Longevity18 Conference

London – 7th September

The effects of climate change on mortality projections

Maria Carannante[1]

Joint research with Valeria D’Amato[1] and Steven Haberman[2]

[1]University of Salerno, ITALY

[2]Faculty of Actuarial Science and Insurance, Bayes Business School, City University of London

Outline

Research relates to the link between climate risk and mortality risk

Climate risk	Use of satellite data to define a complex climate risk
	Definition of ML technique based on sliding windows to model dynamic data
	Definition of bioclimatic indicator
Mortality risk	Use of bioclimatic indicator as a covariate in LC family model
	Estimation of “climate-related” excess of mortality
	Model evaluation, in comparison of standard LC model

Climate risk

Research goal

Proposal of "smart" bioclimatic indicators, based on various artificial intelligence algorithms to handle satellite data, capable of adapting to the microclimates and morphological characteristics of the area

Data sources

NASA-MERRA 2 database

An open dataset that provides a consistent reprocessing of daily or intraday meteorological observations starting from 1980

Harmonized World Soil Database (HWSD)

A database organized by FAO in collaboration with IIASA, ISRIC-World Soil Information, Institute of Soil Science, Chinese Academy of Sciences (ISSCAS), and the Joint Research Centre of the European Commission (JRC), relating to satellite soil characteristics

Data pre-processing overview

To perform tree-based algorithms to define the climatic risk a huge data pre-processing is needed

1. Download of meteorological satellite data and create a data.frame containing information about days, latitudes and longitude and the variables of interest.

2. Download soil satellite data and create a data.frame about latitude and longitude of interest and soil characteristics.

3. Merge of meterological and soil databases in order to select only the data relating to the mainland.

The NASA MERRA-2 data

Step 1. Dataset download

NASA MERRA-2 datasets are available in **NetCDF format**. To build the bioclimatic index we use as meteorological variables the air temperature and the precipitation level from 1980/01/01 to 2020/08/30, totalling **14’854 datasets**.

The dedicated NASA server allows to creation of a link to download files one by one.

For this reason, it is necessary to **automatize the data downloading** from the server.

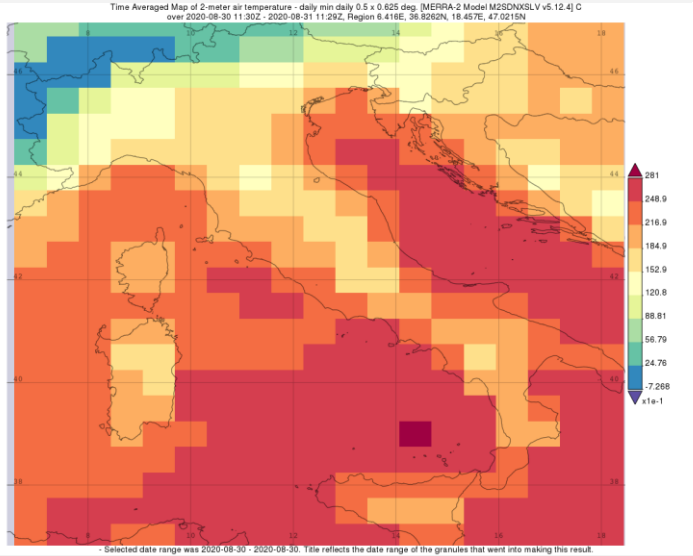
The NASA MERRA-2 data

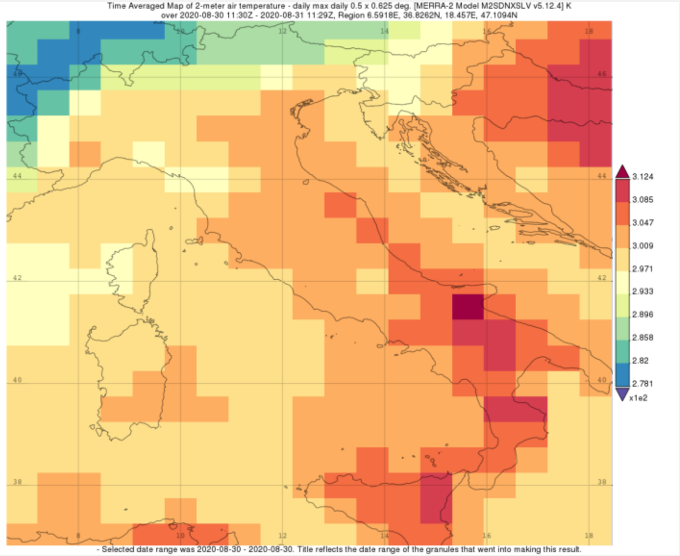
Step 2. Read NetCDF data

Transform the NetCDF files in a data.frame

1. **Selection of latitude and longitude of interest** for each variable. Each couple of latitude and longitude is the statistical unit or pixel.
2. **Store variables and attributes** of variable of interest.
3. Convert text containing dates to a **date object.**
4. **Set the dimension names and values** of the matrix to the appropriate latitude and longitude values.
5. **Merge rows** of each variable by date.
6. **Join variables** in a unique data.frame.

The NASA MERRA-2 data





2020-08-30 daily minimum and daily maximum air temperature. Source: NASA MERRA-2

The HWSD data

Step 1. Dataset download

HWSD is available in raster format. To load the data in the R environment, we use a specifical library, **rhwsd**, that create a SQL connection to database and extract values for specifical selected latitude and longitude.

Since we are interested in Italian data, we refer to the coordinates used to download the MERRA-2 data.

Step 2. Join MERRA-2 and HWSD data:

The starting dataset appears in the R environment as a nested list of geographical coordinates and variables. So, it is necessary to **unnest** the data in order to create the units by variables matrix.

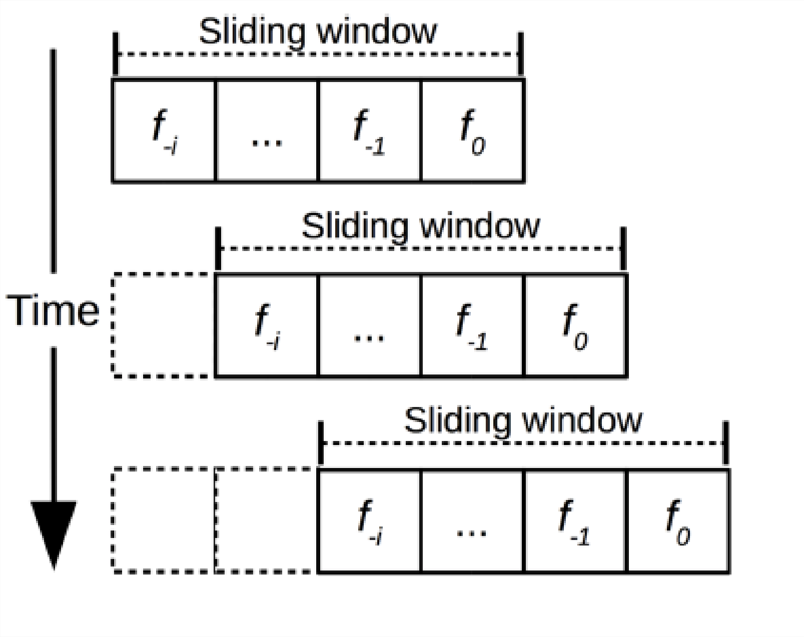
The next step is an **inner join of MERRA-2 and HWSD datasets** and the selection of non-empty observation in the soil ID variable. In this way, **only the mainland pixels** are selected.

Temporal dynamic Forest

To face a time series problem through tree-based algorithms, a **rolling window method** is used.

Features for each target variable correspond to the **previous time step** to which the target variable refers.

The data restructure **preserves the temporal order** between the observations and it continues to be preserved when using this dataset to train a supervised model. (Carannante *et al.* 2023a)



Bagging vs. Boosting

Random Forest (RF) (Breiman 2001) is an ensemble of tree-based techniques working by constructing a multitude of decision trees at training time.

Trees are aggregated using a bootstrap as follows

$$\hat{f}^{RF}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}^{TREE}(\mathbf{X}|b)$$

Extreme Gradient Boosting Algorithm (XGBoost) (Chen and Guestrin 2016) is an ensemble method that works by boosting trees, developed to increase speed and performance.

XGBoost minimizes a regularized objective function

$$\mathcal{L}(\varphi) = \sum_i l\,(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2$

Build climate risk index

Training set creation

Aggregation of data for each latitude and longitude at monthly granularity.

Create the target variable of the model as the number of days in the month in which the risk threshold is exceeded.

Create 12 features for the algorithm by lagging of the month.

Model estimation

For each model class, 12 models, one for each month of the year, are in parallel trained on the same training set, which predicts, for each latitude, longitude and month, the value of the target variable in the following 12 months.

Mortality risk

Research goal

Proposal of a Lee-Carter family model to explain the effects of climate risk on mortality rates and improve mortality projections

Risks in mortality projections

Mortality risk	Risk of heterogeneity in mortality rates
Longevity risk	Risk of people living for longer than expected
Catastrophe risk	Risk of unexpected extreme events affecting mortality rates

Climate Lee-Carter model

Lee-Carter model

Let $\mu_{x,t}$ the force of mortality. A Lee-Carter model (Lee and Carter 1992) is defined as follows:

$$y_{x,t} = \log(\mu_{x,t}) = a_x + b_x k_t + \varepsilon_{xt}$$

Climate Lee-Carter model

We define the force of mortality conditional to climate risk $\mu_{x,t}$ and the relative model:

$$y_{x,t} = \log(\mu_{x,t}) = a_x + g_x c_t + b_x k_t + \varepsilon_{xt}$$

g_x the age-specific factor of climate risk
c_t climate risk

Climate Lee-Carter model

Climate Lee-Carter model

We define the force of mortality conditional to frailty $\mu_{x,t}$ and the relative model:

$$y_{x,t} = \log(\mu_{x,t}) = a_x + g_x c_t + b_x k_t + \varepsilon_{xt}$$

g_x the age-specific factor of climate risk
c_t climate risk

Niu and Melenberg (2014) analyse the relationship between the trend in mortality and the trend in economic growth in different populations.

Boumezoued *et al*. (2022) analyse the effects of exposure to heat on mortality in the French population.

Carannante *et al*. (2023b) analyse how the onset of comorbidities in the population influences mortality trends in the English over-50 population.

Climate Lee-Carter model

Climate Lee-Carter model

We define the force of mortality conditional to frailty $\mu_{x,t}$ and the relative model:

$$y_{x,t} = \log(\mu_{x,t}) = a_x + g_x c_t + b_x k_t + \varepsilon_{xt}$$

g_x the age-specific factor of climate risk
c_t climate risk

We analyse the effects of climate risk indicators on mortality trends.

We compare the results of observed heat exceed data with the **climatic risk indexes** obtained using RF and XGBoost temporal dynamic algorithms.

Climate risk indexes are **rebuilt as time series** from the individual monthly models, in this way we do not lose the temporal dependency structure of data.

Climate Lee-Carter model

Climate Lee-Carter model

We define the force of mortality conditional to frailty $\mu_{x,t}$ and the relative model:

$$y_{x,t} = \log(\mu_{x,t}) = a_x + g_x c_t + b_x k_t + \varepsilon_{xt}$$

g_x the age-specific factor of climate risk
c_t climate risk

Death rates $D_{x,t}$ are assumed to be Poisson random variables

$$D_{x,t} \sim (E_{x,t}\mu_{x,t}), \mu_{x,t} = \exp(a_x + g_x c_t + b_x k_t)$$

With c_t orthogonal to k_t.

c_t is a count random variable measuring **the number of days subjected to climate risk** in a year.

Climate Lee-Carter model

Climate Lee-Carter model

We define the force of mortality conditional to frailty $\mu_{x,t}$ and the relative model:

$$y_{x,t} = \log(\mu_{x,t}) = a_x + g_x c_t + b_x k_t + \varepsilon_{xt}$$

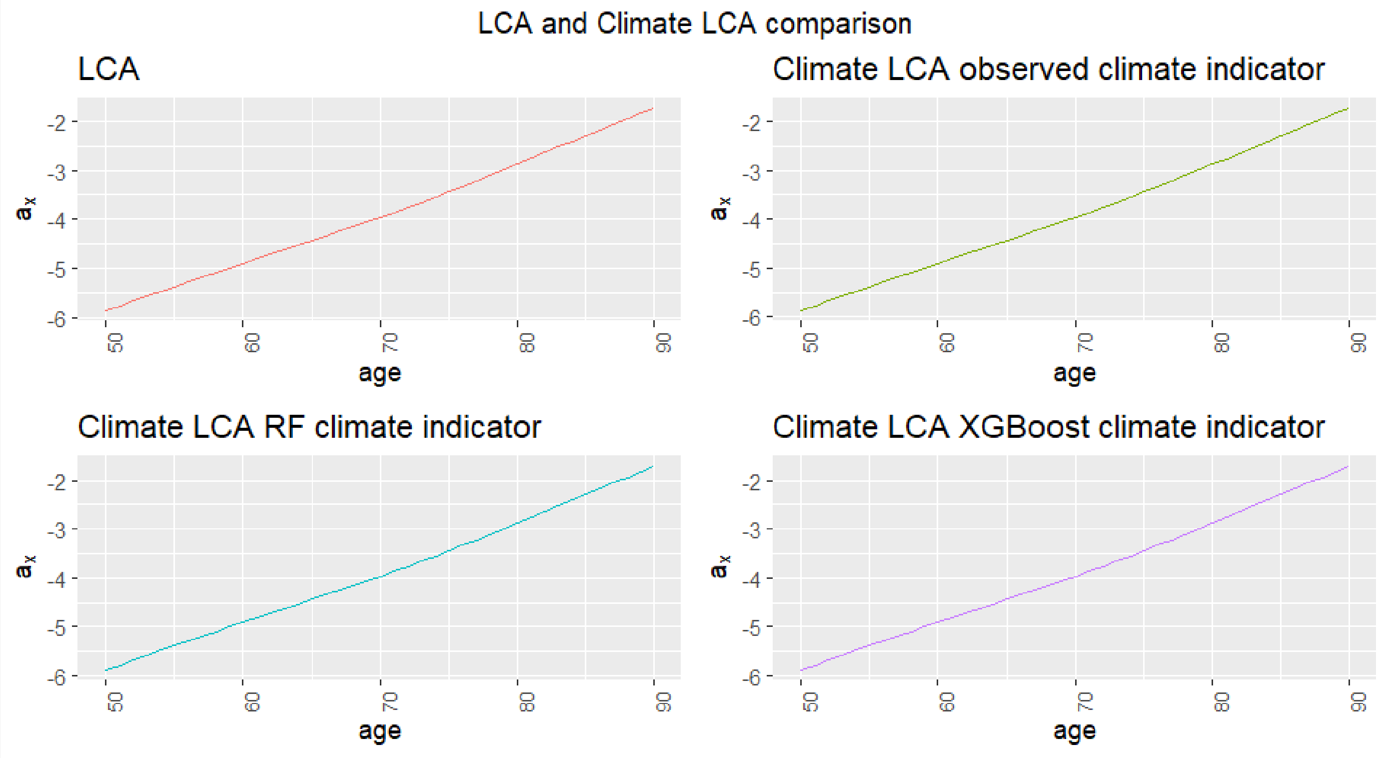
g_x the age-specific factor of climate risk
c_t climate risk

Preliminary results relate to the Italian population aged 50-90 from 1981 to 2020.

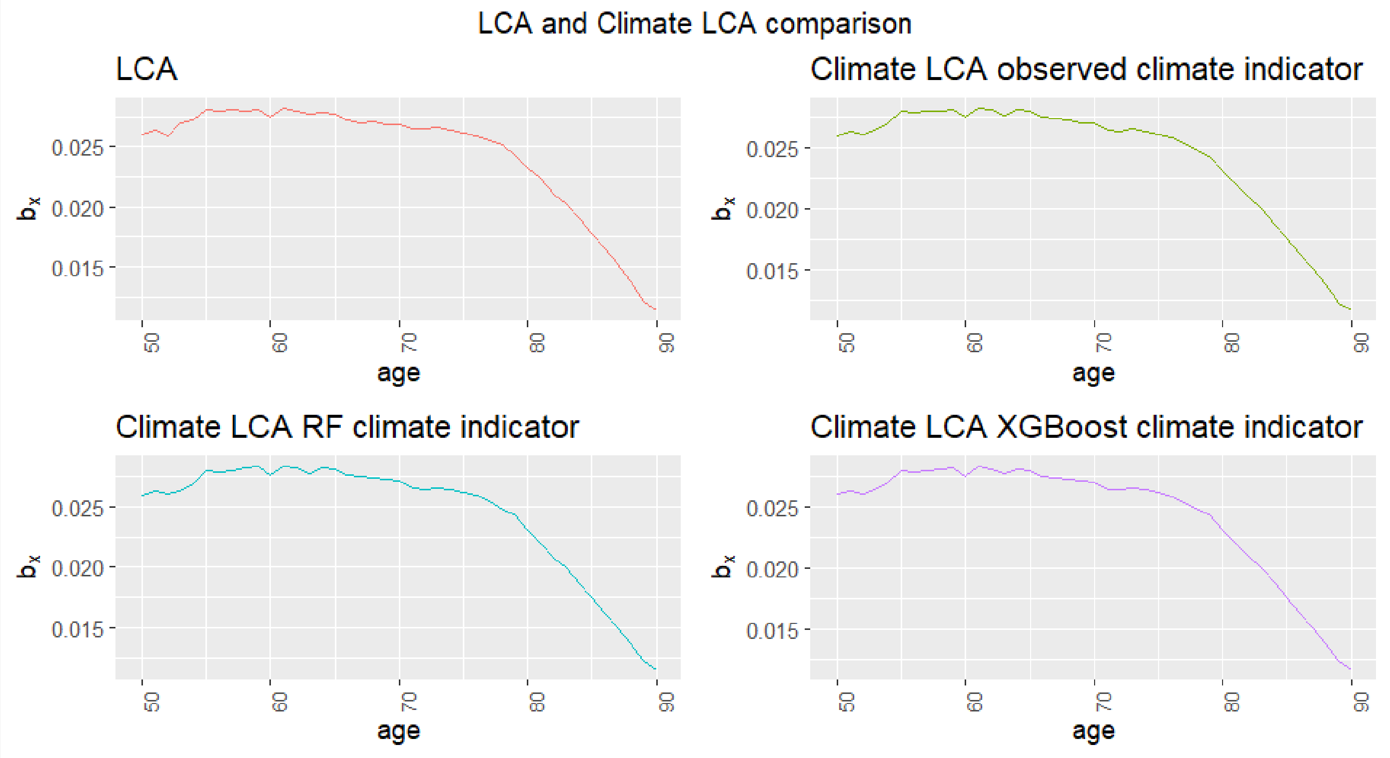
In this stage of analysis, we do not have information about the **spatial dependency** of data.

Our interest in research is to consider the **specific climate risk** of individual geographic areas. E.g., regions in Italy could show greater risks of heat waves than others.

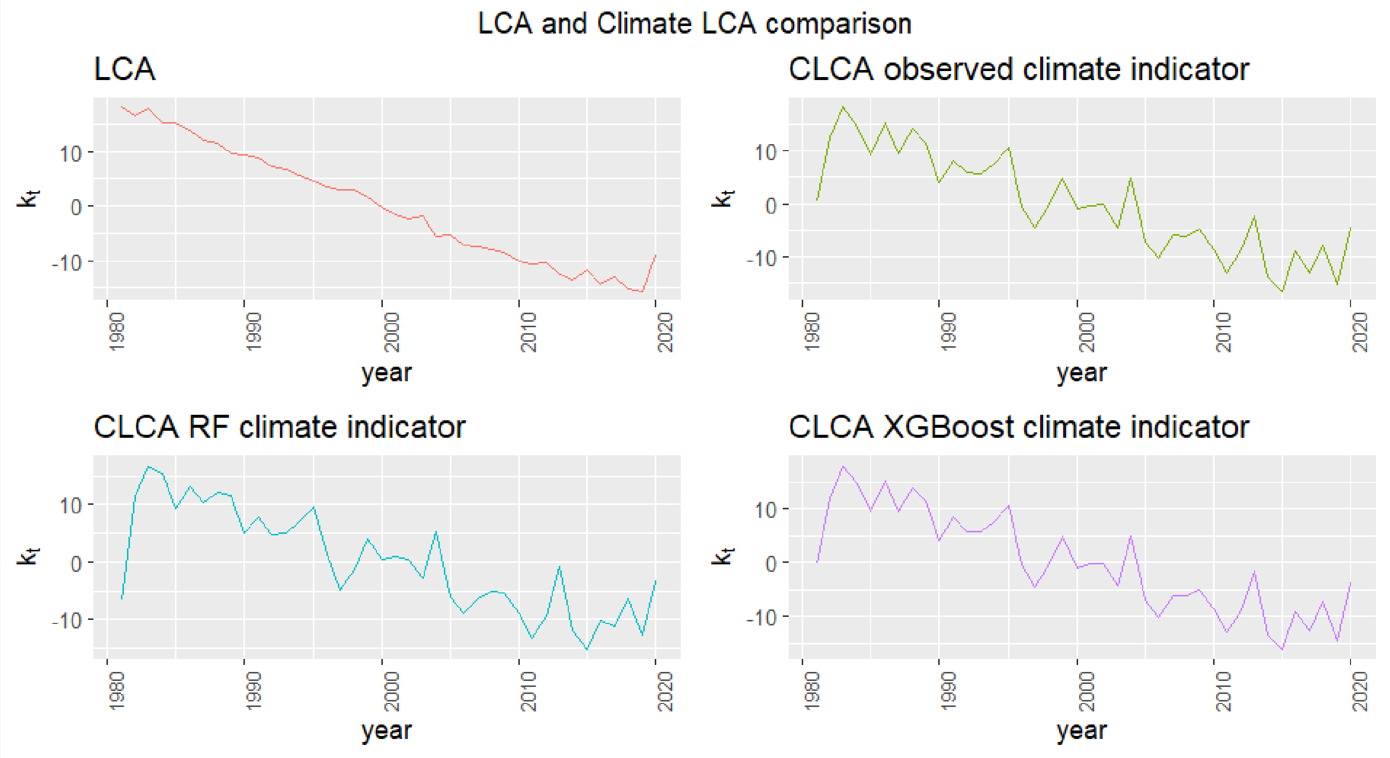
Preliminary results



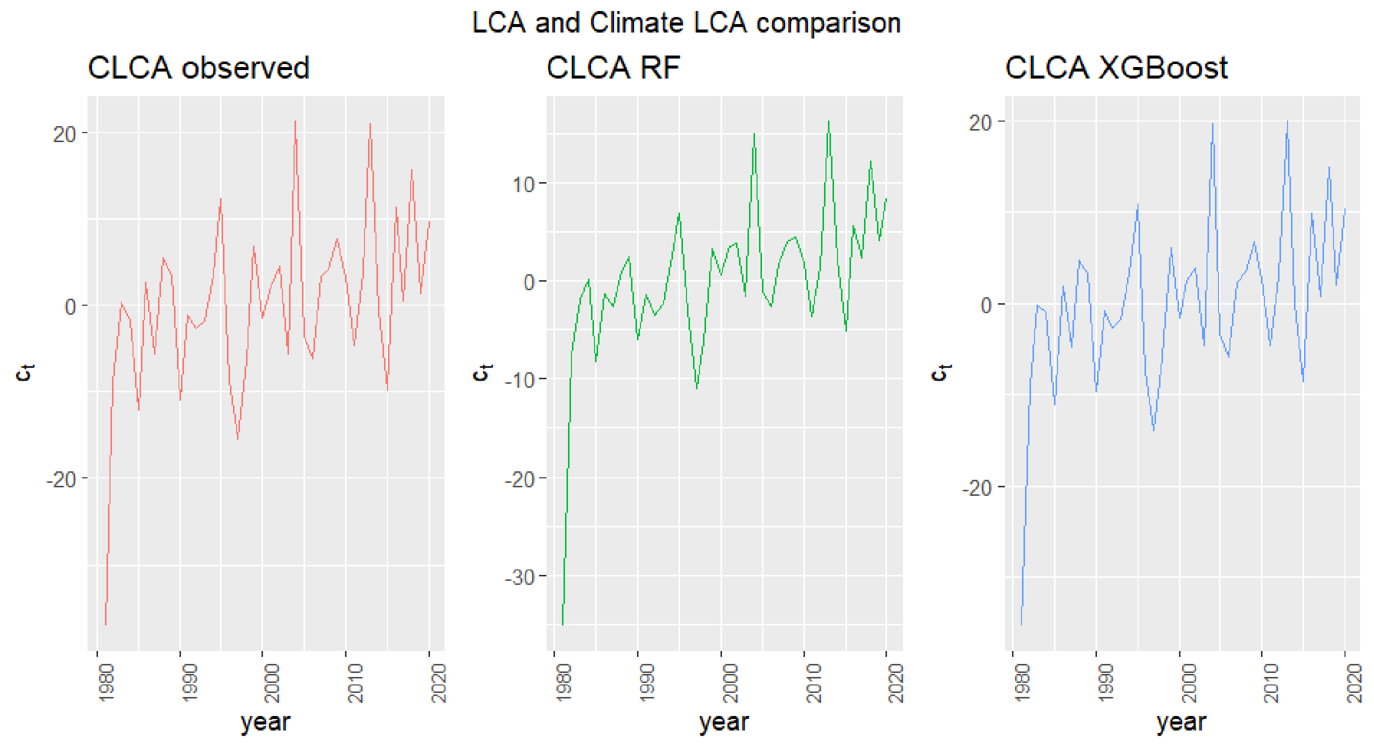
Preliminary results



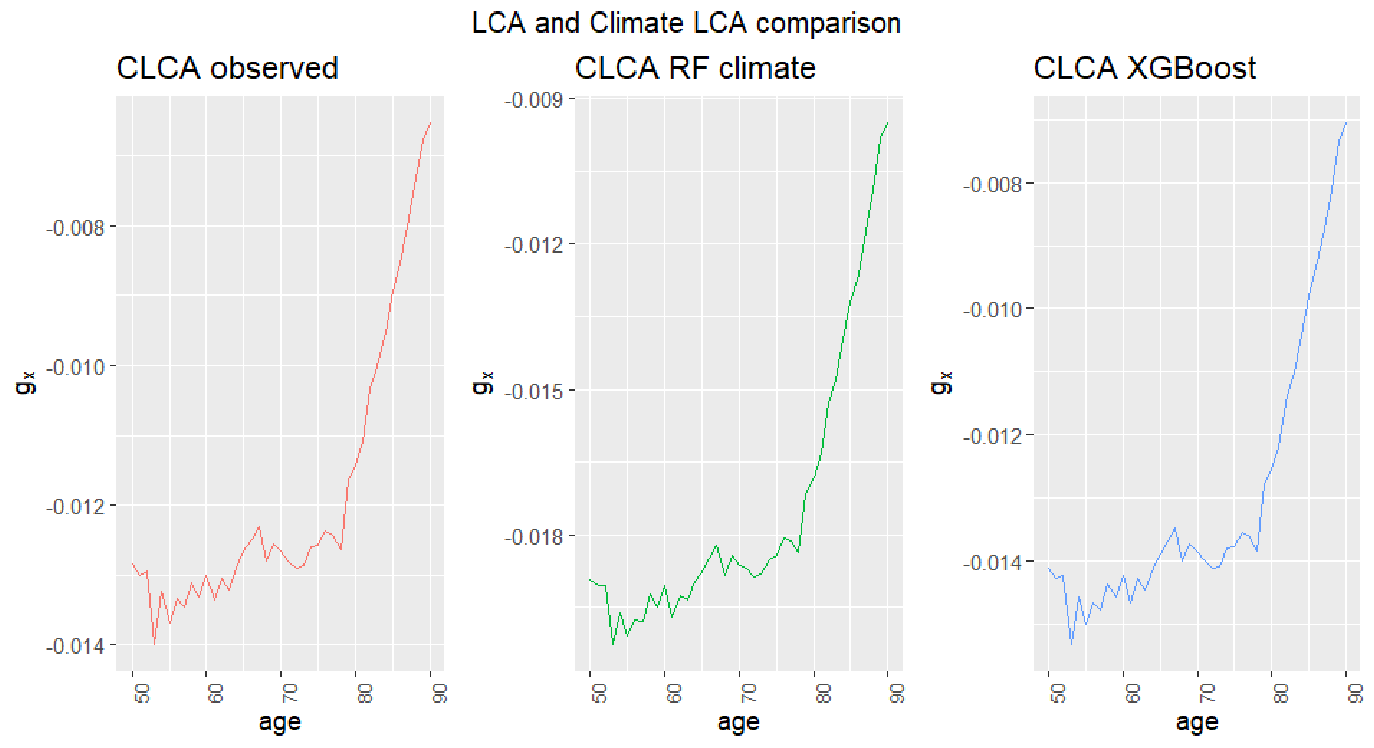
Preliminary results



Preliminary results



Preliminary results



Preliminary results

Model	Deviance	Test vs. LCA	p-value	BIC
LCA	27898.93	-	-	-14392.9
CLCA observed	28130.04	462.22	4.27E-101	-14660.9
CLCA RF	28009.67	221.50	7.99E-49	-14600.7
CLCA XGBoost	28117.27	436.68	1.50E-95	-14654.5

Conclusions

Climate risk

We propose a temporal dynamic forest to build climatic risk indexes that aim to preserve spatial and temporal dependence.

In a previous work (Carannante *et al.* 2023a), we show that the algorithm performs better than standard tree-based algorithms and LSTM networks.

Mortality risk

We propose an alternative functional form of the Lee-Carter family model, considering the morality risk depending on climate risk.

Preliminary results show the use of a climate risk indicator to explain the effects of exceeding heat on mortality trends but with negligible improvements using the climate risk indicator rather than the observed data.

Further research

Climate risk

Climate risk strongly depends on seasonality and geographical area.

The proposed algorithms could be useful to determine the **risk level** for different regions and temporal times.

Mortality risk

Deepen the analysis of the effects of climate risk on mortality risk using **regional** and/or **infra-annual** data.

Assess the **actuarial implications** of mortality risk assessment taking into account the effects of climate risk.

Main references

Boumezoued, A., Elfassihi, A., Germain, V., Titon, E., 2022. Modeling the impact of climate risk on mortality. Milliman White Paper

Bosilovich, M. G., R. Lucchesi, and M. Suarez. 2016. MERRA-2: File Specification. GMAO Office Note No. 9 (Version 1.1). Accessed 14 November 2020 http://gmao.gsfc.nasa.gov/pubs/office_notes

Breiman, L. 2001. Random forest. Machine Learning 45(1): 5–32 https://doi.org/10.1023/a:1010933404324

Carannante, M., D'Amato, V., Fersini, P., Forte, S. 2023a. Climate risk management by using index-based analytics: artificial algorithms on satellite data. Under review

Carannante, M., D'Amato, V., Haberman, S., Menzietti, M. 2023b. Frailty-based Lee Carter family of stochastic mortality models. Quality & Quantity. In print

Carter LR, Lee RD, 1992. Modeling and Forecasting U. S. Mortality. Journal of the American Statistical Association 87(419): 659-671

Chen, T., Guestrin, C. 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, https://doi.org/10.1145/2939672.2939785

Niu, G., Melenberg, B. 2014. Trends in mortality decrease and economic growth. Demography 51(5): 1755–1773 https://doi.org/10.1007/s13524-014-0328-3

Smith K, Woodward A, Campbell-Lendrum D, Chadee D, Honda Y, Liu Q, et al. 2014. Human health: impacts, adaptation, and co-benefits. In: Field CB, Barros V, Dokken D, editors. Climate change 2014: impacts, adaptation, and vulnerability. Vol I: global and sectoral aspects. Contribution of working group II to the fifth assessment report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press

WHO. 2014. Quantitative risk assessment of the effects of climate change on selected causes of death, 2030s and 2050s

Thank you

mcarannante@unisa.it

marcarannante

mar.carannante