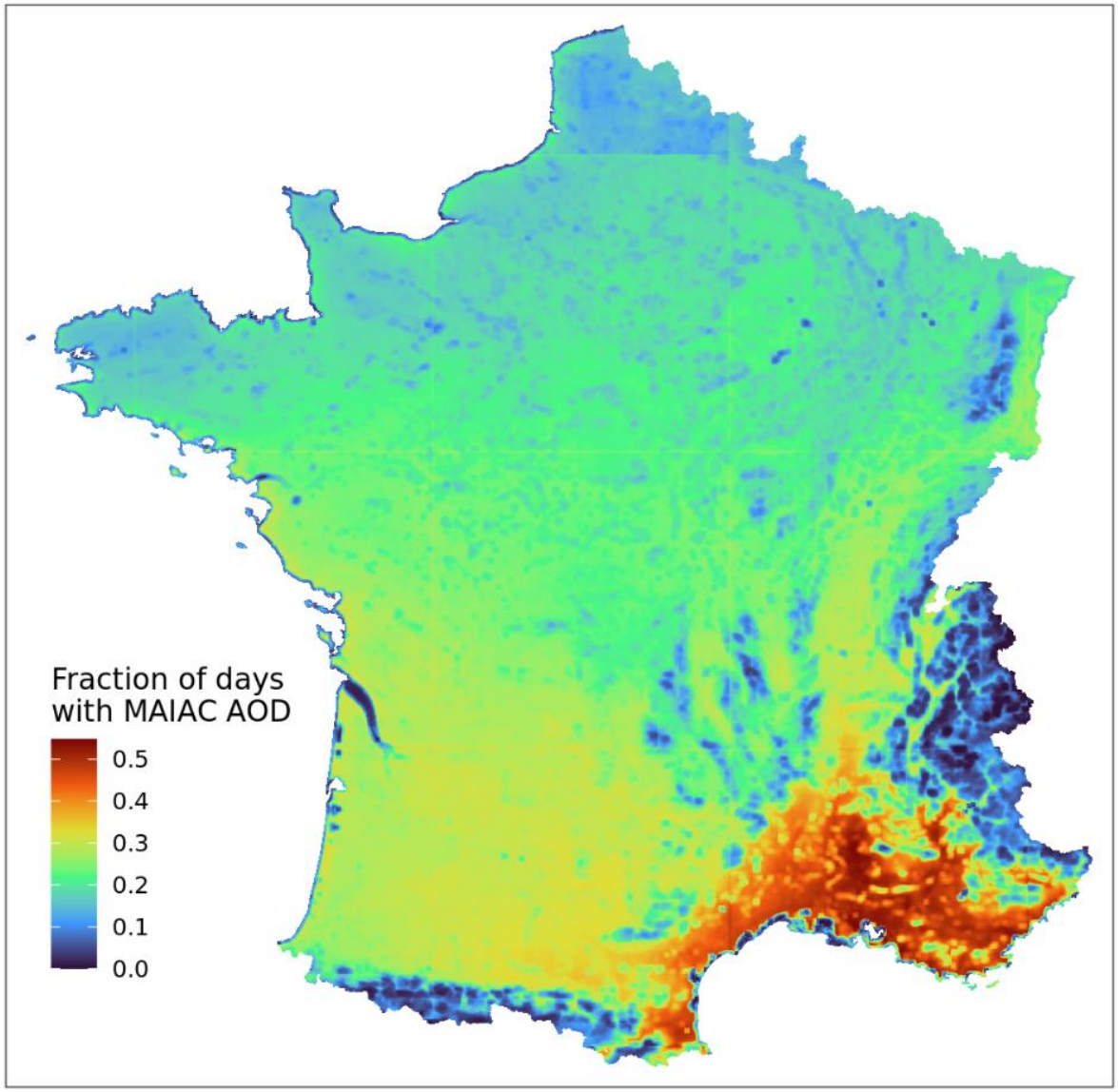

Figure A2-4. Availability of MAIAC AOD over continental France from 2000 to 2019.

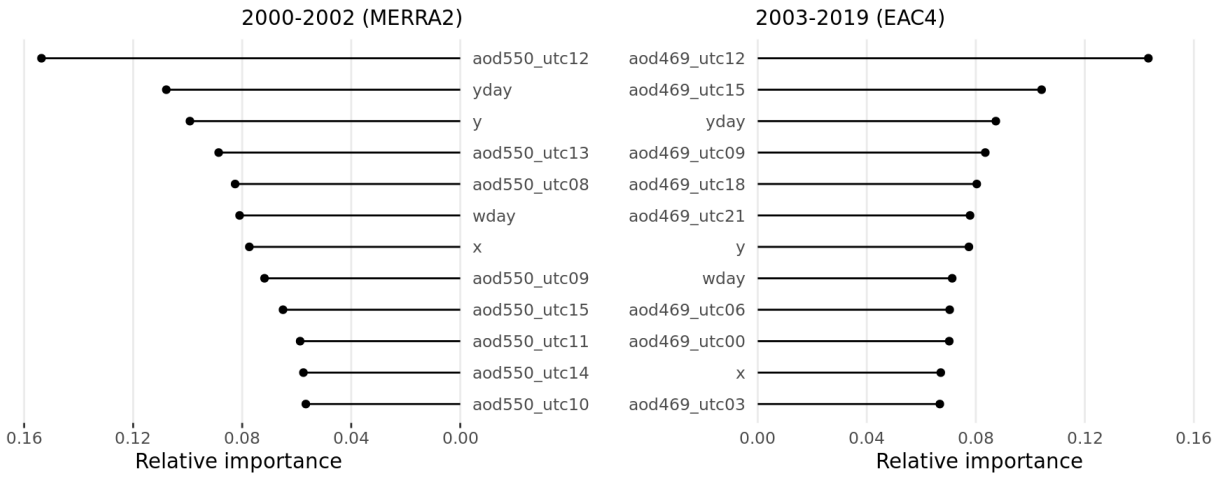

Figure A2-5. Mean relative importance (estimated by permutation) of the predictors used by the stage 2 random forests to fill gaps in MAIAC AOD. Modelled AOD from the MERRA2 atmospheric reanalysis (left) were used in the period predating the EAC4 atmospheric reanalysis (right).

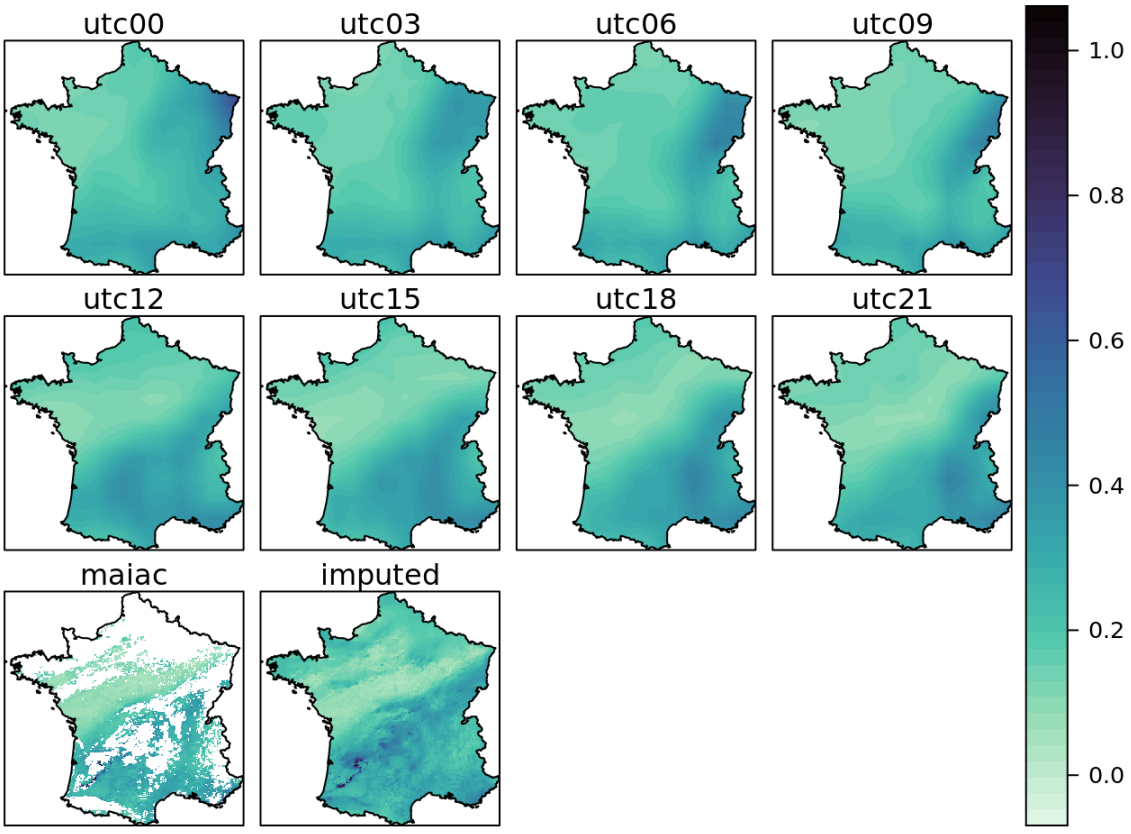

Figure A2-6. Example of filling gaps in MAIAC AOD using the stage 2 random forest. Top two rows (utc00, utc03, …, utc21) show 3-hourly modelled AOD from EAC4 atmospheric reanalysis on 22 August 2012. Bottom left (maiac) shows MAIAC AOD on the same day; white areas are gaps where data is missing due to cloud cover. Bottom right (imputed) shows final AOD after filling gaps with the stage 2 random forest.

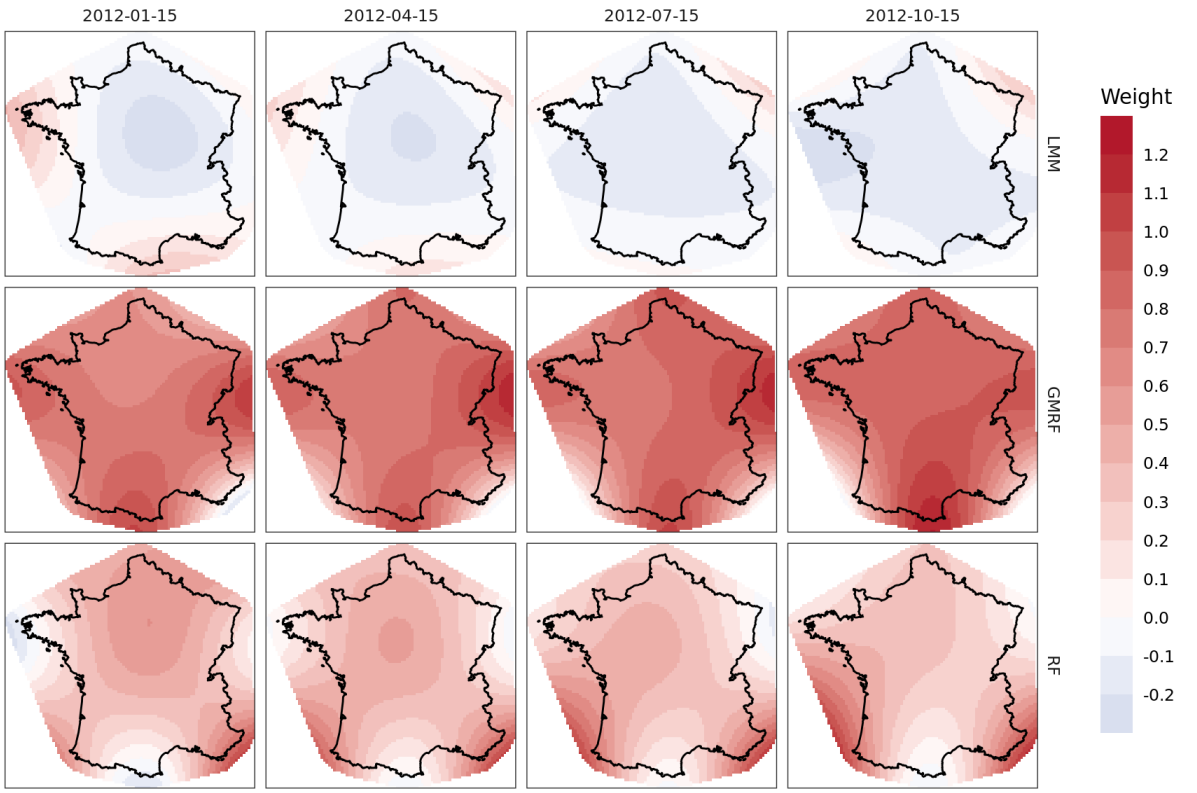

Figure A2-7. Weights used by the GAM ensemble to predict daily 1 km $PM_{2.5}$ in 2012. Rows show the weight assigned to the predictions of each base learner (top, LMM; middle, GMRF; bottom, RF) and columns show four different days in 2012. To predict $PM_{2.5}$ on 15 January 2012, the GAM multiplies the weights in the left column by the corresponding predictions from each base learner and sums. The GMRF was the most important base learner over most of the study area (largest weights); this was the case across all years and both $PM_{2.5}$ and PM_{10}. Note, though, that for this ensemble ($PM_{2.5}$ in 2012) the RF had a higher weight than the GMRF over the southeastern and southwestern corners of continental France.

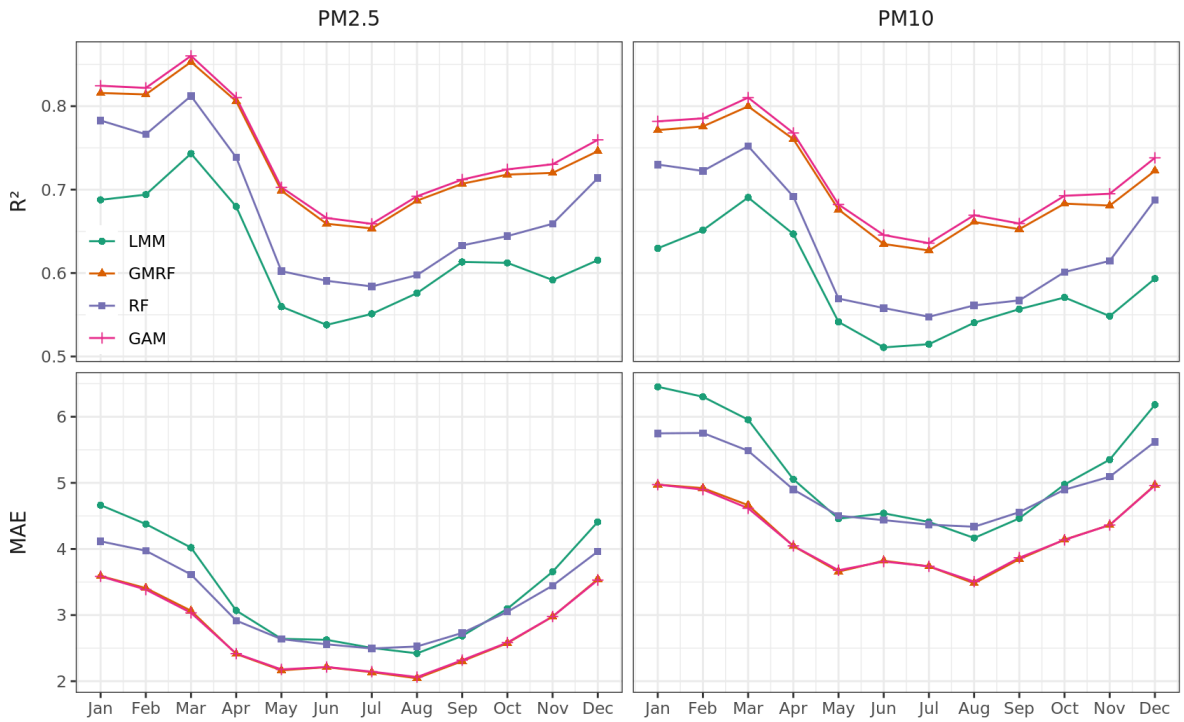

Figure A2-8. Seasonal cross-validated performance of the stage 3 base learners (LMM, GMRF, RF) and stage 4 ensemble (GAM) predicting daily 1 km $PM_{2.5}$ (left) and PM_{10} (right), stratified by month. Top shows R^2 (percent of explained variability); bottom shows MAE (mean absolute error; $\mu g/m^3$). Higher MAE in winter (Dec to Mar) and lower MAE in summer (May to Aug) coincide with higher PM concentrations in winter and lower PM concentrations in summer.

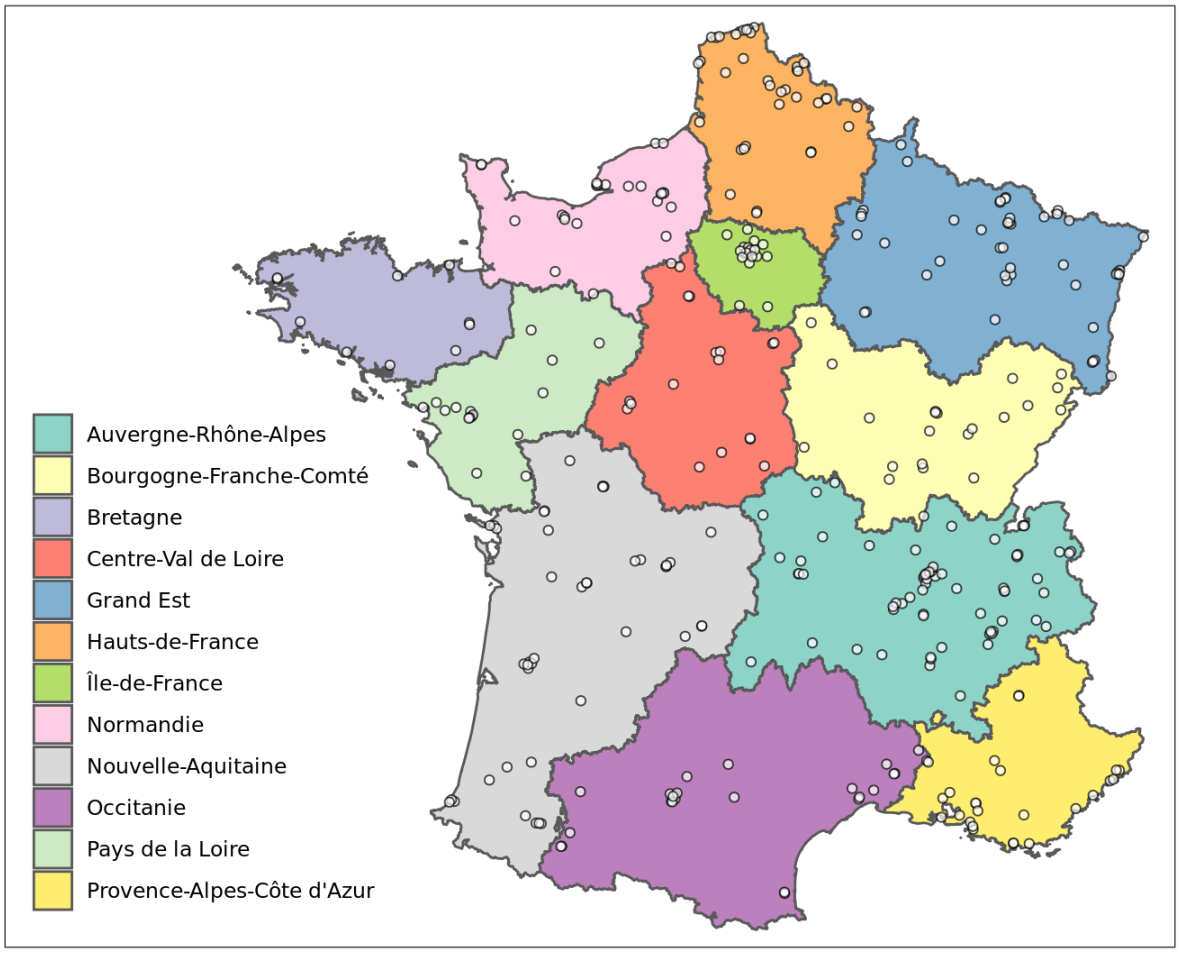

Figure A2-9. Administrative regions of continental France. Points indicate PM monitors.

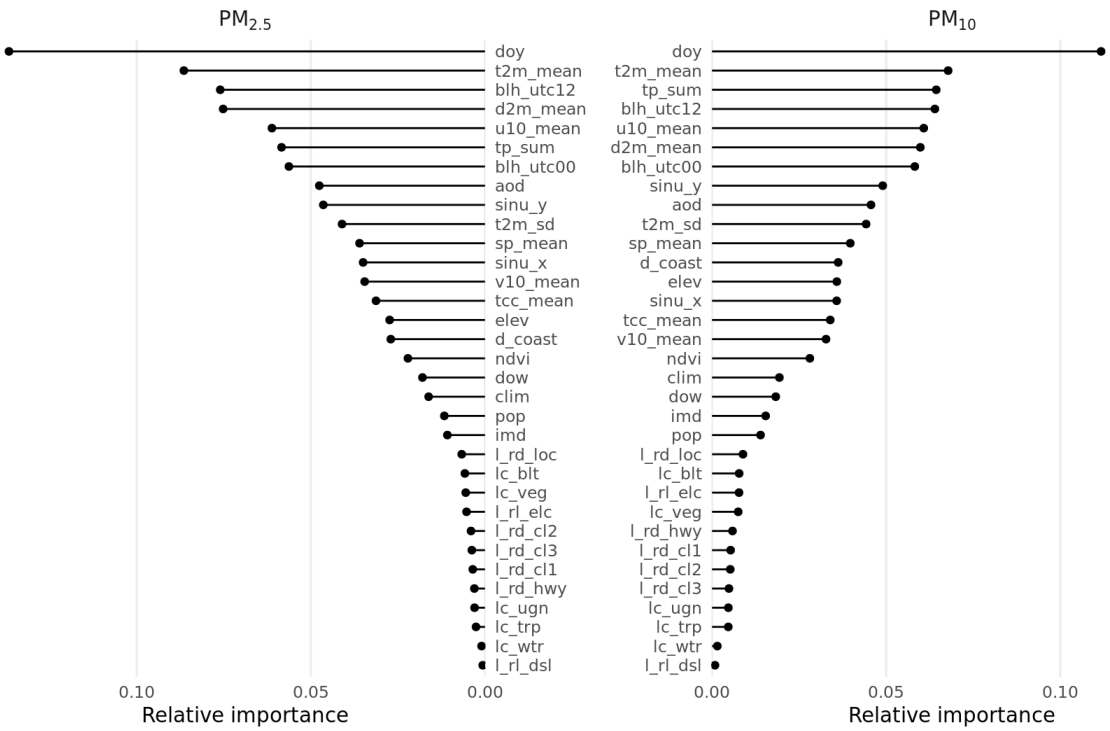

Figure A2-10. Mean relative importance (estimated by permutation) of the predictors used by the stage 3 random forests to predict monitored PM2.5 (left) and PM10 (right) from the spatiotemporal and spatial variables.

Appendix 3: supplement to Chapter 3

Table A3-1. P_{95}: 95th percentile of daily mean temperature over the 5 years preceding the first conception in each cohort area.

Cohort	Area	P_{95}
EDEN	Nancy	21.2
	Poitiers	21.5
PELAGIE	Côtes-d'Armor	18.4
	Finistère	17.8
	Ille-et-Vilaine	20.6
SEPAGES	Grenoble	23.1

Table A3-2. Quantiles for each indicator by exposure period.

Period	Indicator	1%	5%	10%	25%	50%	75%	90%	95%	99%
First 26 weeks following conception	T_{mean} (°C)	-0.8	2.0	4.0	6.9	11.6	16.7	19.9	21.7	25.3
	T_{min} (°C)	-4.2	-1.1	0.6	3.5	7.4	11.6	14.4	15.7	18.3
	T_{max} (°C)	2.7	5.7	7.6	10.9	16.8	23.0	27.2	29.3	33.5
	T_{SD}[a] (°C)	0.5	0.8	1.0	1.3	1.8	2.5	3.3	3.8	4.8
	EHF	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.4	19.4
Last 30 days of pregnancy	T_{mean} (°C)	-1.6	1.2	3.2	7.2	12.1	17.0	20.5	22.6	26.0
	T_{min} (°C)	-5.4	-2.3	-0.4	1.7	7.9	12.1	15.2	16.6	19.0
	T_{max} (°C)	1.6	5.0	7.1	11.3	17.2	23.3	28.1	30.6	35.0
	T_{SD}[a] (°C)	0.5	0.8	1.0	1.3	1.9	2.5	3.3	3.8	4.6
	EHF	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.8	18.1

[a] Weekly standard deviation of daily mean temperature

Table A3-3. Characteristics of subjects included and excluded from the main analyses.

	Included in analysis *n (%) or mean (SD)*	Excluded from analysis *n (%) or mean (SD)*
Participants	5347 (100%)	461 (100.0%)
Unknown date of birth	0 (0.0%)	101 (21.9%)
Preterm births	232 (4.3%)	25 (5.4%)
Duration of pregnancy[a] (weeks)	37.9 (1.7)	37.5 (2.5)
Temperature[b] (°C)	11.8 (3.6)	12 (3.8)
Temperature variability[c] (°C)	2.0 (0.3)	1.9 (0.3)
Excess Heat Factor[d]	0.7 (1.1)	0.9 (1.4)
Cohort		
EDEN	1806 (33.8%)	196 (42.5%)
PELAGIE	3116 (58.3%)	206 (44.7%)
SEPAGES	425 (7.8%)	59 (12.8%)
Child sex		
Boy	2737 (51.2%)	200 (43.4%)
Girl	2610 (48.8%)	156 (33.8%)
Missing	0 (0.0%)	105 (22.8%)
Parity		
0	2374 (44.4%)	169 (36.7%)
1	2021 (37.8%)	121 (26.2%)
>=2	952 (17.8%)	61 (13.2%)
Missing	0 (0.0%)	110 (23.9%)
Maternal age at conception (years)	30 (4.5)	29.4 (4.9)
Missing	0 (0.0%)	115 (25%)
Maternal pre-pregnancy BMI		
<18.5 kg/m^2	412 (7.7%)	23 (5.0%)
18.5 – 25 kg/m^2	3844 (71.9%)	224 (48.6%)
>25 kg/m^2	1091 (20.4%)	67 (14.5%)
Missing	0 (0.0%)	147 (31.9%)
Maternal education		
Baccalaureate or less	1987 (37.2%)	167 (36.2%)
Baccalaureate + 1 or 2 years	1373 (25.7%)	52 (11.3%)
>= Baccalaureate +3 years	1987 (37.2%)	113 (24.5%)
Missing	0 (0.0%)	129 (28.0%)
Smoking status during pregnancy		
None	4006 (74.9%)	191 (41.4%)
Active smoker	1341 (25.1%)	104 (22.6%)
Missing	0 (0.0%)	166 (36.0%)
Urbanicity		
City-centre	896 (16.8%)	57 (12.4%)
Small city-centre or suburban	2021 (37.8%)	105 (22.8%)
Rural	2430 (45.4%)	155 (33.6%)
Missing	0 (0.0%)	144 (31.2%)
Season of conception		
Winter	1321 (24.7%)	117 (25.4%)
Spring	1211 (22.6%)	112 (24.3%)
Summer	1477 (27.6%)	109 (23.6%)
Autumn	1338 (25.0%)	117 (25.4%)
Missing	0 (0.0%)	6 (1.3%)

[a] Conception to delivery

[b] Over the 26 weeks following conception

[c] Weekly standard deviation (SD) over the 26 weeks following conception

[d] Over days 30-181 after conception

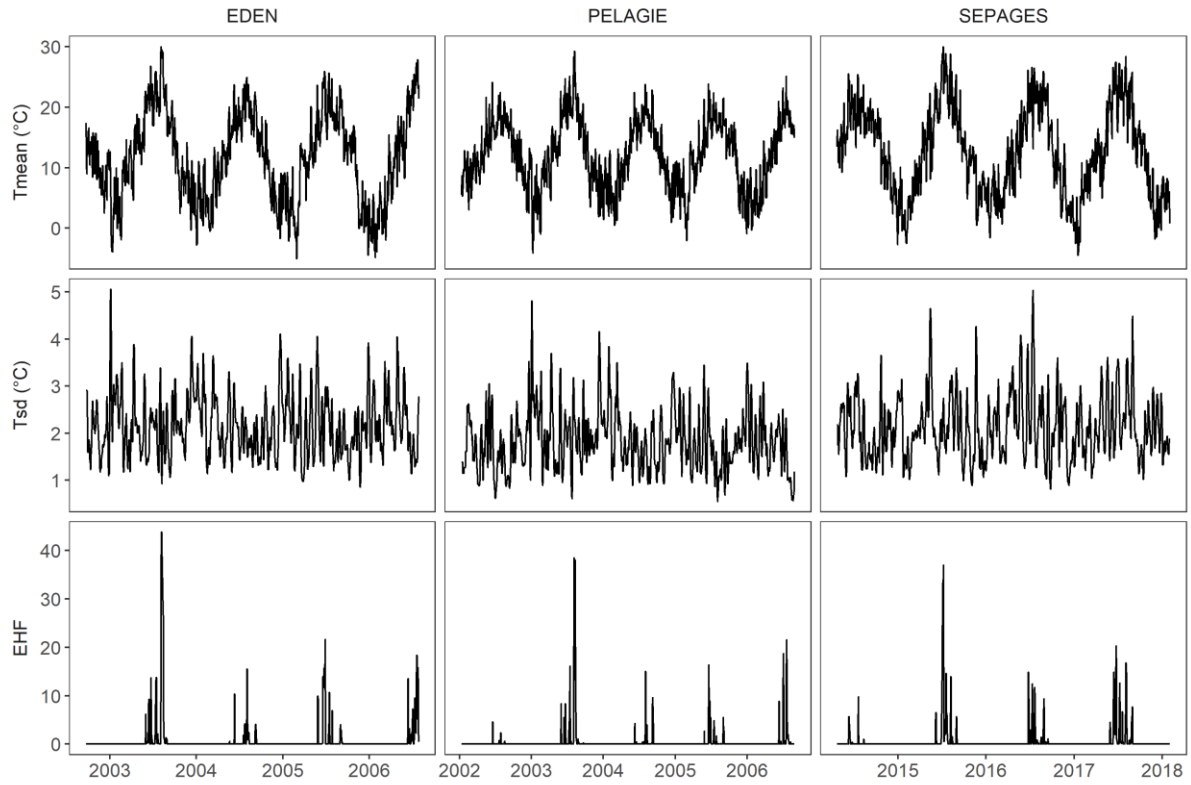

Figure A3-1. Daily mean temperature, temperature variability, and EHF at the home address of study participants from the first conception to last delivery in each cohort.

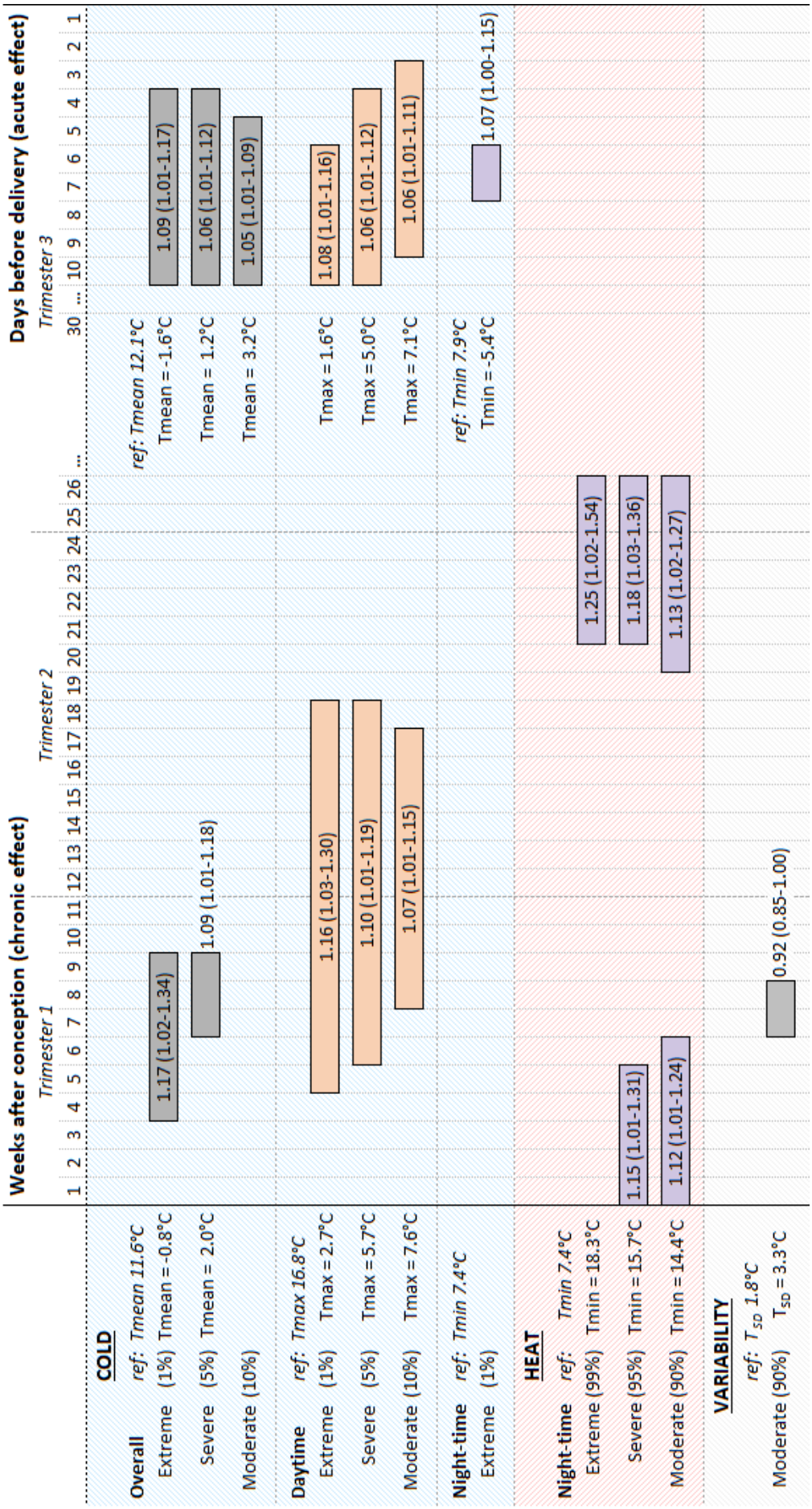

Figure A3-2. Statistically significant critical windows for exposure to indicators of ambient temperature and risk of preterm birth. Bars show timing of critical windows; bar labels show mean adjusted relative risk (95% CI) for preterm birth associated with exposure for one week or day during the critical window.

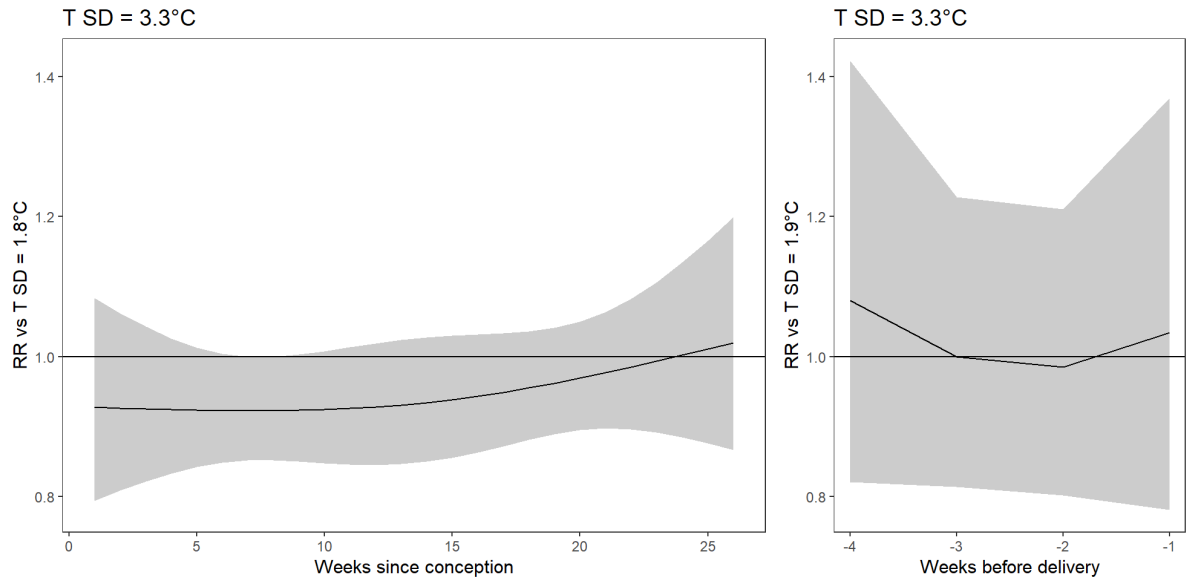

Figure A3-3. Adjusted relative risk (RR) for preterm birth associated with moderate temperature variability (90th percentile of T_{SD}) during each of the 26 weeks following conception (left) and the 30 days ending at delivery (right). Shaded area shows 95% confidence interval.

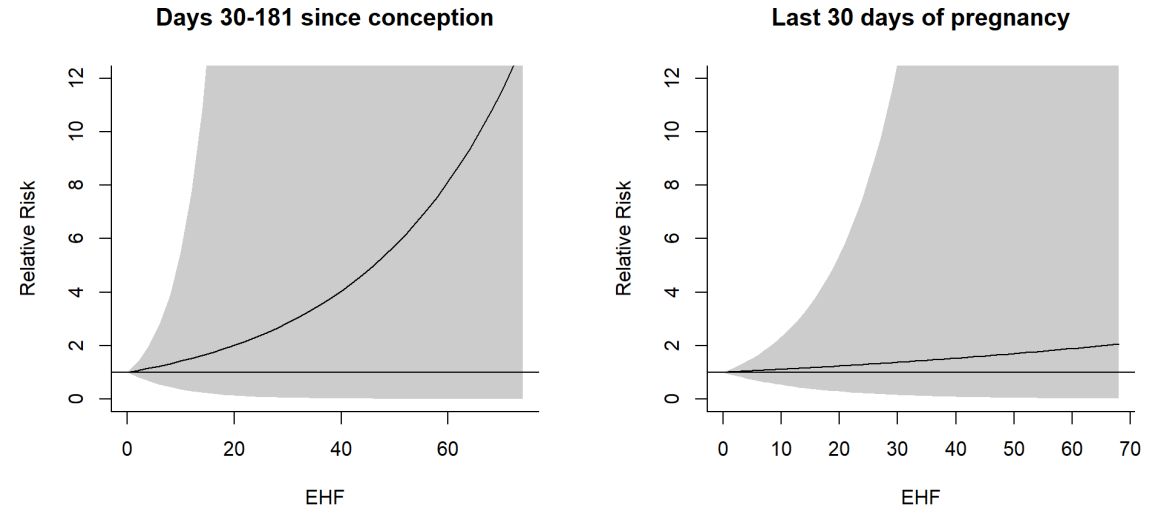

Figure A3-4. Cumulative adjusted relative risk for preterm birth associated with different levels of EHF exposure throughout the 30-181 days after conception (left) or the 30 days ending at delivery (right). Shaded area shows 95% confidence interval.

תקציר (Hebrew abstract)

מחקרים רבים מצביעים על כך שחשיפה לחום, קור או זיהום אוויר (PM) במהלך ההיריון עלול להפחית את משקל הלידה של העובר ואף לקצר את משך ההיריון, אך אין בספרות עדיין הבנה מספקת על קשרים אלו. מכיוון שמשקל לידה נמוך ולידה מוקדמת מעלים את הסיכון לתחלואה ולמחלות כרוניות רבות במהלך הילדות ועד לבגרות יש משמעות רבה להבנה עמוקה של קשרים אלו. עומסי הבריאות העתידיים כתוצאה מעליה צפויה בטמפרטורה וחשיפה מוגברת ל PM צפויים לגדול עם שינויי האקלים והעיור הגלובלי. שיפור ההבנה הכוללת שלנו להשפעות אלו דורש הערכת חשיפה מאוד מדויקת מכיוון שהטיה בחשיפה עלולה להחליש את הקשר עם תוצאות הבריאות.

במסגרת עבודת הדוקטורט ,פיתחנו גישות חדשות וייחודיות שמעלות את הדיוק והרזולוציה המרחבית של המודלים מבוססי לוויין ביחס למודלים קודמים, המאפשרים לנו לשחזר בצורה מדויקת את הטמפרטורה היומית וה PM בצרפת במהלך שני העשורים האחרונים. לאחר מכן השתמשנו בחשיפות אלו כדי לזהות תקופות במהלך ההיריון שבהן חשיפה לקור או לחום עלולה להגביר את הסיכון ללידה מוקדמת.

בראשית המחקר גילינו שחשוב לאפשר לקשר בין טמפרטורת הסביבה המנוטרת לבין LST הנגזר מהלוויין MODISלהשתנות על פני גם המרחב והזמן. התגלית הזאת משפרת משמעותית את הדיוק של המודלים שלנו ביחס למודלים קודמים. השתמשנו בגישה זו כדי למדל את טמפרטורת האוויר היומית (ממוצע [Tmean], מינימום [Tmin] ומקסימום [Tmax]) בצרפת ברזולוציה של קילומטר אחד משנת 2000 עד 2018. כמו כן, הראינו כי שימוש במודלים של Random forest ו Extreme Gradient Boosting הצליחו מאפשרים לנו לרדת ברזולוציה המרחבית של הטמפרטורה היומית באזורים עירוניים על ידי שילוב נתוני לוויין תרמיים . השתמשנו בגישה זו כדי למדל מדי יום Tmean, Tmin ו- Tmax ברזולוציה של 200 מטר ב 103 ערים בצרפת.

שנית, הראנו ששימוש ב Gaussian Markov random fields במודל ושילוב של מודלים שונים של machine learning (כגון Random forest ו Extreme Gradient Boosting) יכולים לעזור בשחזור מדויק יותר של זיהום אוויר ומיוחד באזורים מרוחקים מערים ראשיות היכן שיש מחסור בתחנות ניטור. השתמשנו בגישה זו כדי להריץ מודל של PM2.5 ו PM10 (PM <2.5 ו <10 מיקרון בקוטר, בהתאמה) יומי על פני צרפת ברזולוציה של קילומטר אחד משנת 2000 עד 2019. ביצועי המודל היו טובים מאוד ביחס למודלים קודמים בצרפת. המודל החדשני מאפשר לנו לשחזר חשיפה ל PM2.5 הן לטווח הארוך והן לטווח הקצר למחקרים אפידמיולוגיים גם באזורים מטרופוליטניים רחבי היקף וגם באזורים מרוחקים מערים.

שלישית, ביצענו ניתוח הישרדות באמצעות מודלים של Cox proportional hazards עם distributed lags כדי להעריך את הקשר המשתנה בין טמפרטורת הסביבה באזור המגורים הנאמדת על ידי מודל החשיפה שלנו לבין הסיכון ללידה מוקדמת בקרב 5347 לידות (4.3% מתוכם לידה מוקדמות) מקבוצות ההיריון הפוטנציאליות בצרפת. בדקנו במקביל חשיפה כרונית (במהלך 26 השבועות שלאחר ההתעברות) וחשיפה קצרת מועד (במהלך 30 הימים שלפני הלידה), ובחנו את ההשפעות של השתנות הטמפרטורה והתאקלמות במיקום ובעונה במהלך גלי חום וקור. חשיפה לקור וחום בלילה הגדילו את הסיכון ללידה מוקדמת, כשהרגישות מתחילה כבר בעת ההתעברות והמשך לאורך חלקים מהטרימסטרים 1 ו -2.

הגישות החדשות שלנו יכולות לשפר את מודלי החשיפה המבוססים על לוויין בתחומים אחרים ומערך הנתונים הרב שנתי שלנו יהיה בעל ערך רב לאפידמיולוגים, חוקרי אקלים, מתכננים וקובעי מדיניות ציבורית. בהקשר של עליית טמפרטורות ומפגעי מזג אוויר תכופים יותר, הממצאים שלנו על הסיכונים שחום וקור מהווים לנשים בהריון אמורים להוסיף ידע רב שיעזרו לקבוע מדיניות של בריאות הציבור בנושאי בריאות במהלך ההיריון.

<u>הצהרת תלמיד המחקר עם הגשת עבודת הדוקטור לשיפוט</u>

אני החתום מטה מצהיר/ה בזאת:

_**X**_ חיברתי את חיבורי בעצמי, להוציא עזרת ההדרכה שקיבלתי מאת מנחה/ים.

_**X**_ החומר המדעי הנכלל בעבודה זו הינו פרי מחקרי מתקופת היותי תלמיד/ת מחקר.

____ בעבודה נכלל חומר מחקרי שהוא פרי שיתוף עם אחרים, למעט עזרה טכנית הנהוגה בעבודה ניסיונית. לפי כך מצורפת בזאת הצהרה על תרומתי ותרומת שותפי למחקר, שאושרה על ידם ומוגשת בהסכמתם.

תאריך **2021-10-19** שם התלמיד/ה **Ian Hough** חתימה 

טמפרטורה, זיהום אוויר ובריאות הריונית
בהקשר של שינוי אקלים:
חשיפה והשפעות באוכלוסיות רגישות

מחקר לשם מילוי חלקי של הדרישות לקבלת תואר "דוקטור לפילוסופיה"

מאת

איאן הופ

הוגש לסנאט אוניברסיטת בן גוריון בנגב

תשפ"ב טבת י"ב **2021-12-16**

באר שבע