

# Impact of Environmental Factors

## on Acute Myocardial Infarction

Gabriela Zemenčíková

June, 2024



UNIVERSITAT DE  
BARCELONA

# Dataset & Data Preparation

---

- The environmental dataset and census data taken from the **Statistical Institute of Catalonia**.

# Dataset & Data Preparation

---

- The environmental dataset and census data taken from the **Statistical Institute of Catalonia**.
- AMI dataset taken from ten main hospitals across Catalonia.

# Dataset & Data Preparation

---

- The environmental dataset and census data taken from the **Statistical Institute of Catalonia**.
- AMI dataset taken from ten main hospitals across Catalonia.
- 22,812 hospital admissions at the scale of the 948 municipalities in Catalonia stratified by province, sex and age

# Dataset & Data Preparation

---

- The environmental dataset and census data taken from the **Statistical Institute of Catalonia**.
- AMI dataset taken from ten main hospitals across Catalonia.
- 22,812 hospital admissions at the scale of the 948 municipalities in Catalonia stratified by province, sex and age
- **Response variable:** Age Standardised Incidence Rate (ASIR)

# Dataset & Data Preparation

---

- The environmental dataset and census data taken from the **Statistical Institute of Catalonia**.
- AMI dataset taken from ten main hospitals across Catalonia.
- 22,812 hospital admissions at the scale of the 948 municipalities in Catalonia stratified by province, sex and age
- **Response variable:** Age Standardised Incidence Rate (ASIR)
- A total of **5** predictors - Humidity, Temperature, Ozone levels, Particulate Matter, Public holidays

# Choice of Age-Standardized Incidence Rates (ASIR)

---

- Ensures comparability of events across populations.

# Choice of Age-Standardized Incidence Rates (ASIR)

---

- Ensures comparability of events across populations.
- Ensures comparability across different age distributions.



# Choice of Age-Standardized Incidence Rates (ASIR)

---

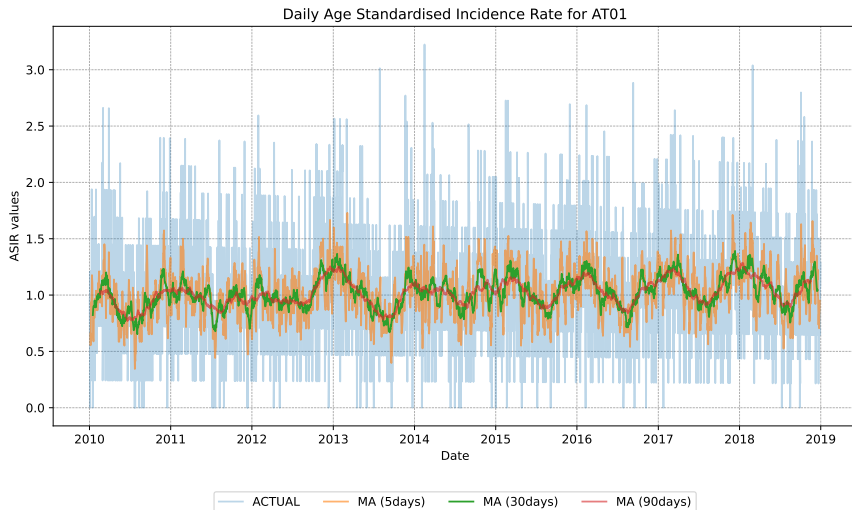
- Ensures comparability of events across populations.
- Ensures comparability across different age distributions.
- Adjusts for age as a confounding factor, providing a more accurate representation of AMI incidence.

# Choice of Age-Standardized Incidence Rates (ASIR)

---

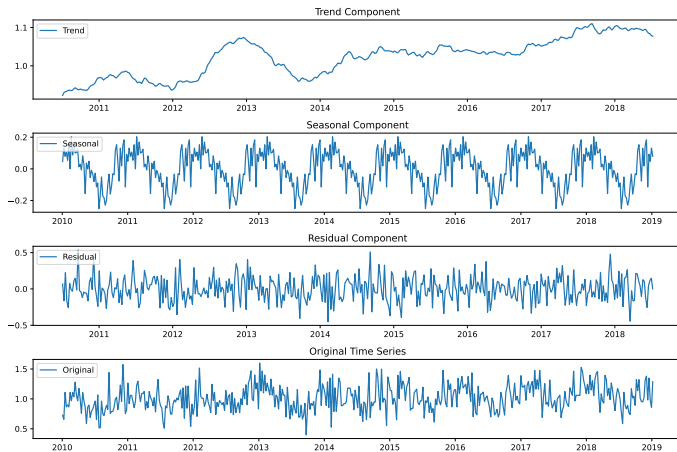
- Ensures comparability of events across populations.
- Ensures comparability across different age distributions.
- Adjusts for age as a confounding factor, providing a more accurate representation of AMI incidence.
- Allows comparisons between regions or over time periods.

# Target Variable - ASIR



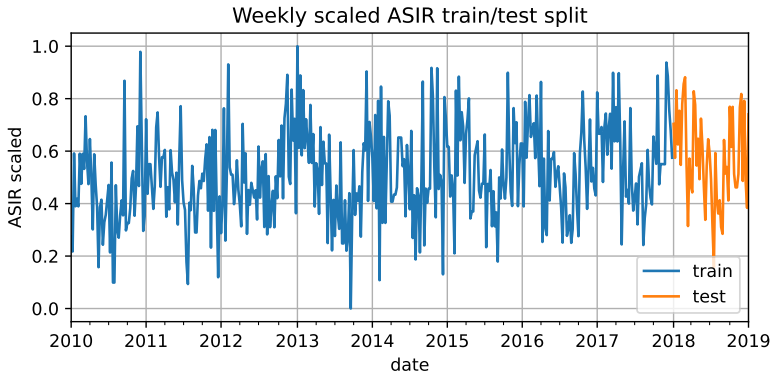
# Target Variable - ASIR

---



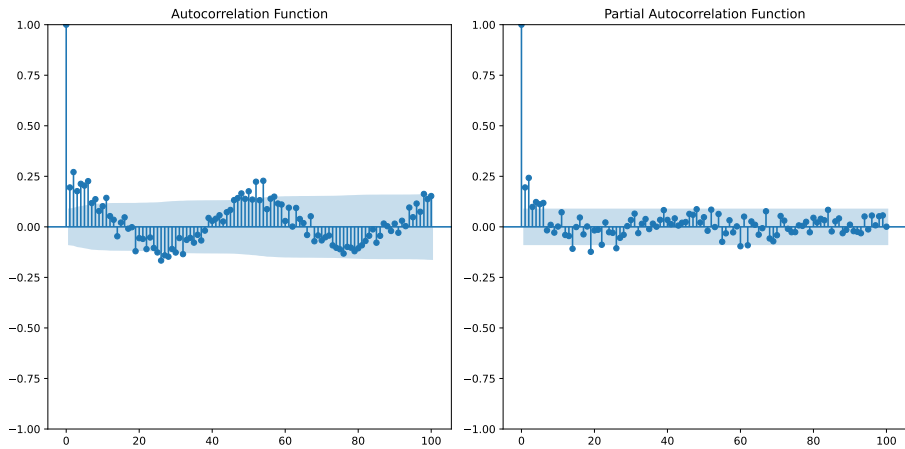
# Target Variable - ASIR

---



# Target Variable - ASIR

---



# SARIMAX

---

SARIMAX model can be represented as:

$$Y_t = \beta X_t + \epsilon_t$$

- $Y_t$  is the observed ASIR at time  $t$ ,
- $X_t$  represents the vector of exogenous variables at time  $t$ ,
- $\beta$  is the vector of coefficients for the exogenous variables,
- $\epsilon_t$  is the error term following a stationary ARIMA process.

# SARIMAX

---

SARIMAX model can be represented as:

$$Y_t = \beta X_t + \epsilon_t$$

- $Y_t$  is the observed ASIR at time  $t$ ,
- $X_t$  represents the vector of exogenous variables at time  $t$ ,
- $\beta$  is the vector of coefficients for the exogenous variables,
- $\epsilon_t$  is the error term following a stationary ARIMA process.

By incorporating exogenous variables, it allows for a more comprehensive modeling of the relationship between environmental factors and ASIR.



# SARIMAX

---

SARIMAX model can be represented as:

$$Y_t = \beta X_t + \epsilon_t$$

- $Y_t$  is the observed ASIR at time  $t$ ,
- $X_t$  represents the vector of exogenous variables at time  $t$ ,
- $\beta$  is the vector of coefficients for the exogenous variables,
- $\epsilon_t$  is the error term following a stationary ARIMA process.

By incorporating exogenous variables, it allows for a more comprehensive modeling of the relationship between environmental factors and ASIR.

By integrating both temporal and environmental variables, it enables more robust forecasting, taking into consideration seasonality

# LSTM

---

- a type of recurrent neural network (RNN) architecture specifically designed to model sequence data while addressing the vanishing gradient problem.

# LSTM

---

- a type of recurrent neural network (RNN) architecture specifically designed to model sequence data while addressing the vanishing gradient problem.
- The LSTM cell has a memory cell and three gates: input gate, forget gate, and output gate

# LSTM

---

- a type of recurrent neural network (RNN) architecture specifically designed to model sequence data while addressing the vanishing gradient problem.
- The LSTM cell has a memory cell and three gates: input gate, forget gate, and output gate
- LSTM networks are capable of learning long-term dependencies in sequential data, making them suitable for time series forecasting tasks.

# LSTM

---

- a type of recurrent neural network (RNN) architecture specifically designed to model sequence data while addressing the vanishing gradient problem.
- The LSTM cell has a memory cell and three gates: input gate, forget gate, and output gate
- LSTM networks are capable of learning long-term dependencies in sequential data, making them suitable for time series forecasting tasks.
- can capture complex patterns and relationships to forecast future ASIR.

# Hyperparameter optimization

---

## SARIMAX Model Optimization:

- Utilized grid search over a range of parameters ( $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$ ,  $Q$ ,  $s$ ).
- Evaluated using time series cross-validation (TimeSeriesSplit).
- Identified best parameters:
  - $p = 1, d = 0, q = 1$
  - $P = 1, D = 1, Q = 1, s = 52$

# Hyperparameter optimization

---

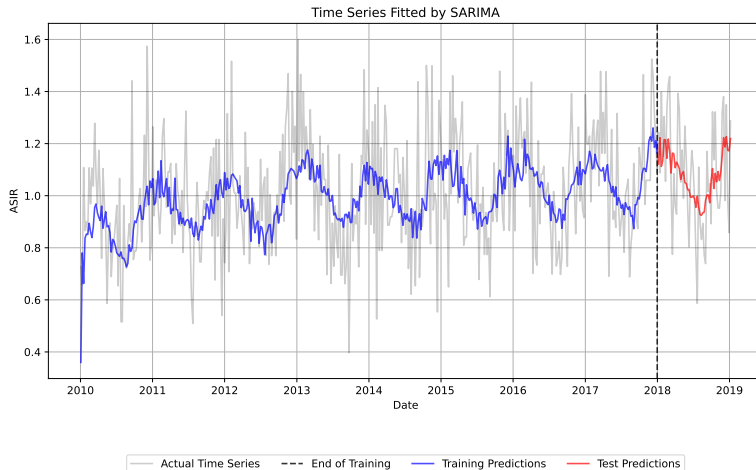
## SARIMAX Model Optimization:

- Utilized grid search over a range of parameters ( $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$ ,  $Q$ ,  $s$ ).
- Evaluated using time series cross-validation (TimeSeriesSplit).
- Identified best parameters:
  - $p = 1$ ,  $d = 0$ ,  $q = 1$
  - $P = 1$ ,  $D = 1$ ,  $Q = 1$ ,  $s = 52$

## LSTM Model Optimization:

- Conducted random search over the search space for hyperparameters.
- Executed multiple trials with different configurations.
- Best hyperparameters:
  - Units: **150**
  - Activation: **ReLU**
  - Dropout: **0.2**
  - Number of layers: **1**
  - LSTM layer 0 units: **150**
  - Optimizer: **Adam**
  - LSTM layer 1 units: **150**

# Results - SARIMAX





# Results - SARIMAX

---

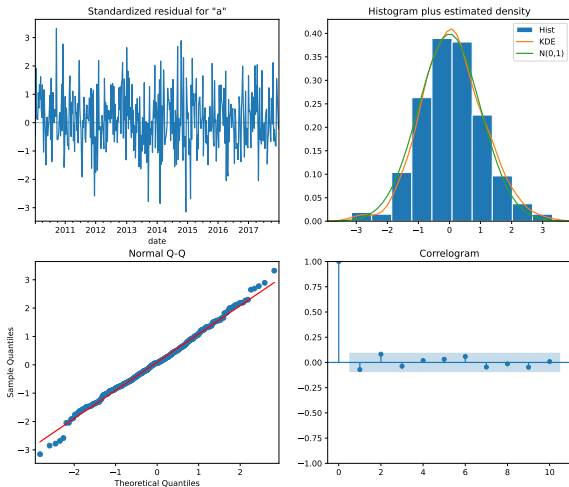
- Maximum temperature (`max_temp`): A decrease of 1°C in maximum temperature is associated with a decrease of approximately 0.345 units in the Age-Standardized Incidence Rate (ASIR) of AMI.

# Results - SARIMAX

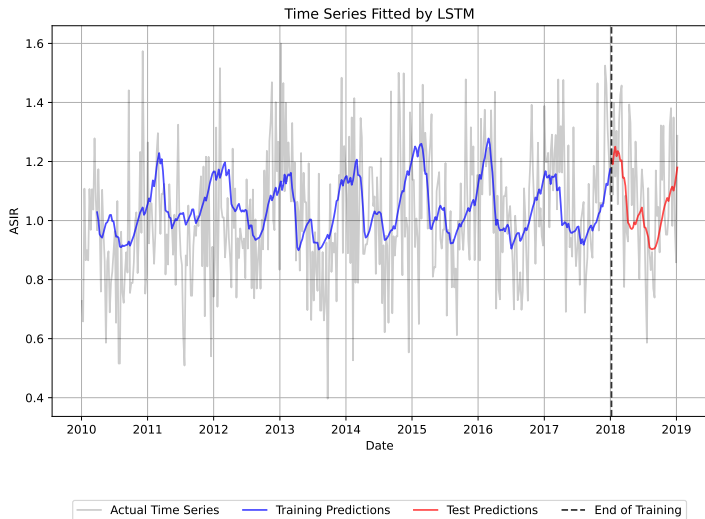
---

- Maximum temperature (`max_temp`): A decrease of  $1^{\circ}\text{C}$  in maximum temperature is associated with a decrease of approximately 0.345 units in the Age-Standardized Incidence Rate (ASIR) of AMI.
- Mean PM10 concentration (`mean_PM10`): An increase of  $1 \mu\text{g}/\text{m}^3$  in mean PM10 concentration is associated with an increase of approximately 0.146 units in ASIR of AMI.

# Diagnostics - SARIMAX



# Results - LSTM



# Comparison

---

Table 1: Results

	SARIMAX	LSTM
MAE	<b>0.12</b>	0.15
MSE	<b>0.02</b>	0.03
RMSE	<b>0.15</b>	0.19
MAPE	26.60	14.53

# Comparison

---

Table 1: Results

	SARIMAX	LSTM
MAE	<b>0.12</b>	0.15
MSE	<b>0.02</b>	0.03
RMSE	<b>0.15</b>	0.19
MAPE	26.60	14.53

- SARIMAX outperforms LSTM in terms of both MAE and MSE, making it a more suitable choice for this forecasting task.

# Comparison

---

Table 1: Results

	SARIMAX	LSTM
MAE	<b>0.12</b>	0.15
MSE	<b>0.02</b>	0.03
RMSE	<b>0.15</b>	0.19
MAPE	26.60	14.53

- SARIMAX outperforms LSTM in terms of both MAE and MSE, making it a more suitable choice for this forecasting task.
- However, the LSTM model exhibits substantially lower AIC and BIC values compared to SARIMAX, indicating a potentially better fit to the data and superior long-term forecasting capabilities

Thank you for your attention