Atmos. Meas. Tech. Discuss., 8, C3785–C3787, 2015
www.atmos-meas-tech-discuss.net/8/C3785/2015/

Atmospheric Measurement Techniques
Discussions

Interactive comment on "Self-Nowcast Model of extreme precipitation events for operational meteorology" *by* G. B. França et al.

Anonymous Referee #1

Received and published: 11 November 2015

This paper is interesting and quite well written. It addresses a topic of great importance in nowcasting research. I would like to reconsider the paper for publication if major revisions will be done, especially related to the major concerns that I will describe in what follows.

Major concerns

The division of the total dataset into a training set and a validation set only does not guarantee the reliability of the final results. Due to this fact and to the large number of hidden neurons allowed in the network structure, it is very probable that an overfitting problem arises in your investigation, so giving outputs with an overestimated goodness.

The standard way of acting in dealing with these problems is to consider a training , a

validation and a test set, by stopping the iterative training cycles when the error begins to increase in the validation set. Only a procedure like this could guarantee that the test data are not overfitted. Thus, consider this procedure! You will probably see that also the optimal number of hidden neurons will decrease. In particular, the class-frequency statistics on the validation and test sets should be the same of the total dataset.

Furthermore, an alternative procedure, very useful for small datasets, has been recently developed. See, for instance, Pasini and Modugno (2013), Atmospheric Science Letters 14, 301-305; Pasini (2015), Journal of Thoracic Disease 7, 953-960. In the latter paper also a treatment of the overfitting problem has been given in terms of training, validation and test sets. I suggest to apply also the so-called generalized leave-one-out procedure described in these papers, or at least to cite them as a reference to another useful procedure that could be adopted for avoiding overfitting.

Furthermore, there is no explanation about the way in which you choose the optimal number of hidden neurons. Empirical choice? Please, specify.

Again, the structure of the networks used has not been sufficiently specified. For instance, have you considered nonlinear transfer functions at hidden neurons and linear ones at the output? Please, give more details on probabilistic neural networks, too. Readers could be not familiar with them.

Finally, did you adopt an objective method for pruning the inputs? Do you know that the presence of collinear inputs bring no new information and could decrease the network performance? A pruning performed starting from consideration about linear and nonlinear correlations (through the so-called correlation ratio) will be welcome. See, for instance, Pasini and Ameli (2003), Geophysical Research Letters 30, 7, 1386, where this problem is addressed.

Minor comments

P. 10640, rows 1-7. Several other references should be given for neural network applications to environmental studies. Refer to S. Haupt et al. (2009), Artificial intelligence methods in the environmental sciences, Springer; W. Hsieh (2009), Machine learning methods in the environmental sciences, Cambridge.

P. 10640, rows 9-14. You talk about three simple tasks but describe just two of them.

P. 10648, row 16. Substitute "Study case" with "Case study". More generally, sometimes English is not up to standards. Please, ask help to a mother tongue colleague.

Interactive comment on Atmos. Meas. Tech. Discuss., 8, 10635, 2015.