

## ***Interactive comment on “Predicting the residual aluminum level in water treatment process” by J. Tomperi et al.***

**Anonymous Referee #2**

Received and published: 5 April 2013

Nice article about a common problem with drinking water production plants: you would like to know in due time if the quality of the product will exceed a certain level for a specific parameter. Prediction models can be a solution, especially when they are combined with Early Warning Systems. The authors explain how they tackled the issue in the case of residual aluminum by applying different types of models and techniques to get the best results. They do this by searching for a simple model. I can imagine that a prediction model has to be reliable, but why should it also be simple. This is not fully explained. Moreover, adding some complexity, for instance time lags and autoregressive and moving-average elements like is done in ARIMA/Box-Jenkins modelling could greatly improve on the quality of the prediction model. So at the end of 2.5 some major drawbacks of MLR as compared to more complex techniques should be added. One comment i would like to add: in statistics it is good practice to analyse the errors,

C337

the part of the output variable the model could not explain. Preferrably this noise shows up as "white noise". Page 244, Line 24 "because the good ability" > "because of its"; Page 245, Line 17 "certainly" > "certainty"; Page 245, Line 27 "variables to the amount" > "variables in a prediction model"; Page 246, Line 1 "to affect" > "have an effect on"; Page 246, Line 21 "ANNs has" > "ANN models have"; Page 248 Line 5 "effect on" > "have an effect on"; Page 249 line 5 : why can incorrect conclusions be found, please explain this; Page 249 Line 16 : "(xj), the values" > "(xj), of the values"; Page 251, Line 23 "due to it is very easy to understand" please modify this; Page 254, Line 2 "of original" > "of the original"; Page 254 Line 2/3 "Not a Number values" only Matlab users will be familiar with this term; Page 254 Line 15 It is stated that the accuracy is naturally affected by the preprocessing method, i don't understand "naturally", please add some explanation here; Page 254 Line 24 at the end of the article it is explained why variables of the drinking water were not used, "as early stage of the water" doesn't give me a clue, please use better description; Page 255 Line 28 "were found to overcome" > "showed a better performance than"; Page 256 Line 9 at the end of the article it is mentioned that peak values cannot be predicted, but the goal was to predict only the baseline, so this aspect is of no concern, maybe this goal can be added at the beginning of the article or in the abstract; Page 256 Line 17 suggestion: "for residual" > "for the baseline of residual".

---

Interactive comment on Drink. Water Eng. Sci. Discuss., 5, 243, 2012.