Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/nhess-2021-389-RC2, 2022

Comment on nhess-2021-389

Francesco Sera (Referee)

Referee comment on "Machine learning models to predict myocardial infarctions from past climatic and environmental conditions" by Lennart Marien et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-389-RC2, 2022

This paper evaluate different machine algorithm to predict myocardial infarctions (MI) using environmental exposures variables.

The paperi is well structured and clear. The authors found a poor performance on predicting daily or weekly MI levels, but a good perfoemne on predicting the yearly values.

I think the analysis and the intrperetation of the results could be improved considering the following aspects:

- The authors considered a linear respresentation of the exposure (features). This is highly questionable especially for metereological variables. It is well known that the effect of ambient temperature on several health outcomes in non linear with higher helh impact for coldest and hottest temperature. This could be modelled using splines reparametrisation or in a simpler way using 3-piece linear spline (segmented linear) function (e.g. see Armstrong, Ben Models for the Relationship Between Ambient Temperature and Daily Mortality, Epidemiology: November 2006 - Volume 17 - Issue 6 - p 624-631). Mi suggestioni s to consider a reparametrisation of tem perature related variables considring both the heat and cold effects.
- The authors considered a lag of 3 days, and this would be enough for the heat effect of temeprature and for most pollutants, but it has been shown that the cold effect could have a longer delayed effects (up to 3 or 4 weeks). I suggest to increase the lag up to at leaast 21 days.
- I think that the good performance at yearly level is totally expcted as you are considering year and day of the year as features in your models, so environmental features doesn’t seems having any role here. To see if the environmental features have a role on predicting the yearly values models without time variables should be tested.
- Some demographic features were considered and I think they are reatevely stable over time, so I think they shouldn’t contain any information at daily or weekly level, Changes on the demograhic structure should be captured by the trend (year) variable.

If the objective was to standardise the outcome the authors could consider to use as outcome the daily incidence of MI, perhaps considering logarithm values in order to have a more symmetric distribution, without considering demoraphic predictors.

- I diagree with sentences in line 90-95. Actually it is possible considering case-controls desing nested within time-series using case-crossover design, that is each case is matched with days before and after the case day and the association measured conditioning on those risk set.