Atmos. Meas. Tech. Discuss., 7, C2801–C2803, 2014
www.atmos-meas-tech-discuss.net/7/C2801/2014/

Atmospheric Measurement Techniques Discussions

Interactive comment on "Regression models tolerant to massively missing data: a case study in solar radiation nowcasting" *by* I. Žliobaitė et al.

I. Žliobaitė et al.

zliobaite@gmail.com

Received and published: 22 September 2014

First of all, let us thank for your time committed to reviewing the manuscript, and thank you for your comments.

Indeed, the article has a stronger focus on the numerical aspects of the modeling, leaving the physical underpinning of the regression problem out of focus. In fact, such a focus is intentional. We have not systematically explored possible non-linear transformations of variables, or manual variable selection/combination in our case study. Since our primary goal is to investigate the techniques for dealing with missing values in the streaming data scenario, rather than producing a specific model for solar radiation prediction, we believe that such an investigation would introduce a distraction from the

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

main focus of the paper, and the principles of dealing with missing values would be lost among details of polishing the resulting model. Keeping this in mind, we would prefer to keep the focus on the numerical aspects, and resort to application of linear features.

We fully agree that nowcasting and forecasting are two different problems from the practical perspective, and our statement may be misleading. When revising the manuscript we will explicitly state, as you are suggesting, that forecasting is a different issue, and will rephrase the sentence regarding availability of related research addressing missing value issues.

Regarding uncertainty, it is possible that the measurements have different uncertainty and different level of noise due to different nature of the phenomenons measured as well as, potentially, due to different wear-and-tear of the measurement instrumentation. The true uncertainty is unknown, unless an alternative source of measurements for the same variable is available. But even if we knew the uncertainty associated with each variable, we have no immediate ideas how and whether we would be able to make use of that information for improving linear regression model accuracy, since uncertainly is already implicitly considered by the model: if a measurement is very noisy, the regression coefficient for that variable will be weak. Therefore, we believe we do not need to assume that all the measurements are equally uncertain. On the other hand, we do implicitly assume that the uncertainties are stable over time. We will add a clarification of this aspect in the revised version. In fact, variable uncertainties over time would make an interesting extension of the current work, if we had some way of knowing, how the uncertainties vary. The strength of uncertainty could be measured by from 0 to 1, where 0 means perfect certainty, 1 means missing value, and everything in between means a noisy measurement. In such a case, missing value could be considered as a special case of uncertainty. We will discuss this scenario in the future work section.

Indre Zliobaite on behalf of the authors

AMTD

7, C2801–C2803, 2014

Interactive comment on Atmos. Meas. Tech. Discuss., 7, 7137, 2014.

AMTD

7, C2801–C2803, 2014