Modelling and quantifying mortality and longevity risk

Module C.2 on Statistical and machine learning methods for fine-grained mortality data

Katrien Antonio & Michel Vellekoop

September 18-19, 2025

University of Amsterdam

RCLR Research Centre for Longevity Risk

Learning outcomes of this module

In this module you will learn:

- pros and cons of using machine and deep learning methods for modelling mortality data
- how to specify a baseline model (of type Serfling) for short-term mortality
- how to use gradient boosting as a learning method to exlore associations between short term mortality and a high-dimensional set of weather and pollution features.

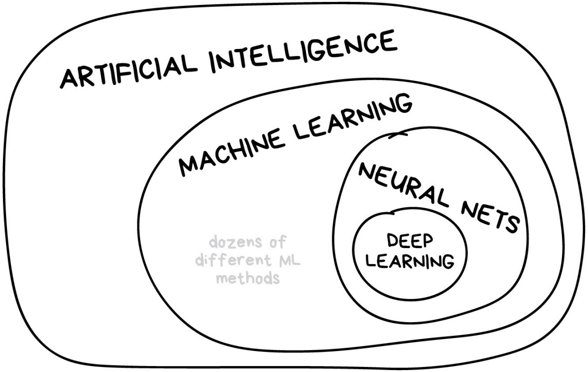
Outline

Use of machine learning techniques in mortality modelling

ML, DL and mortality modelling

In Module 6 on portfolio data we reflected on the use of machine or deep learning methods:

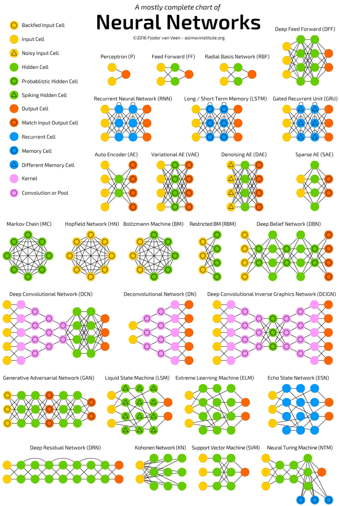
- for death counts $D_{x,t,g,\ell}$ under Poisson assumption, in the presence of risk factors or covariates $\boldsymbol{x}_\ell$
- with techniques such as:
  - Random Forests (RFs)
  - Gradient Boosting Machines (GBMs)
  - Combined Actuarial Neural Networks (CANNs) with Artificial Neural Networks (ANNs) on top of e.g. GLM baseline.



Picture taken from Machine learning for everyone. In simple words. With real-world examples. Yes, again.

(Deep) Feed-forward neural nets:

- as in Richman and Wüthrich [2021]
- with country, gender, age and year as input features
- Lee & Carter and Li & Lee revisited with a network version + its calibration (no hidden layers)
- deep feed-forward neural nets are explored as well
- focus on improving a loss function, residual plots, . . .



(Deep) Recurrent neural nets (RNNs), Long short-term memory nets (LSTMs)

- as in Lindholm and Palmborg [2022], with focus on LSTM for the $\hat{\kappa}_t$'s
- as in Perla et al. [2021] where the $\mu_{x,t}^{(i)}$'s are modelled with (deep) RNNs and LSTMs for different populations (i) (i.e., genders and countries)
- as in Euthum et al. [2024] where the $\mu_{x,t}^{(i)}$'s are modelled with (deep) RNNs and LSTMs using inputs age, gender, year and deprivation group.

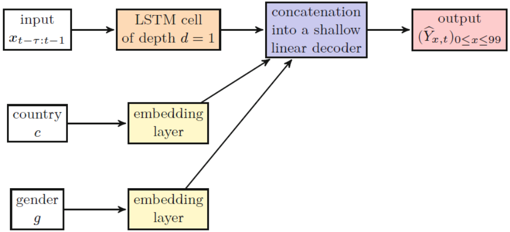


Fig. 8.13 LSTM architecture used to process the raw mortality rates $(M_{x,t})_{x,t}$

Picture from Wüthrich and Merz [2023].

(Deep) Convolutional neural nets (CNNs):

- as in Perla et al. [2021] where the $\mu_{x,t}^{(i)}$'s are modelled with CNN-type layers for the $\mu_{x,t}^{(i)}$s, combined with input features gender and region for population (i)
- as in Wang et al. [2021] who use CNNs with as input a 2-dimensional 'image' around $\mu_{x,t}$ with mortality information for ages in $[x - x_1, x + x_2]$, lagging k years $1 \leq k \leq s$.

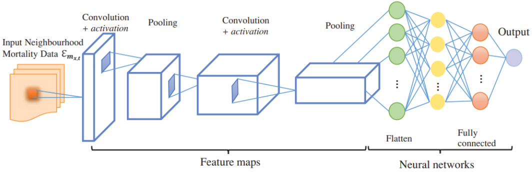


FIGURE 6: An illustrative CNN architecture in constructing neighbouring prediction models.

Today's approach

For details on analyzing mortality data with neural networks, we refer to:

- Lee and Carter go Machine Learning: Recurrent Neural Networks, a tutorial by the Swiss Actuarial Data Science working group
- the chapters on neural nets in Wüthrich and Merz [2023].

Today, we take a deep dive into a recent working paper by Jens, Katrien and Torsten:

- try to explain weekly death counts across European regions
- with a baseline Serfling model (e.g., alike EuroMoMo)
- combined with a (high-dimensional) set of weather and pollution features.

Short-term mortality and environmental variables

I will present today our JRSS A paper:

The association between environmental variables and short-term mortality: evidence from Europe

working paper on arxiv, code on GitHub, by Jens Robben (UvA, RCLR), Katrien Antonio and Torsten Kleinow (UvA, RCLR).

Research goals

1. Identify the primary environmental factors contributing to the estimation of mortality deviations from a pre-defined baseline level.

Research goals

1. Identify the primary environmental factors contributing to the estimation of mortality deviations from a pre-defined baseline level.

2. Investigate the marginal impact of an environmental factor on deviations from the mortality baseline level.

Research goals

1. Identify the primary environmental factors contributing to the estimation of mortality deviations from a pre-defined baseline level.

2. Investigate the marginal impact of an environmental factor on deviations from the mortality baseline level.

3. Investigate how environmental factors interact when modelling mortality rates. Are there harvesting effects present?

1. Identify the primary environmental factors contributing to the estimation of mortality deviations from a pre-defined baseline level.
2. Investigate the marginal impact of an environmental factor on deviations from the mortality baseline level.
3. Investigate how environmental factors interact when modelling mortality rates. Are there harvesting effects present?
4. Quantify the added contribution of environmental factors when aggregating the resulting estimated weekly mortality rates on an annual or country-level basis.

Short-term association between environmental variables and mortality rates

Introduction

Plan of attack

Using fine-grained open data, study the association between environmental factors and weekly mortality rates in European regions.

Plan of attack

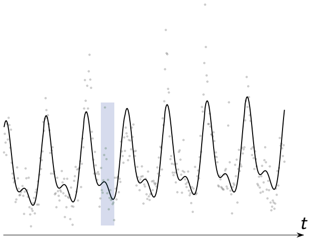
Using fine-grained open data, study the association between environmental factors and weekly mortality rates in European regions.

Proposed framework:

Using fine-grained open data, study the association between environmental factors and weekly mortality rates in European regions.

Proposed framework:

1. a weekly, region-specific baseline mortality model to capture overall seasonal trends.



Introduction

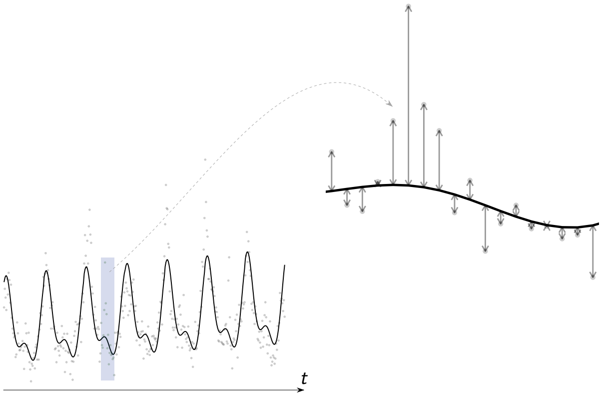
Plan of attack

Using fine-grained open data, study the association between environmental factors and weekly mortality rates in European regions.

Proposed framework:

1. a weekly, region-specific baseline mortality model to capture overall seasonal trends.

2. a predictive model to analyze mortality deviations from the baseline model using region-specific environmental factors.

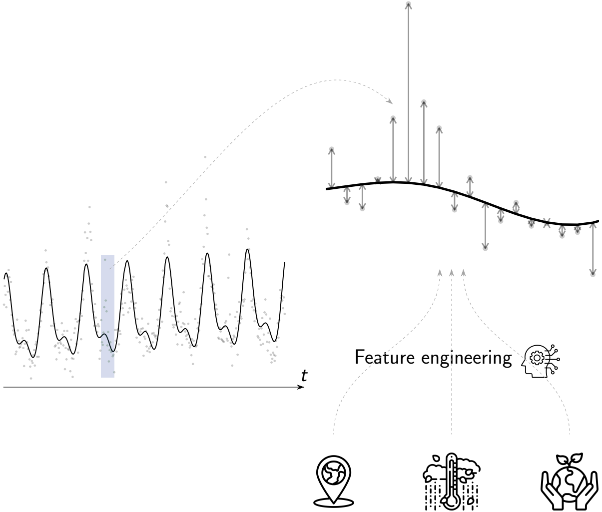


Using fine-grained open data, study the association between environmental factors and weekly mortality rates in European regions.

Proposed framework:

1. a weekly, region-specific baseline mortality model to capture overall seasonal trends.

2. a predictive model to analyze mortality deviations from the baseline model using region-specific environmental factors.



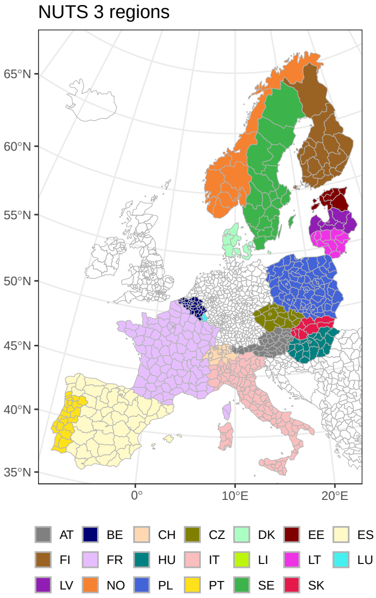
Data

Death counts

Eurostat: deaths by week, gender, 5-year age group and NUTS 3 region from 20 European countries throughout the years 2013-2019 ($>$ 500 regions).

Death counts

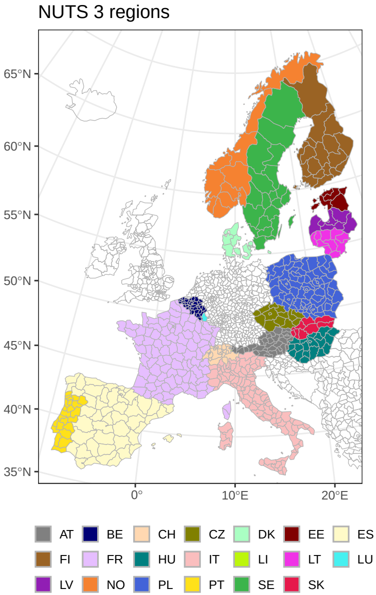
Eurostat: deaths by week, gender, 5-year age group and NUTS 3 region from 20 European countries throughout the years 2013-2019 ($>$ 500 regions).



Death counts

Eurostat: deaths by week, gender, 5-year age group and NUTS 3 region from 20 European countries throughout the years 2013-2019 ($>$ 500 regions).

Focus on old age group 65+, unisex.

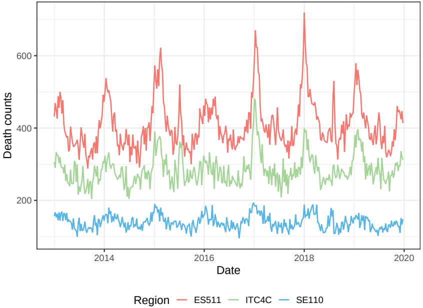


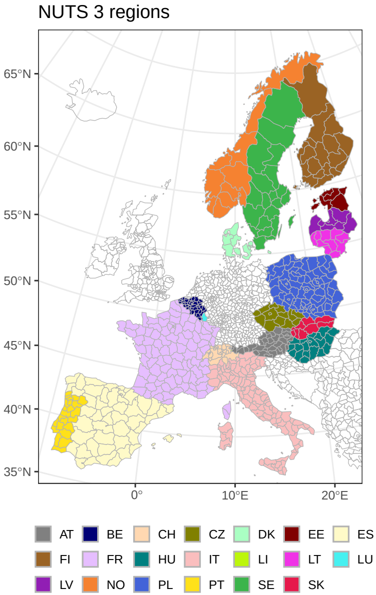
Death counts

Eurostat: deaths by week, gender, 5-year age group and NUTS 3 region from 20 European countries throughout the years 2013-2019 ($>$ 500 regions).

Focus on old age group 65+, unisex.

Seasonal trend:





Weather data

E-OBS land-only, gridded meteorological data for Europe from the Copernicus Climate Data Store.

Weather data

E-OBS land-only, gridded meteorological data for Europe from the Copernicus Climate Data Store.

Daily, high-resolution gridded dataset, defined on a grid with a spatial resolution of 0.10° (≈ 9 km).

Weather data

E-OBS land-only, gridded meteorological data for Europe from the Copernicus Climate Data Store.

Daily, high-resolution gridded dataset, defined on a grid with a spatial resolution of 0.10° (≈ 9 km).

Weather factors:

- `Tmax`: daily maximum temperature
- `Tmin`: daily minimum temperature
- `Hum`: daily average relative humidity
- `Rain`: total daily precipitation
- `Wind`: daily average wind speed

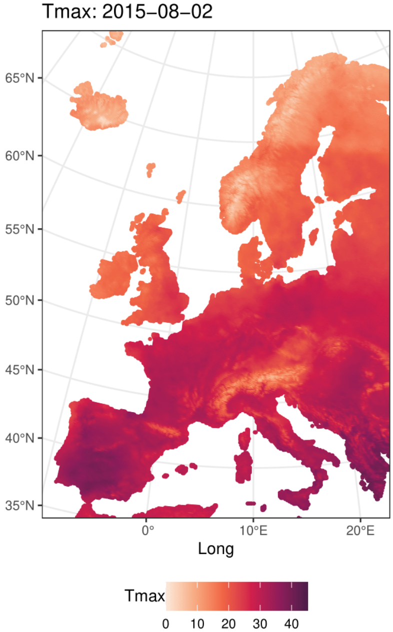
Weather data

E-OBS land-only, gridded meteorological data for Europe from the Copernicus Climate Data Store.

Daily, high-resolution gridded dataset, defined on a grid with a spatial resolution of 0.10° (≈ 9 km).

Weather factors:

- `Tmax`: daily maximum temperature
- `Tmin`: daily minimum temperature
- `Hum`: daily average relative humidity
- `Rain`: total daily precipitation
- `Wind`: daily average wind speed



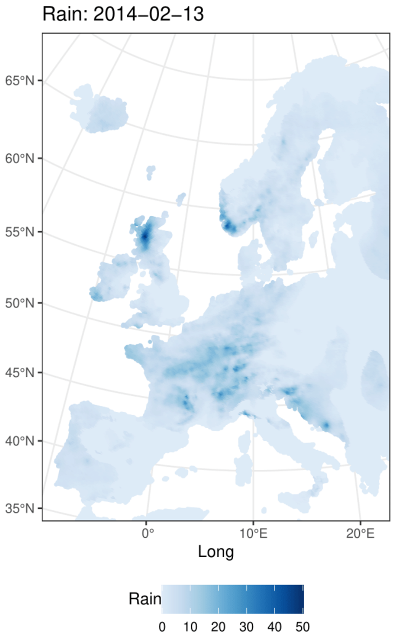
Weather data

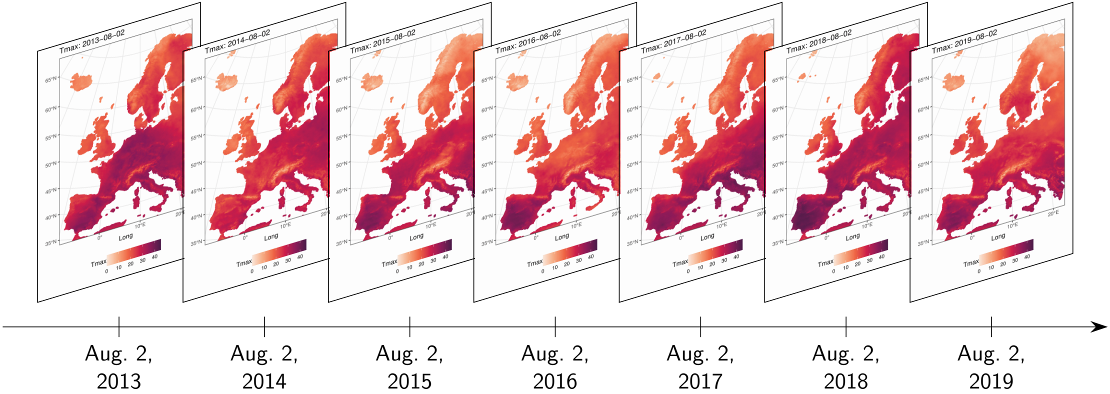
E-OBS land-only, gridded meteorological data for Europe from the Copernicus Climate Data Store.

Daily, high-resolution gridded dataset, defined on a grid with a spatial resolution of 0.10° (≈ 9 km).

Weather factors:

- `Tmax`: daily maximum temperature
- `Tmin`: daily minimum temperature
- `Hum`: daily average relative humidity
- `Rain`: total daily precipitation
- `Wind`: daily average wind speed




Tmax: 2013–08–02
Tmax: 2014–08–02
Tmax: 2015–08–02
Tmax: 2016–08–02
Tmax: 2017–08–02
Tmax: 2018–08–02
Tmax: 2019–08–02
Long
Tmax
0 10 20 30 40
Aug. 2, 2013
Aug. 2, 2014
Aug. 2, 2015
Aug. 2, 2016
Aug. 2, 2017
Aug. 2, 2018
Aug. 2, 2019

Air pollution data

CAMS European air quality reanalyses dataset from the Copernicus Atmosphere Monitoring Service (land + sea).

Air pollution data

CAMS European air quality reanalyses dataset from the Copernicus Atmosphere Monitoring Service (land + sea).

Hourly, high-resolution air quality reanalyses, defined on a grid with a spatial resolution of 0.10° (≈ 9 km).

Air pollution data

CAMS European air quality reanalyses dataset from the Copernicus Atmosphere Monitoring Service (land + sea).

Hourly, high-resolution air quality reanalyses, defined on a grid with a spatial resolution of 0.10° ($\approx$ 9 km).

Air pollutants ($\mu g/m^3$):

- `O3`: hourly ozone levels.
- `NO2`: hourly nitrogen dioxide levels.
- `PM10`: hourly particular matter (10 microns wide).
- `PM2.5`: hourly particular matter (2.5 microns wide).

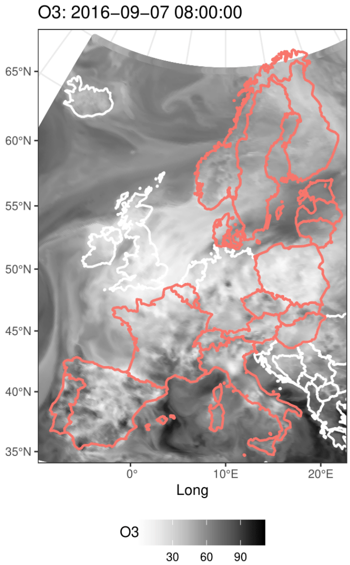
Air pollution data

CAMS European air quality reanalyses dataset from the Copernicus Atmosphere Monitoring Service (land + sea).

Hourly, high-resolution air quality reanalyses, defined on a grid with a spatial resolution of 0.10° ($\approx$ 9 km).

Air pollutants ($\mu g/m^3$):

- `O3`: hourly ozone levels.
- `NO2`: hourly nitrogen dioxide levels.
- `PM10`: hourly particular matter (10 microns wide).
- `PM2.5`: hourly particular matter (2.5 microns wide).



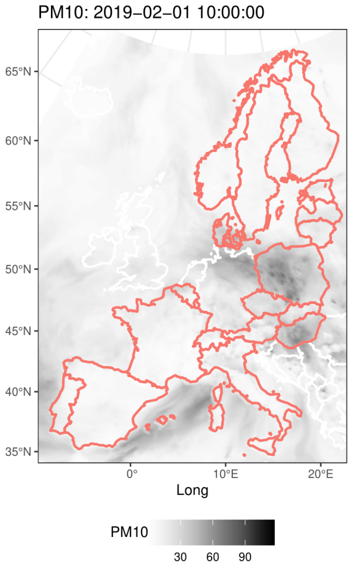
Air pollution data

CAMS European air quality reanalyses dataset from the Copernicus Atmosphere Monitoring Service (land + sea).

Hourly, high-resolution air quality reanalyses, defined on a grid with a spatial resolution of 0.10° ($\approx$ 9 km).

Air pollutants ($\mu g/m^3$):

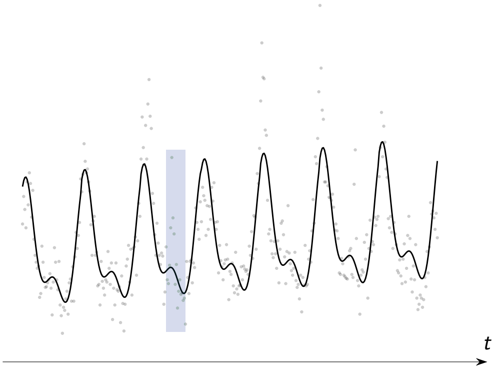
- `O3`: hourly ozone levels.
- `NO2`: hourly nitrogen dioxide levels.
- `PM10`: hourly particular matter (10 microns wide).
- `PM2.5`: hourly particular matter (2.5 microns wide).



Model specification

Weekly, region-specific baseline mortality model

A weekly, region-specific baseline mortality model to capture the overall seasonal trends in the considered regions.



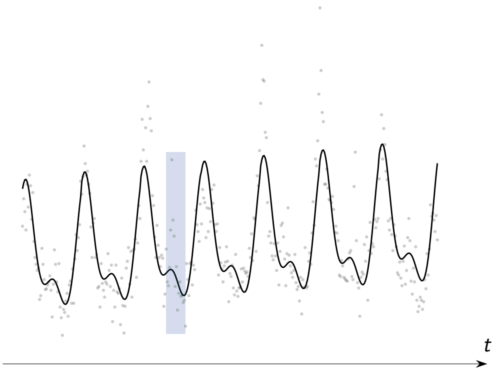
Weekly, region-specific baseline mortality model

A weekly, region-specific baseline mortality model to capture the overall seasonal trends in the considered regions.

Incorporate seasonality through Fourier terms Serfling [1963]:

$$D_{t,w}^{(r)} \sim \text{Poisson}\left(E_{t,w}^{(r)} \cdot \mu_{t,w}^{(r)}\right),$$

$$\log \mu_{t,w}^{(r)} = \beta_0^{(r)} + \beta_1^{(r)} t + \beta_2^{(r)} \sin\left(\frac{2\pi w}{52}\right) + \beta_3^{(r)} \cos\left(\frac{2\pi w}{52}\right) + \beta_4^{(r)} \sin\left(\frac{2\pi w}{26}\right) + \beta_5^{(r)} \cos\left(\frac{2\pi w}{26}\right).$$



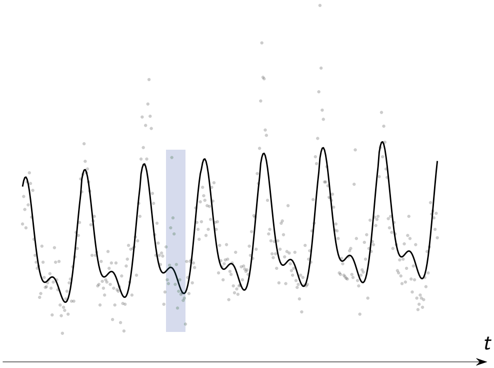
Weekly, region-specific baseline mortality model

A weekly, region-specific baseline mortality model to capture the overall seasonal trends in the considered regions.

Incorporate seasonality through Fourier terms Serfling [1963]:

$$D_{t,w}^{(r)} \sim \text{Poisson}\left(E_{t,w}^{(r)} \cdot \mu_{t,w}^{(r)}\right),$$

$$\log \mu_{t,w}^{(r)} = \beta_0^{(r)} + \beta_1^{(r)} t + \beta_2^{(r)} \sin\left(\frac{2\pi w}{52}\right) + \beta_3^{(r)} \cos\left(\frac{2\pi w}{52}\right) + \beta_4^{(r)} \sin\left(\frac{2\pi w}{26}\right) + \beta_5^{(r)} \cos\left(\frac{2\pi w}{26}\right).$$



Region-specific population exposures $E_{t,w}^{(r)}$ from Eurostat.

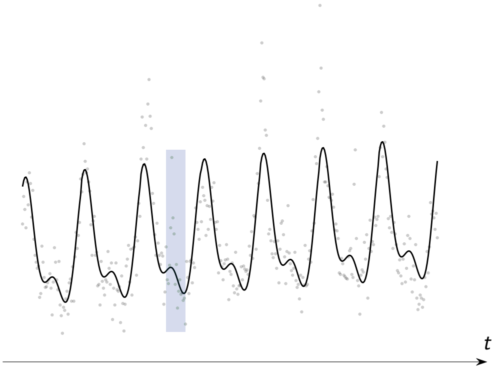
Weekly, region-specific baseline mortality model

A weekly, region-specific baseline mortality model to capture the overall seasonal trends in the considered regions.

Incorporate seasonality through Fourier terms Serfling [1963]:

$$D_{t,w}^{(r)} \sim \text{Poisson}\left(E_{t,w}^{(r)} \cdot \mu_{t,w}^{(r)}\right),$$

$$\log \mu_{t,w}^{(r)} = \beta_0^{(r)} + \beta_1^{(r)} t + \beta_2^{(r)} \sin\left(\frac{2\pi w}{52}\right) + \beta_3^{(r)} \cos\left(\frac{2\pi w}{52}\right) + \beta_4^{(r)} \sin\left(\frac{2\pi w}{26}\right) + \beta_5^{(r)} \cos\left(\frac{2\pi w}{26}\right).$$

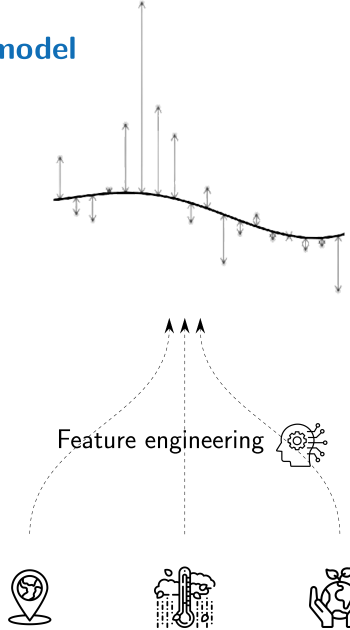


Region-specific population exposures $E_{t,w}^{(r)}$ from Eurostat.

Estimated baseline death counts: $\hat{b}_{t,w}^{(r)} := E_{t,w}^{(r)} \cdot \hat{\mu}_{t,w}^{(r)}$.

Modelling mortality deviations from the baseline model

Explain observed deviations from the baseline deaths using region-specific environmental features.



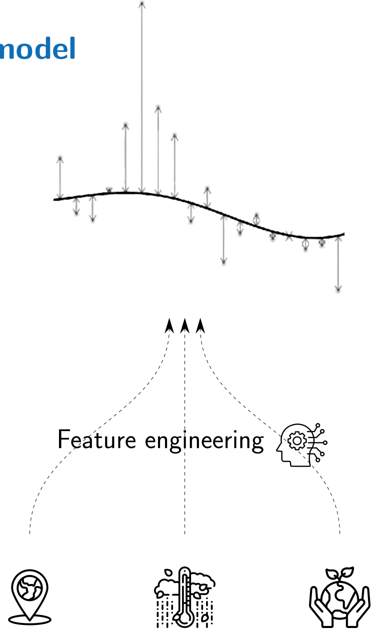
Modelling mortality deviations from the baseline model

Explain observed deviations from the baseline deaths using region-specific environmental features.

Fix estimated baseline deaths and impose distributional assumption:

$$D_{t,w}^{(r)} \sim \text{Poisson}\left(\hat{b}_{t,w}^{(r)}\,\phi_{t,w}^{(r)}\right),$$

$$\phi_{t,w}^{(r)} = f\Big(\texttt{long}^{(r)},\ \texttt{lat}^{(r)},\ \texttt{season}_{t,w},\ \boldsymbol{c}_{t,w}^{(r)},\ \boldsymbol{e}_{t,w}^{(r)},$$
$$l^1\left(\boldsymbol{c}_{t,w}^{(r)}\right),\ l^1\left(\boldsymbol{e}_{t,w}^{(r)}\right),\ \dots,\ l^s\left(\boldsymbol{c}_{t,w}^{(r)}\right),\ l^s\left(\boldsymbol{e}_{t,w}^{(r)}\right)\Big).$$



Modelling mortality deviations from the baseline model

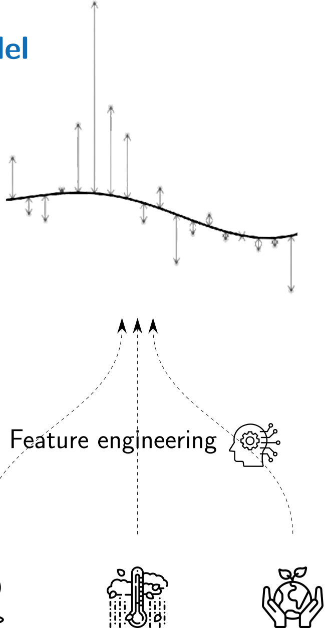
Explain observed deviations from the baseline deaths using region-specific environmental features.

Fix estimated baseline deaths and impose distributional assumption:

$$D^{(r)}_{t,w} \sim \text{Poisson}\left(\hat{b}^{(r)}_{t,w}\, \phi^{(r)}_{t,w}\right),$$

$$\phi^{(r)}_{t,w} = f\Big(\texttt{long}^{(r)}, \texttt{lat}^{(r)}, \texttt{season}_{t,w}, \boldsymbol{c}^{(r)}_{t,w}, \boldsymbol{e}^{(r)}_{t,w}, l^1\left(\boldsymbol{c}^{(r)}_{t,w}\right), l^1\left(\boldsymbol{e}^{(r)}_{t,w}\right), \ldots, l^s\left(\boldsymbol{c}^{(r)}_{t,w}\right), l^s\left(\boldsymbol{e}^{(r)}_{t,w}\right)\Big).$$

$f(\cdot)$ is a selected predictive modelling technique.



Explain observed deviations from the baseline deaths using region-specific environmental features.

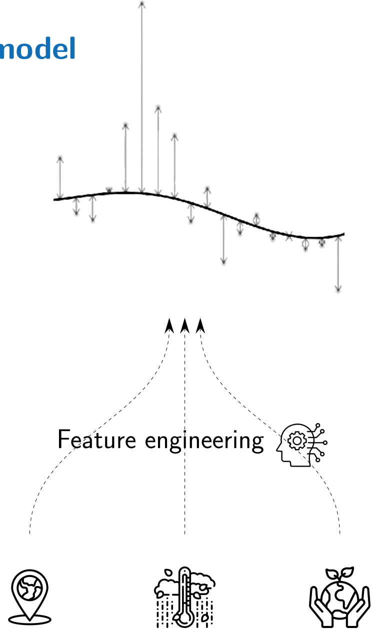
Fix estimated baseline deaths and impose distributional assumption:

$$D^{(r)}_{t,w} \sim \text{Poisson}\left(\hat{b}^{(r)}_{t,w}\, \phi^{(r)}_{t,w}\right),$$

$$\phi^{(r)}_{t,w} = f\Big(\texttt{long}^{(r)},\ \texttt{lat}^{(r)},\ \texttt{season}_{t,w},\ \boldsymbol{c}^{(r)}_{t,w},\ \boldsymbol{e}^{(r)}_{t,w},\ l^{1}\left(\boldsymbol{c}^{(r)}_{t,w}\right),\ l^{1}\left(\boldsymbol{e}^{(r)}_{t,w}\right),\ \ldots,\ l^{s}\left(\boldsymbol{c}^{(r)}_{t,w}\right),\ l^{s}\left(\boldsymbol{e}^{(r)}_{t,w}\right)\Big).$$

$f(\cdot)$ is a selected predictive modelling technique.

We opt for a machine learning model.



Model calibration

Fit a Poisson GLM jointly across all regions, add a smoothness penalty:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left(-\log \mathcal{L}_P(\boldsymbol{\beta}) + \sum_{p=0}^{5} \lambda_p \boldsymbol{\beta}_p^T \boldsymbol{S} \boldsymbol{\beta}_p \right),$$

Fit a Poisson GLM jointly across all regions, add a smoothness penalty:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left(-\log \mathcal{L}_P(\boldsymbol{\beta}) + \sum_{p=0}^{5} \lambda_p \boldsymbol{\beta}_p^T \boldsymbol{S} \boldsymbol{\beta}_p \right),$$

where:

$\boldsymbol{\beta}$: parameter vector

$\log \mathcal{L}_P(\boldsymbol{\beta})$: Poisson log-likelihood

$\boldsymbol{\beta}_p^T \boldsymbol{S} \boldsymbol{\beta}_p$: penalizes sum of squared differences between coefs of adjacent regions

λ_p: smoothing or penalty parameter.

Fit a Poisson GLM jointly across all regions, add a smoothness penalty:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left(-\log \mathcal{L}_P(\boldsymbol{\beta}) + \sum_{p=0}^{5} \lambda_p \boldsymbol{\beta}_p^T \boldsymbol{S} \boldsymbol{\beta}_p \right),$$

where:

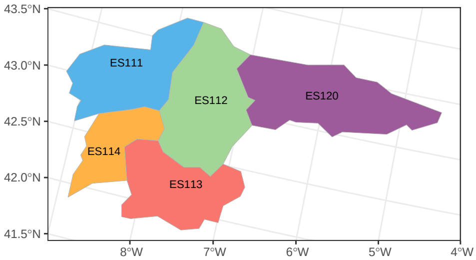
$\boldsymbol{\beta}$: parameter vector

$\log \mathcal{L}_P(\boldsymbol{\beta})$: Poisson log-likelihood

$\boldsymbol{\beta}_p^T \boldsymbol{S} \boldsymbol{\beta}_p$: penalizes sum of squared differences between coefs of adjacent regions

λ_p: smoothing or penalty parameter.

Example (5 Spanish NUTS 3 regions):



Calibrating the baseline model

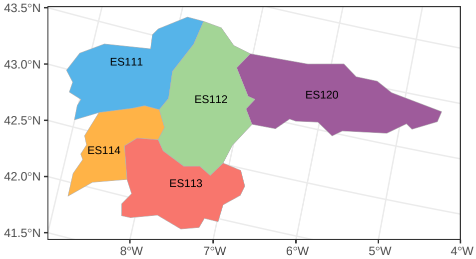
Fit a Poisson GLM jointly across all regions, add a smoothness penalty:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \left(-\log \mathcal{L}_P(\boldsymbol{\beta}) + \sum_{p=0}^{5} \lambda_p \boldsymbol{\beta}_p^T \boldsymbol{S} \boldsymbol{\beta}_p \right),$$

where:

- $\boldsymbol{\beta}$: parameter vector
- $\log \mathcal{L}_P(\boldsymbol{\beta})$: Poisson log-likelihood
- $\boldsymbol{\beta}_p^T \boldsymbol{S} \boldsymbol{\beta}_p$: penalizes sum of squared differences between coefs of adjacent regions
- λ_p: smoothing or penalty parameter.

Example (5 Spanish NUTS 3 regions):



Penalty matrix $\boldsymbol{S}$:

$$\begin{array}{c} \\ ES111 \\ ES112 \\ ES113 \\ ES114 \\ ES120 \end{array} \begin{array}{c} \begin{array}{ccccc} ES111 & ES112 & ES113 & ES114 & ES120 \end{array} \\ \begin{pmatrix} 2 & -1 & 0 & -1 & 0 \\ -1 & 4 & -1 & -1 & -1 \\ 0 & -1 & 2 & -1 & 0 \\ -1 & -1 & -1 & 3 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Calibrating the mortality deviations model

XGBoost: flexible and efficient implementation of gradient boosting.

Calibrating the mortality deviations model

XGBoost: flexible and efficient implementation of gradient boosting.

Tuning parameters:

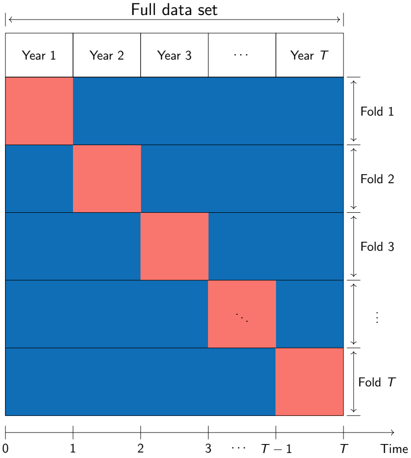
`nrounds`:	number of boosting iterations.
`eta`:	learning rate.
`max_depth`:	the maximum depth of a tree.
`subsample`:	subsample ratio of the training data.
`colsample_bytree`:	subsample ratio of the features.

XGBoost: flexible and efficient implementation of gradient boosting.

Tuning parameters:

`nrounds`:	number of boosting iterations.
`eta`:	learning rate.
`max_depth`:	the maximum depth of a tree.
`subsample`:	subsample ratio of the training data.
`colsample_bytree`:	subsample ratio of the features.

Parameter tuning with T-fold cross-validation.



Calibrating the mortality deviations model

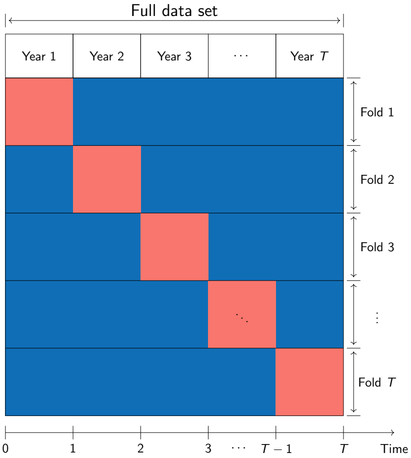
XGBoost: flexible and efficient implementation of gradient boosting.

Tuning parameters:

`nrounds`:	number of boosting iterations.
`eta`:	learning rate.
`max_depth`:	the maximum depth of a tree.
`subsample`:	subsample ratio of the training data.
`colsample_bytree`:	subsample ratio of the features.

Parameter tuning with T-fold cross-validation.

Calibrate XGBoost model on entire training data with optimal parameter configuration.



XGBoost: flexible and efficient implementation of gradient boosting.

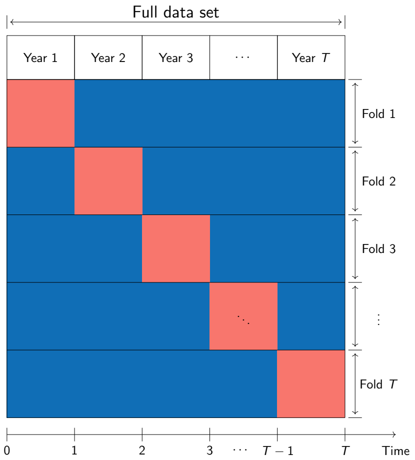
Tuning parameters:

`nrounds`:	number of boosting iterations.
`eta`:	learning rate.
`max_depth`:	the maximum depth of a tree.
`subsample`:	subsample ratio of the training data.
`colsample_bytree`:	subsample ratio of the features.

Parameter tuning with T-fold cross-validation.

Calibrate XGBoost model on entire training data with optimal parameter configuration.

Interpretation tools to gain insights: VIP, ALE.



Feature engineering

Feature engineering

Motivation

Differences in spatial and temporal dimension across the data sources:

- data on death counts: weekly, NUTS 3 scale.
- environmental data: hourly or daily time scale, spatial grid.

Feature engineering

Motivation

Differences in spatial and temporal dimension across the data sources:

- data on death counts: weekly, NUTS 3 scale.
- environmental data: hourly or daily time scale, spatial grid.

Goal of feature engineering:

- convert the temporal and spatial dimensions of the environmental data into features on a weekly, NUTS 3 scale

Feature engineering

Motivation

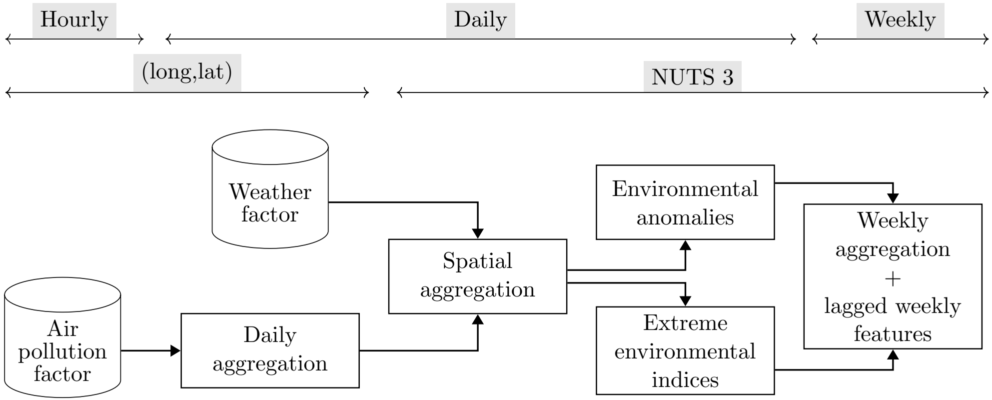
Differences in spatial and temporal dimension across the data sources:

- data on death counts: weekly, NUTS 3 scale.
- environmental data: hourly or daily time scale, spatial grid.

Goal of feature engineering:

- convert the temporal and spatial dimensions of the environmental data into features on a weekly, NUTS 3 scale
- create features that measure deviations from baseline conditions from environmental data to explain excess or deficit mortality (other ideas welcome!).

Flow chart


Hourly
Daily
Weekly
(long,lat)
NUTS 3
Weather factor
Air pollution factor
Daily aggregation
Spatial aggregation
Environmental anomalies
Extreme environmental indices
Weekly aggregation
+
lagged weekly features

Extreme environmental indices

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

Extreme environmental indices

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

Calculate region-specific 5% and 95% quantiles of the daily historical temperature or air pollution observations over the years 2013-2019.

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

Calculate region-specific 5% and 95% quantiles of the daily historical temperature or air pollution observations over the years 2013-2019.

Define extreme high temperature index (hot-day index):

$$\text{T.ind}_{t,w,d}^{(r,95\%)} = \mathbb{1}\left\{\texttt{Tmax}_{t,w,d}^{(r)} \geq q_{\texttt{Tmax}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tavg}_{t,w,d}^{(r)} \geq q_{\texttt{Tavg}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tmin}_{t,w,d}^{(r)} \geq q_{\texttt{Tmin}}^{(r,95\%)}\right\}.$$

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

Calculate region-specific 5% and 95% quantiles of the daily historical temperature or air pollution observations over the years 2013-2019.

Define extreme high temperature index (hot-day index):

$$\text{T.ind}_{t,w,d}^{(r,95\%)} = \mathbb{1}\left\{\texttt{Tmax}_{t,w,d}^{(r)} \geq q_{\texttt{Tmax}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tavg}_{t,w,d}^{(r)} \geq q_{\texttt{Tavg}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tmin}_{t,w,d}^{(r)} \geq q_{\texttt{Tmin}}^{(r,95\%)}\right\}.$$

Index values: 0-3, indicating the severity of hot days.

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

Calculate region-specific 5% and 95% quantiles of the daily historical temperature or air pollution observations over the years 2013-2019.

Define extreme high temperature index (hot-day index):

$$\text{T.ind}_{t,w,d}^{(r,95\%)} = \mathbb{1}\left\{\texttt{Tmax}_{t,w,d}^{(r)} \geq q_{\texttt{Tmax}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tavg}_{t,w,d}^{(r)} \geq q_{\texttt{Tavg}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tmin}_{t,w,d}^{(r)} \geq q_{\texttt{Tmin}}^{(r,95\%)}\right\}.$$

Index values: 0-3, indicating the severity of hot days.

Similar extreme indices are created for the other daily weather and air pollution factors.

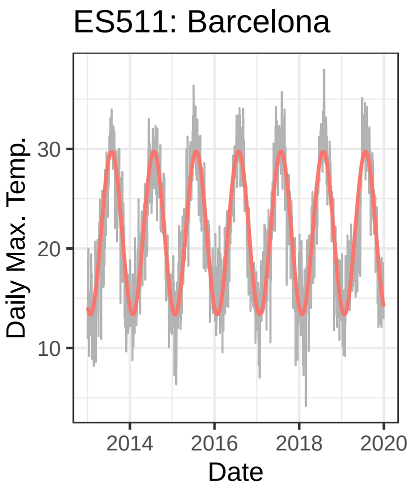
Environmental anomalies

Create features that quantify deviations from typical, baseline conditions for each day throughout the year.

Create features that quantify deviations from typical, baseline conditions for each day throughout the year.

Robust linear regression to capture baseline:

$$\tilde{x}^{(r)}_{t,w,d} = \alpha^{(r)}_0 + \alpha^{(r)}_1 \sin\left(\frac{2\pi d}{365.25}\right) + \alpha^{(r)}_2 \cos\left(\frac{2\pi d}{365.25}\right) + \epsilon^{(r)}_{t,w,d},$$



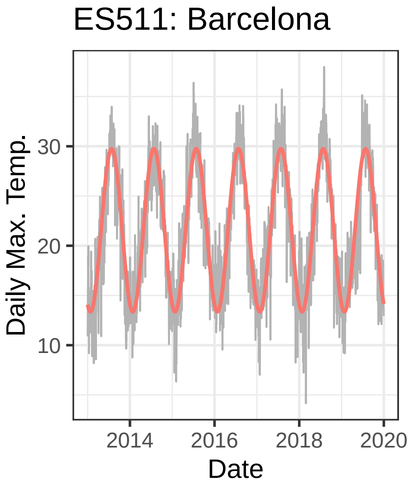
Create features that quantify deviations from typical, baseline conditions for each day throughout the year.

Robust linear regression to capture baseline:

$$\tilde{x}^{(r)}_{t,w,d} = \alpha^{(r)}_0 + \alpha^{(r)}_1 \sin\left(\frac{2\pi d}{365.25}\right) + \alpha^{(r)}_2 \cos\left(\frac{2\pi d}{365.25}\right) + \epsilon^{(r)}_{t,w,d},$$

In the paper, we work with excesses or deviations from the baseline (anomalies):

$$\tilde{x}^{(r)}_{t,w,d} - \hat{\tilde{x}}^{(r)}_{t,w,d}$$



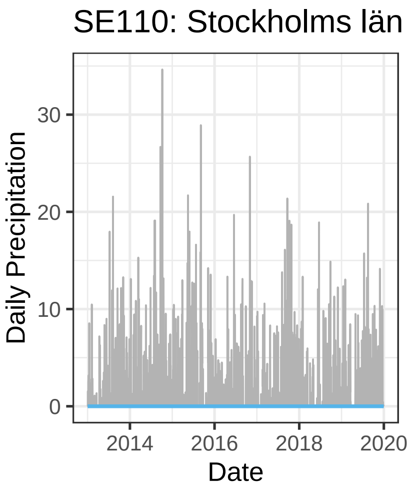
Create features that quantify deviations from typical, baseline conditions for each day throughout the year.

Robust linear regression to capture baseline:

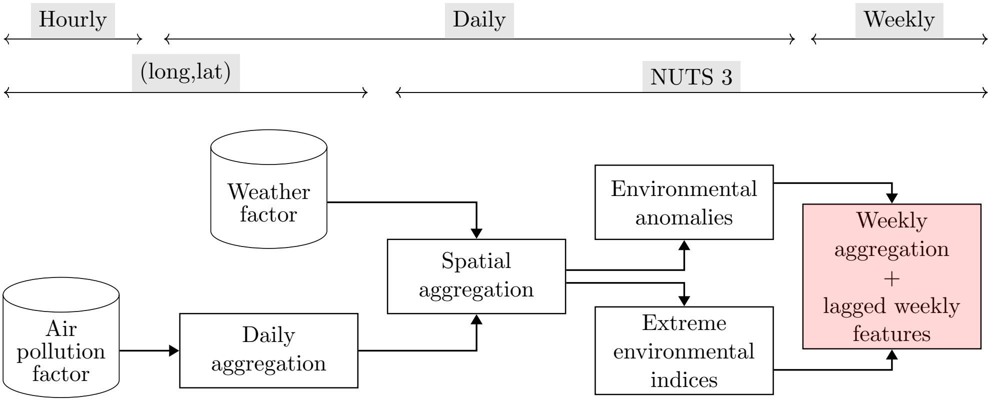
$$\tilde{x}^{(r)}_{t,w,d} = \alpha^{(r)}_0 + \alpha^{(r)}_1 \sin\left(\frac{2\pi w}{365.25}\right) + \alpha^{(r)}_2 \cos\left(\frac{2\pi w}{365.25}\right) + \epsilon^{(r)}_{t,w,d},$$

In the paper we work with excesses or deviations from the baseline (anomalies):

$$\tilde{x}^{(r)}_{t,w,d} - \hat{\tilde{x}}^{(r)}_{t,w,d}$$

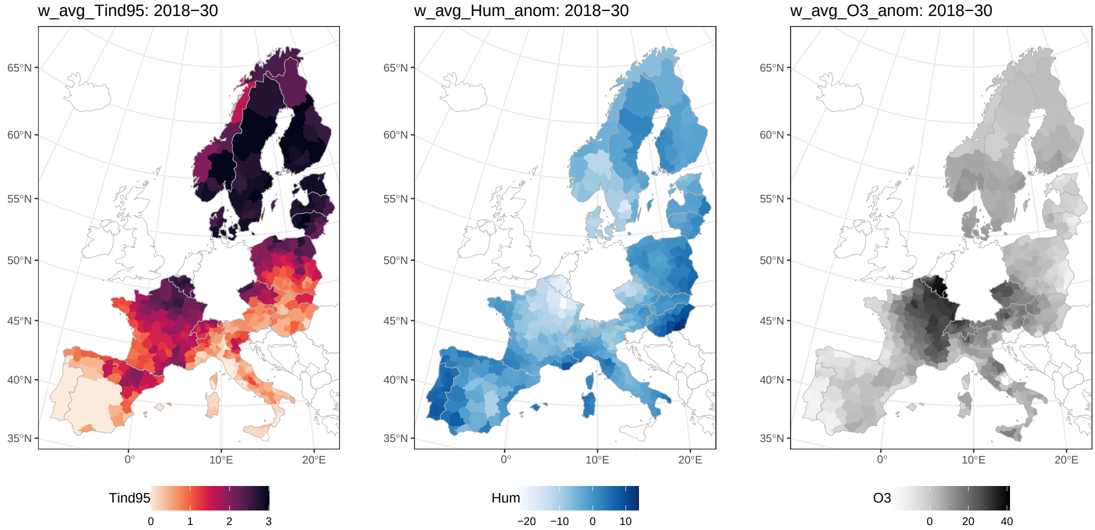


Flow chart



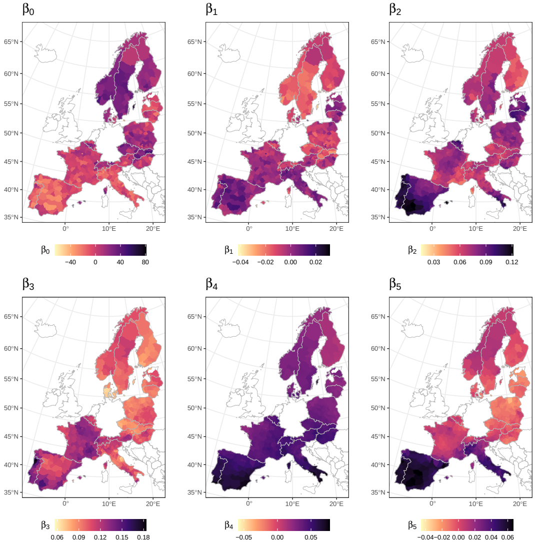
Hourly
Daily
Weekly
(long,lat)
NUTS 3
Weather factor
Air pollution factor
Daily aggregation
Spatial aggregation
Environmental anomalies
Extreme environmental indices
Weekly aggregation + lagged weekly features

Weekly aggregation



Calibration results

Baseline model



Machine learning model

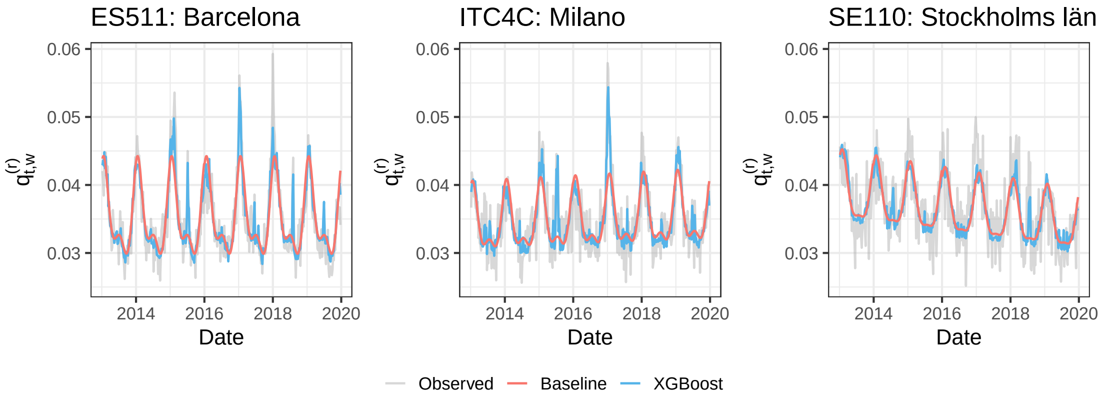
Input features: longitude-latitude coordinates, season, (one-week lagged) environmental anomalies and extreme indices.

Tuning by 7-fold cross validation over the years 2013-2019 using an extensive tuning grid.

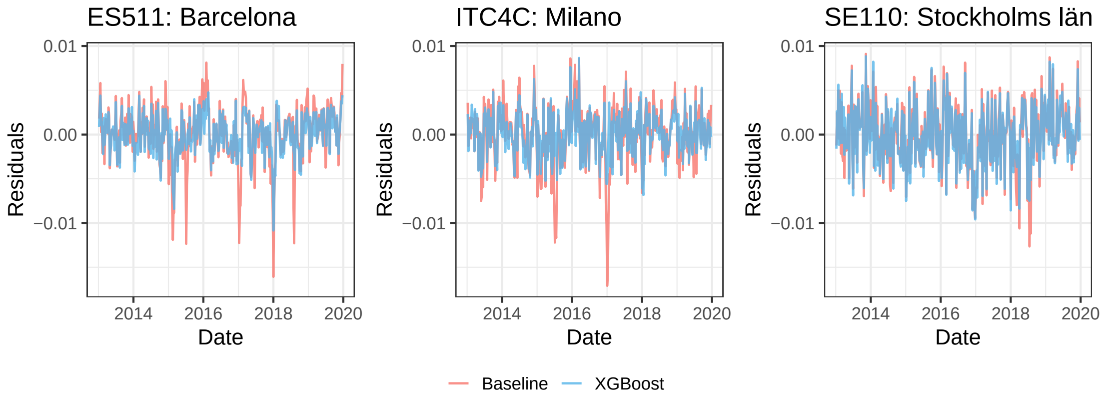
Tuning parameters: `nrounds` (490), `eta` (0.01), `min_child_weight` (1000), `max.depth` (7), `subsample` (0.75), `colsample_bytree` (0.50).

Insights in the machine-learning model

Observed and estimated mortality rates (baseline + XGBoost):



Residuals of the estimated weekly mortality rates (baseline + XGBoost):



Feature importance

Machine learning techniques perform automatic feature selection.

Feature importance

Machine learning techniques perform automatic feature selection.

Which features do significantly contribute to the predictions?

Machine learning techniques perform automatic feature selection.

Which features do significantly contribute to the predictions?

We calculate the feature importance of each feature X_l as:

$$\mathcal{V}_{\text{imp}}(X_l) = \frac{1}{\texttt{nrounds}} \sum_{n=1}^{\texttt{nrounds}} \Delta\mathcal{L}_n(X_l),$$

with $\Delta\mathcal{L}_n(X_l)$ the total reduction in the Poisson loss function, caused by splits associated to feature X_l in the tree built during iteration n of the XGBoost algorithm.

Feature importance

Machine learning techniques perform automatic feature selection.

Which features do significantly contribute to the predictions?

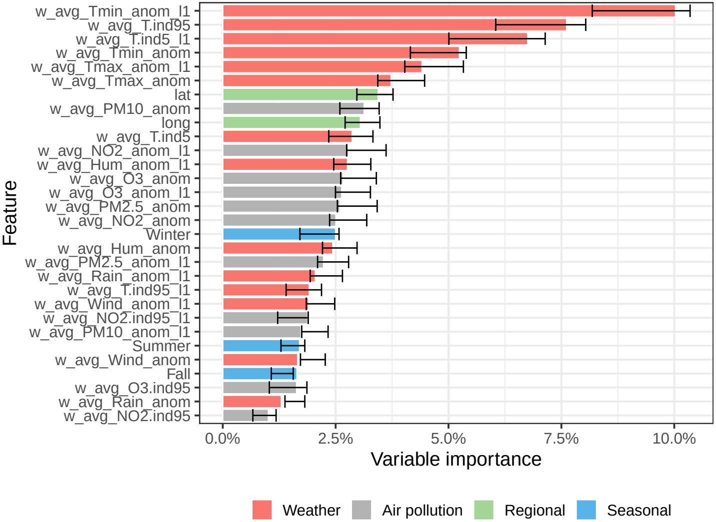
We calculate the feature importance of each feature X_l as:

$$\mathcal{V}_{\text{imp}}(X_l) = \frac{1}{\texttt{nrounds}} \sum_{n=1}^{\texttt{nrounds}} \Delta\mathcal{L}_n(X_l),$$

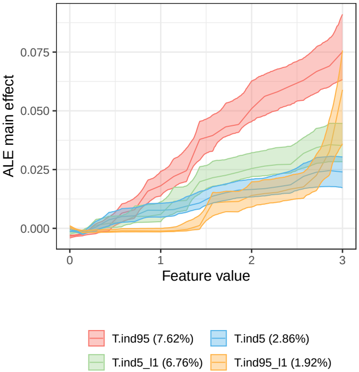
with $\Delta\mathcal{L}_n(X_l)$ the total reduction in the Poisson loss function, caused by splits associated to feature X_l in the tree built during iteration n of the XGBoost algorithm.

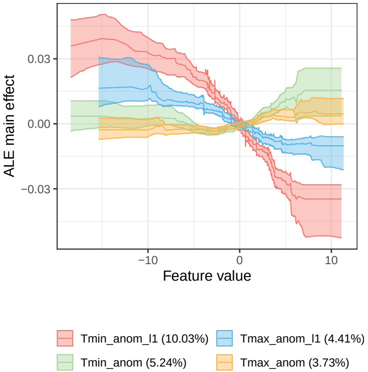
Features with a high importance appear often and high in the tree.

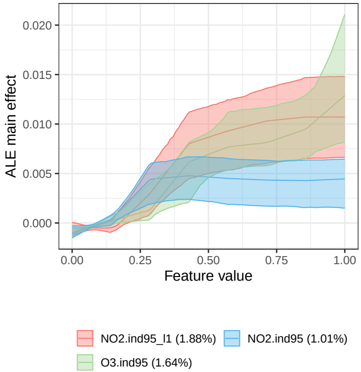
Feature importance



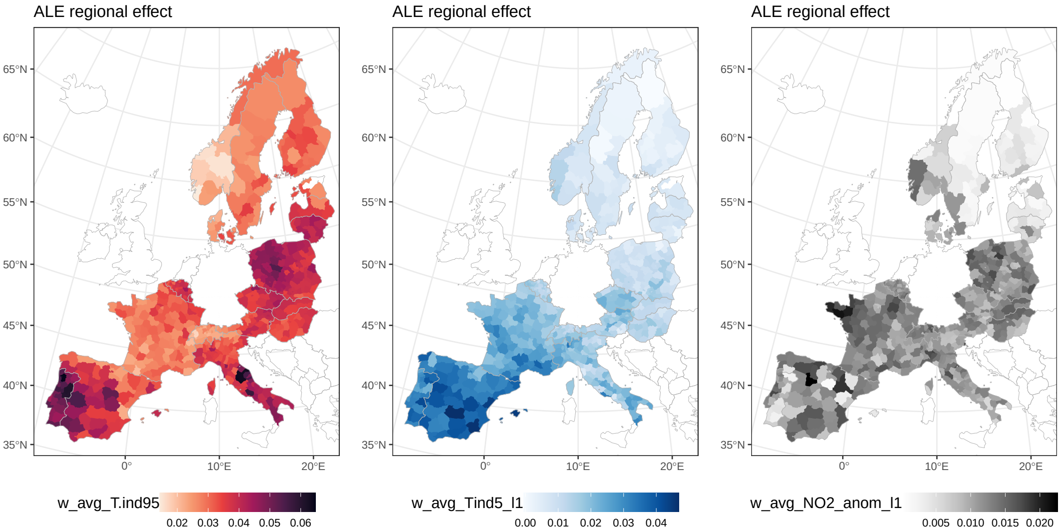
ALE main effects

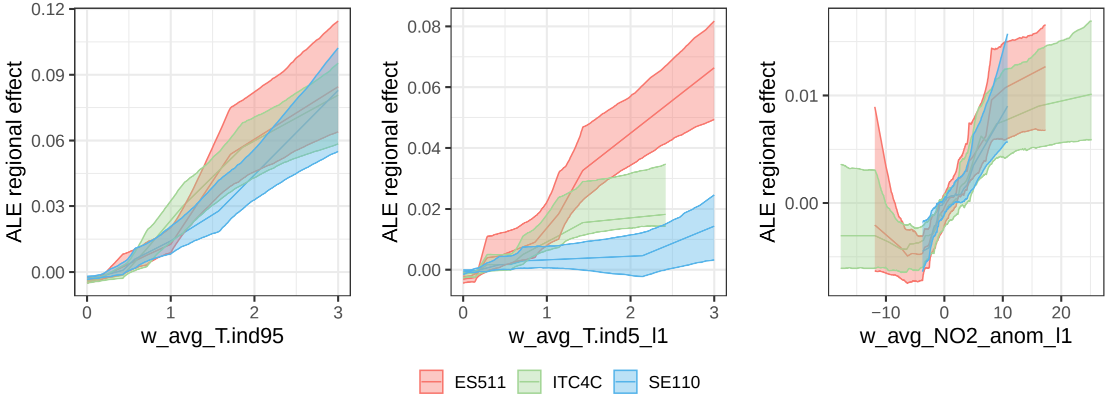




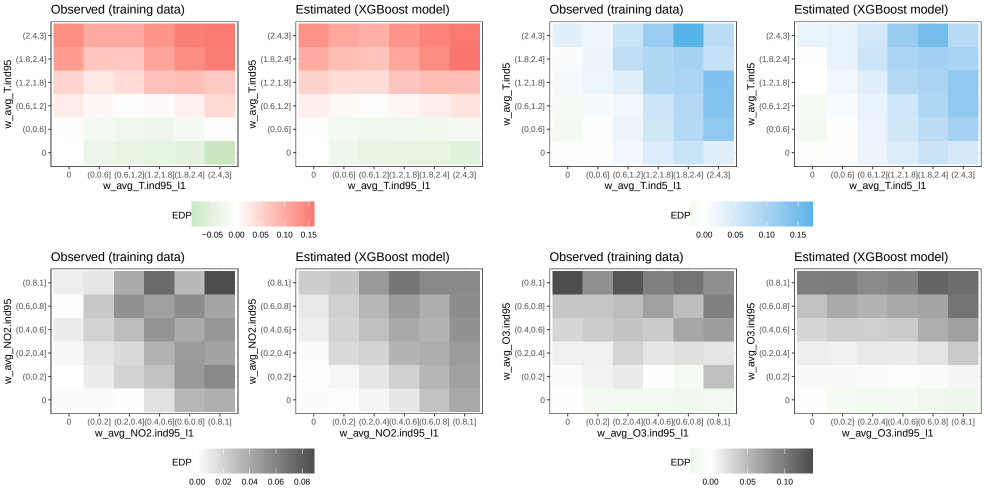


ALE regional effects




ALE regional effect
0.12
0.09
0.06
0.03
0.00
0 1 2 3
w_avg_T.ind95
0.08
0.06
0.04
0.02
0.00
w_avg_T.ind5_l1
0.01
0.00
−10 0 10 20
w_avg_NO2_anom_l1
ES511
ITC4C
SE110

Harvesting effects



Wrap up

We have (multiple) additional visuals and analyses in the working paper, as well as a detailed discussion of related literature.

Wrap up

We have (multiple) additional visuals and analyses in the working paper, as well as a detailed discussion of related literature.

Limitation: focus on short-term associations only!

Wrap up

We have (multiple) additional visuals and analyses in the working paper, as well as a detailed discussion of related literature.

Limitation: focus on short-term associations only!

Exciting opportunities ahead using sophisticated learning methods and fine-grained (open) data.

We have (multiple) additional visuals and analyses in the working paper, as well as a detailed discussion of related literature.

Limitation: focus on short-term associations only!

Exciting opportunities ahead using sophisticated learning methods and fine-grained (open) data.

However, feature engineering, (proper) interpretation tools, interplay of more traditional actuarial + statistical learning and sophisticated machine learning methods are key!

M. Euthum, M. Scherer, and F. Ungolo. A neural network approach for the mortality analysis of multiple populations: a case study on data of the italian population. *European Actuarial Journal*, 2024.

Mathias Lindholm and Lisa Palmborg. Efficient use of data for lstm mortality forecasting. *European Actuarial Journal*, 12:749–778, 2022.

Francesca Perla, Ronald Richman, Salvatore Scognamiglio, and Mario V. Wüthrich. Time-series forecasting of mortality rates using deep learning. *Scandinanivan Actuarial Journal*, 7(2): 572–598, 2021.

Ronald Richman and Mario V. Wüthrich. A neural network extension of the lee–carter model to multiple populations. *Annals of Actuarial Science*, 15(2):346–366, 2021.

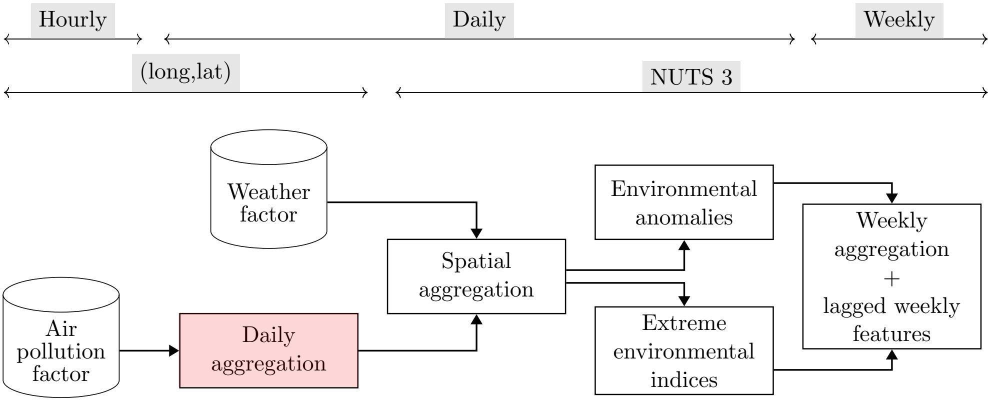
Robert E Serfling. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public health reports*, 78(6):494, 1963.

Chou-Wen Wang, Jinggong Zhang, and Wenjun Zhu. Neighbouring prediction for mortality. *ASTIN Bulletin*, 51(3):689–718, 2021.

References II

Mario Wüthrich and Michael Merz. *Statistical foundations of actuarial learning and its applications*. Springer, 2023.

Flow chart



Hourly
Daily
Weekly
(long,lat)
NUTS 3
Weather factor
Air pollution factor
Daily aggregation
Spatial aggregation
Environmental anomalies
Extreme environmental indices
Weekly aggregation + lagged weekly features

Daily aggregation

Consider an air pollution factor and denote its concentration at hour h of day d in week w of year t and located at longitude-latitude coordinates (long,lat) as $x_{t,w,d,h}^{(\text{long,lat})}$.

Consider an air pollution factor and denote its concentration at hour h of day d in week w of year t and located at longitude-latitude coordinates (long,lat) as $x_{t,w,d,h}^{(\text{long,lat})}$.

Compute the daily minimum, average, and maximum concentrations of the air pollutant, measured at the coordinates (long,lat) as:

$$\hat{x}_{t,w,d}^{(\text{long,lat})} = \min\left\{x_{t,w,d,h}^{(\text{long,lat})} \mid h = 0, 1, ..., 23\right\}$$
$$\overline{x}_{t,w,d}^{(\text{long,lat})} = \text{avg}\left\{x_{t,w,d,h}^{(\text{long,lat})} \mid h = 0, 1, ..., 23\right\}$$
$$\overset{\vee}{x}_{t,w,d}^{(\text{long,lat})} = \max\left\{x_{t,w,d,h}^{(\text{long,lat})} \mid h = 0, 1, ..., 23\right\}.$$

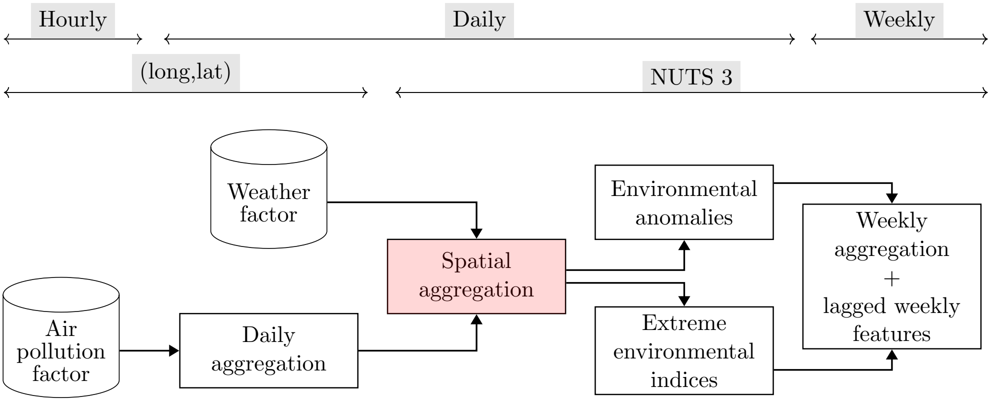
Consider an air pollution factor and denote its concentration at hour h of day d in week w of year t and located at longitude-latitude coordinates (long,lat) as $x_{t,w,d,h}^{(\text{long,lat})}$.

Compute the daily minimum, average, and maximum concentrations of the air pollutant, measured at the coordinates (long,lat) as:

$$\begin{aligned}
\hat{x}{}_{t,w,d}^{(\text{long,lat})} &= \min\left\{x_{t,w,d,h}^{(\text{long,lat})} \mid h = 0, 1, \ldots, 23\right\} \\
\overline{x}_{t,w,d}^{(\text{long,lat})} &= \operatorname{avg}\left\{x_{t,w,d,h}^{(\text{long,lat})} \mid h = 0, 1, \ldots, 23\right\} \\
\overset{\vee}{x}{}_{t,w,d}^{(\text{long,lat})} &= \max\left\{x_{t,w,d,h}^{(\text{long,lat})} \mid h = 0, 1, \ldots, 23\right\}.
\end{aligned}$$

Weather factors already available at the daily level (no need for daily aggregation).

Flow chart


Hourly
Daily
Weekly
(long,lat)
NUTS 3
Weather factor
Air pollution factor
Daily aggregation
Spatial aggregation
Environmental anomalies
Extreme environmental indices
Weekly aggregation + lagged weekly features

Spatial aggregation

$\tilde{x}_{t,w,d}^{(\text{long,lat})}$: daily level of a specific environmental feature at coordinates (long, lat) for year t, week w, and day d.

$\tilde{x}_{t,w,d}^{(\text{long,lat})}$: daily level of a specific environmental feature at coordinates (long, lat) for year t, week w, and day d.

Construct feature on NUTS 3 scale:

$$\tilde{x}_{t,w,d}^{(r)} = \sum_{(\text{long,lat}) \in \mathcal{I}_1(r)} \omega_{(\text{long,lat})} \cdot \tilde{x}_{t,w,d}^{(\text{long,lat})},$$

$\tilde{x}^{(\text{long,lat})}_{t,w,d}$: daily level of a specific environmental feature at coordinates (long, lat) for year t, week w, and day d.

Construct feature on NUTS 3 scale:

$$\tilde{x}^{(r)}_{t,w,d} = \sum_{(\text{long,lat}) \in \mathcal{I}_1(r)} \omega_{(\text{long,lat})} \cdot \tilde{x}^{(\text{long,lat})}_{t,w,d},$$

where:

$\omega_{(\text{long,lat})}$: population weights using gridded population data from the Socioeconomic Data and Applications Center (NASA's EOSDIS)

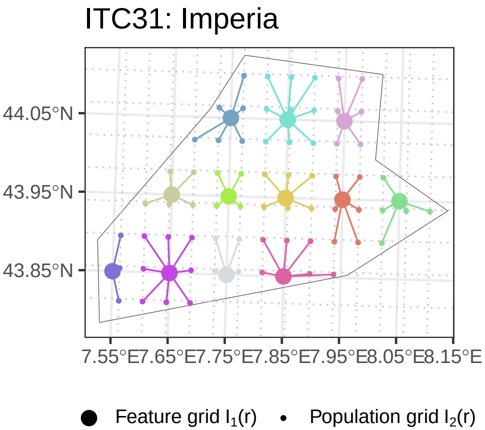
$\tilde{x}_{t,w,d}^{(\text{long,lat})}$: daily level of a specific environmental feature at coordinates (long, lat) for year t, week w, and day d.

Construct feature on NUTS 3 scale:

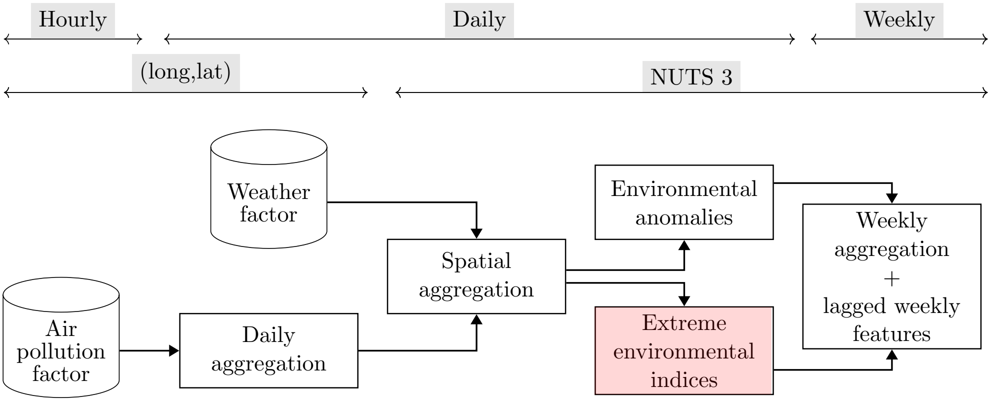
$$\tilde{x}_{t,w,d}^{(r)} = \sum_{(\text{long,lat}) \in \mathcal{I}_1(r)} \omega_{(\text{long,lat})} \cdot \tilde{x}_{t,w,d}^{(\text{long,lat})},$$

where:

- $\omega_{(\text{long,lat})}$: population weights using gridded population data from the Socioeconomic Data and Applications Center (NASA's EOSDIS)



Flow chart


Hourly
Daily
Weekly
(long,lat)
NUTS 3
Weather factor
Air pollution factor
Daily aggregation
Spatial aggregation
Environmental anomalies
Extreme environmental indices
Weekly aggregation
+
lagged weekly features

Extreme environmental indices

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

Extreme environmental indices

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

Calculate region-specific 5% and 95% quantiles of the daily historical temperature or air pollution observations over the years 2013-2019.

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

Calculate region-specific 5% and 95% quantiles of the daily historical temperature or air pollution observations over the years 2013-2019.

Define extreme high temperature index (hot-day index):

$$\text{T.ind}_{t,w,d}^{(r,95\%)} = \mathbb{1}\left\{\texttt{Tmax}_{t,w,d}^{(r)} \geq q_{\texttt{Tmax}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tavg}_{t,w,d}^{(r)} \geq q_{\texttt{Tavg}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tmin}_{t,w,d}^{(r)} \geq q_{\texttt{Tmin}}^{(r,95\%)}\right\}.$$

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

Calculate region-specific 5% and 95% quantiles of the daily historical temperature or air pollution observations over the years 2013-2019.

Define extreme high temperature index (hot-day index):

$$\text{T.ind}_{t,w,d}^{(r,95\%)} = \mathbb{1}\left\{\texttt{Tmax}_{t,w,d}^{(r)} \geq q_{\texttt{Tmax}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tavg}_{t,w,d}^{(r)} \geq q_{\texttt{Tavg}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tmin}_{t,w,d}^{(r)} \geq q_{\texttt{Tmin}}^{(r,95\%)}\right\}.$$

Index values: 0-3, indicating the severity of hot days.

Extreme environmental indices

Aim: to capture the effects of extreme environmental conditions on mortality baseline deviations.

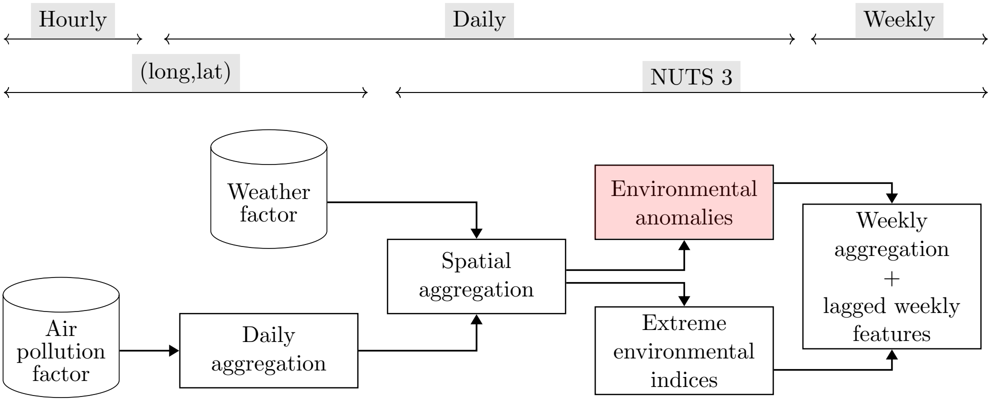
Calculate region-specific 5% and 95% quantiles of the daily historical temperature or air pollution observations over the years 2013-2019.

Define extreme high temperature index (hot-day index):

$$\text{T.ind}_{t,w,d}^{(r,95\%)} = \mathbb{1}\left\{\texttt{Tmax}_{t,w,d}^{(r)} \geq q_{\texttt{Tmax}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tavg}_{t,w,d}^{(r)} \geq q_{\texttt{Tavg}}^{(r,95\%)}\right\} + \mathbb{1}\left\{\texttt{Tmin}_{t,w,d}^{(r)} \geq q_{\texttt{Tmin}}^{(r,95\%)}\right\}.$$

Index values: 0-3, indicating the severity of hot days.

Similar extreme indices are created for the other daily weather and air pollution factors.



Hourly
Daily
Weekly
(long,lat)
NUTS 3
Weather factor
Air pollution factor
Daily aggregation
Spatial aggregation
Environmental anomalies
Extreme environmental indices
Weekly aggregation + lagged weekly features

Environmental anomalies

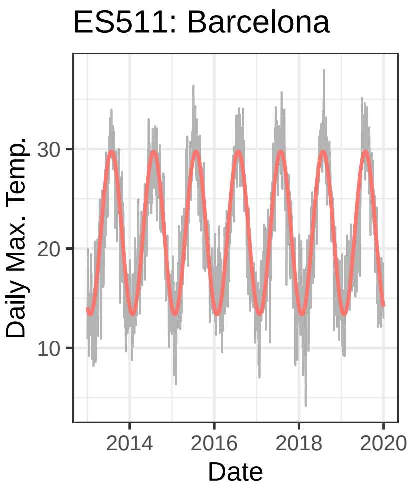
Environmental anomalies

Create features that quantify deviations from typical, baseline conditions for each day throughout the year.

Create features that quantify deviations from typical, baseline conditions for each day throughout the year.

Robust linear regression to capture baseline:

$$\tilde{x}^{(r)}_{t,w,d} = \alpha^{(r)}_0 + \alpha^{(r)}_1 \sin\left(\frac{2\pi d}{365.25}\right) + \alpha^{(r)}_2 \cos\left(\frac{2\pi d}{365.25}\right) + \epsilon^{(r)}_{t,w,d},$$



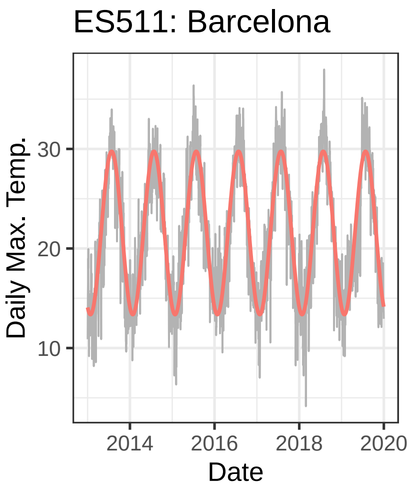
Create features that quantify deviations from typical, baseline conditions for each day throughout the year.

Robust linear regression to capture baseline:

$$\tilde{x}^{(r)}_{t,w,d} = \alpha^{(r)}_0 + \alpha^{(r)}_1 \sin\left(\frac{2\pi d}{365.25}\right) + \alpha^{(r)}_2 \cos\left(\frac{2\pi d}{365.25}\right) + \epsilon^{(r)}_{t,w,d},$$

In the paper, we work with excesses or deviations from the baseline (anomalies):

$$\tilde{x}^{(r)}_{t,w,d} - \hat{\tilde{x}}^{(r)}_{t,w,d}$$



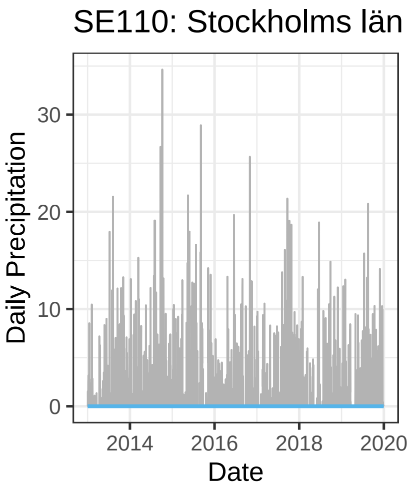
Create features that quantify deviations from typical, baseline conditions for each day throughout the year.

Robust linear regression to capture baseline:

$$\tilde{x}^{(r)}_{t,w,d} = \alpha^{(r)}_0 + \alpha^{(r)}_1 \sin\left(\frac{2\pi w}{365.25}\right) + \alpha^{(r)}_2 \cos\left(\frac{2\pi w}{365.25}\right) + \epsilon^{(r)}_{t,w,d},$$

In the paper we work with excesses or deviations from the baseline (anomalies):

$$\tilde{x}^{(r)}_{t,w,d} - \hat{\tilde{x}}^{(r)}_{t,w,d}$$



How does each feature impact the predictions produced by the machine learning model?

Partial dependence plots interpret the marginal effect of a feature X_l on the model predictions as:

$$f_{l,\mathrm{PDP}}(x) = \frac{1}{n}\sum_{j=1}^{n} \hat{f}(x, \boldsymbol{x}_j^{(-l)}),$$

where $\boldsymbol{x}_j^{(-l)}$ is the j-th training observation where the l-th feature is omitted.

Problem: correlated features may lead to unrealistic data points.

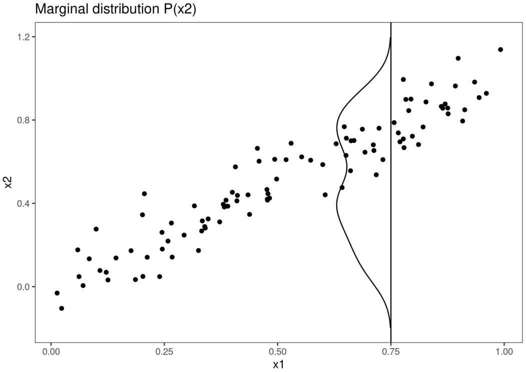


Figure: Visualisation taken from Molnar (2019).

Accumulated local effects - motivation

How does each feature impact the predictions generated by a machine learning model?

Partial dependence plots interpret the marginal effect of a feature X_l on the model predictions as:

$$f_{l,\text{PDP}}(x) = \frac{1}{n}\sum_{j=1}^{n} \hat{f}(x, \boldsymbol{x}_j^{(-l)}),$$

where $\boldsymbol{x}_j^{(-l)}$ is the j-th training observation where the l-th feature is omitted.

Problem: correlated features may lead to unrealistic data points.

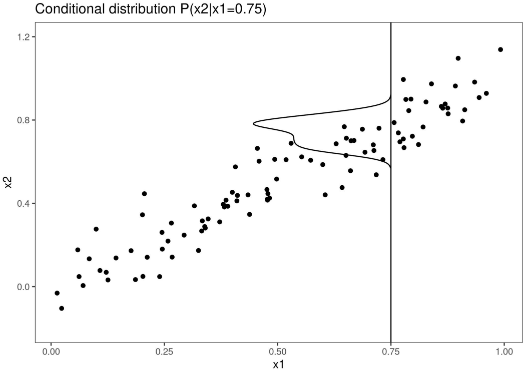


Figure: Visualisation taken from Molnar (2019).

How does each feature impact the predictions generated by a machine learning model?

Partial dependence plots interpret the marginal effect of a feature X_l on the model predictions as:

$$f_{l,\mathrm{PDP}}(x) = \frac{1}{n}\sum_{j=1}^{n} \hat{f}(x, \boldsymbol{x}_j^{(-l)}),$$

where $\boldsymbol{x}_j^{(-l)}$ is the j-th training observation where the l-th feature is omitted.

Problem: correlated features may lead to unrealistic data points.

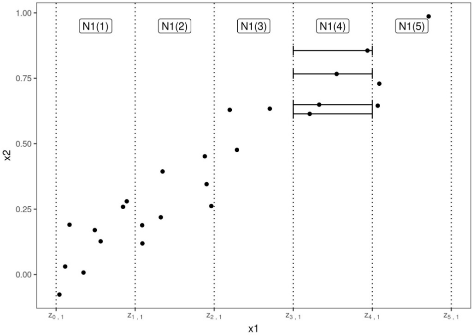


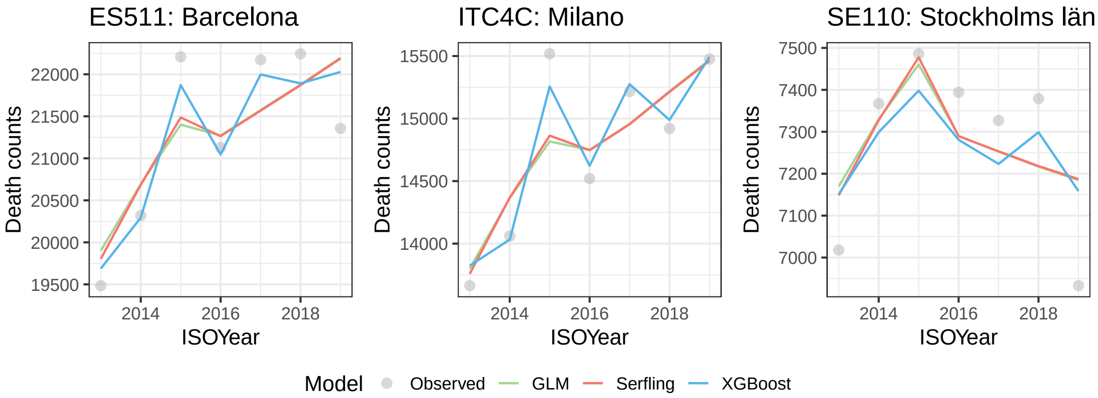
Figure: Visualisation taken from Molnar (2019).

Accumulated local effects

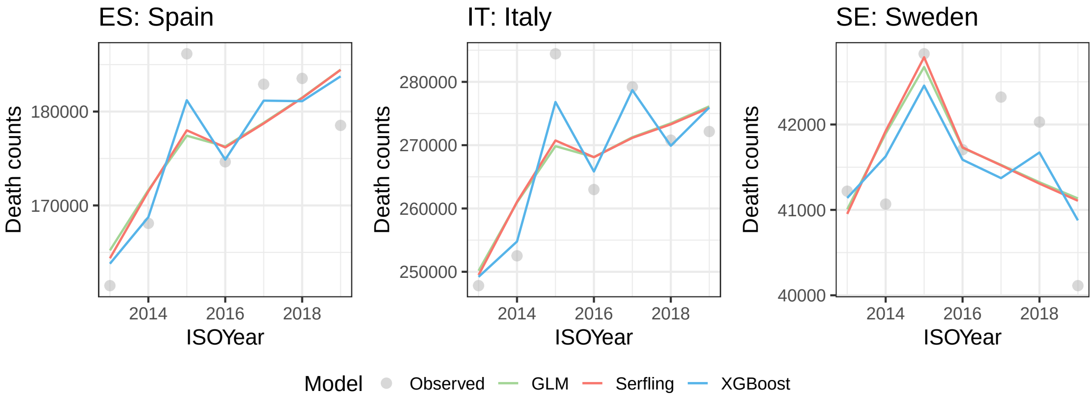
Accumulated Local Effects avoid the use of unrealistic data instances, average changes in predictions and accumulate them. For a feature X_l, the ALE effect equals:

$$f_{l,\text{ALE}}(x) = \int_{z_{0,l}}^{x} \mathbb{E}\left[\frac{\partial f(X_1, X_2, ..., X_p)}{\partial X_l}\middle| X_l = z_l\right] dz_l - c_l,$$

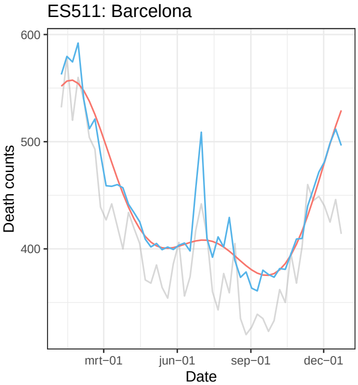
Instead of averaging predictions, ALE average changes of predictions (via the partial derivatives). The integral over z_l accumulates the differences in the predictions when moving over the range of X_l. A constant c_l is subtracted to center the ALE plot and make sure the mean effect is zero.



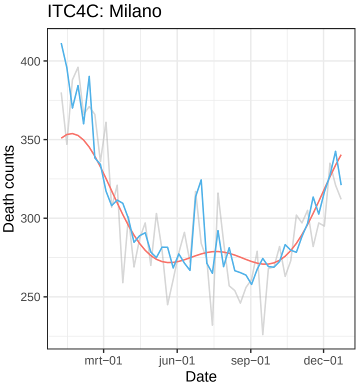
ES511: Barcelona
ITC4C: Milano
SE110: Stockholms län
Death counts
ISOYear
22000
21500
21000
20500
20000
19500
15500
15000
14500
14000
7500
7400
7300
7200
7100
7000
2014
2016
2018
Model
Observed
GLM
Serfling
XGBoost



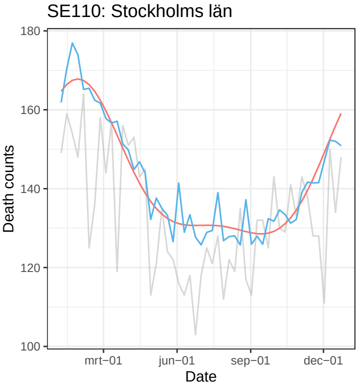
ES: Spain
IT: Italy
SE: Sweden
Death counts
180000
170000
280000
270000
260000
250000
42000
41000
40000
2014
2016
2018
ISOYear
Model
Observed
GLM
Serfling
XGBoost



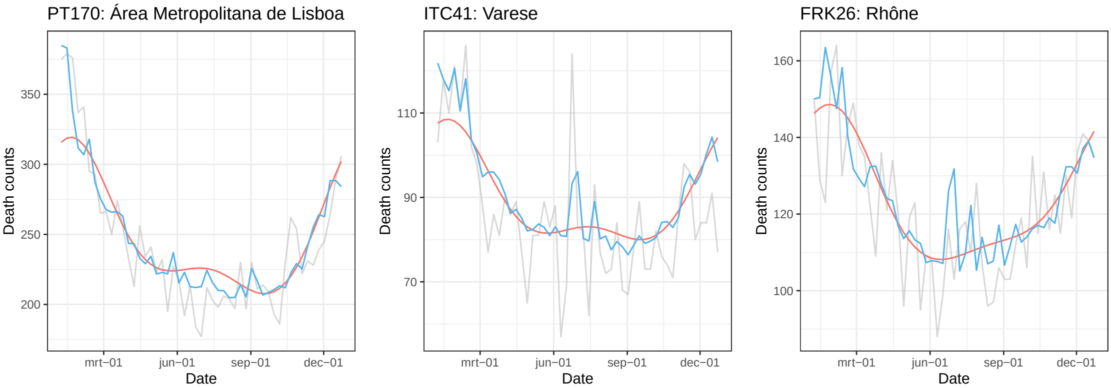
ES511: Barcelona
600
500
400
Death counts
mrt−01
jun−01
sep−01
dec−01
Date



ITC4C: Milano
400
350
300
250
Death counts
mrt−01
jun−01
sep−01
dec−01
Date



SE110: Stockholms län
180
160
140
120
100
Death counts
mrt−01
jun−01
sep−01
dec−01
Date



PT170: Área Metropolitana de Lisboa
ITC41: Varese
FRK26: Rhône
Death counts
Date
mrt−01
jun−01
sep−01
dec−01
350
300
250
200
110
90
70
160
140
120
100