PO. 0109

On the wind energy production data deficiencies: simulation from statistical models

E. Pelikan, M. Brabec, K. Eben, O. Konar
Institute of Computer Sciences, Czech Academy of Sciences, Prague, Czech Republic, email: pelikan@cs.cas.cz

Abstract

It is a common practice in wind energy production modeling and forecasting to build various statistical models for electrical power produced from wind as if the data were perfect. That is, as if they contained only the variability connected with wind and power production processes.
On the other hand, the real data can contain variability related to various nuisance processes. For instance, the relation between wind speed and power produced can be easily modulated by superimposed wind farm management and maintenance processes.
It is clear that if the available data contain precise information e.g. about the periods in which maintenance of individual turbines was carried out, it is easy to reflect this knowledge when building a statistical model and hence to filter the problem out.
On the other hand, if such information is missing or if it is imprecise (e.g. in terms of timing of intervals when individual turbines are on and off for maintenance), the data are not providing unbiased account of the output curve and other features typically used for prediction purposes. In this study, we use simulations from formalized statistical models to assess how important the problem is and to show the practical consequences.

Objectives

Assessment of detrimental effect that an undocumented turbine switch-off might have upon the quality of the forecasting model. Check the potential biases and variance inflation potentially caused by the undocumented and unintentional power data modulation via simulation from a statistical model.

Ultimately, the pragmatic question is:

Is it worthwhile to ensure higher data quality (at higher costs) in terms of insisting on correct concomitant info about precise turbine maintenance timing, or not?

Methods

A GAM (generalized additive model) is used (selected via AIC):

$$y_t = \beta_0 + s_1(x_{t-1}) + s_2(x_{t-2}) + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t \qquad (1)$$

where:

y_t is electric power produced at time (hour) *t*,

x_t is wind velocity at time *t*,

$s_k(.), r_k(.), k = 1,2$ are smooth functions to be estimated from data (as smoothing splines),

$\beta_0, \beta_1, \beta_2$ are unknown coefficients to be estimated from data,

$\varepsilon_t \sim N(0, \sigma^2)$ is a random, (normally) distributed error term.

It pays off to reparametrize to:

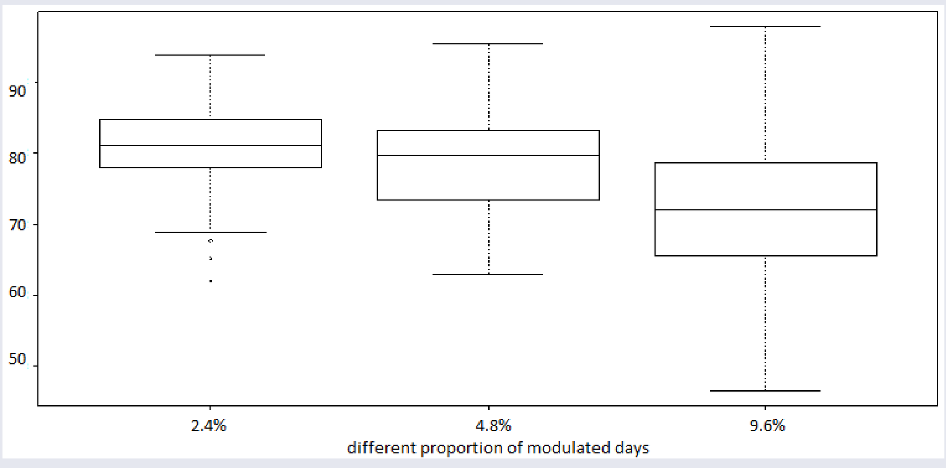
$$y_t = \beta_0 + r_1(x_{t-1}) + r_2(x_{t-2} - x_{t-1}) + \beta_1 y_{t-1} + \beta_2(y_{t-2} - y_{t-1}) + \varepsilon_t \qquad (2)$$

In our experiment we used hourly measured data from a wind farm in the Czech Republic from the period January 2008-March 2009. We simulated undocumented switch-off's on randomly selected periods. The proportion of modulated days attained three different levels.

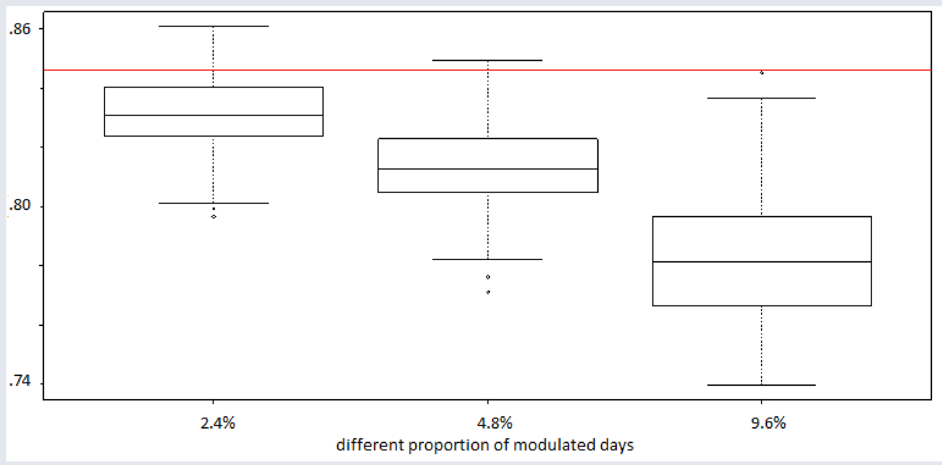
Results

Influence of the undocumented switch-off upon the estimated parameters of the model (2):

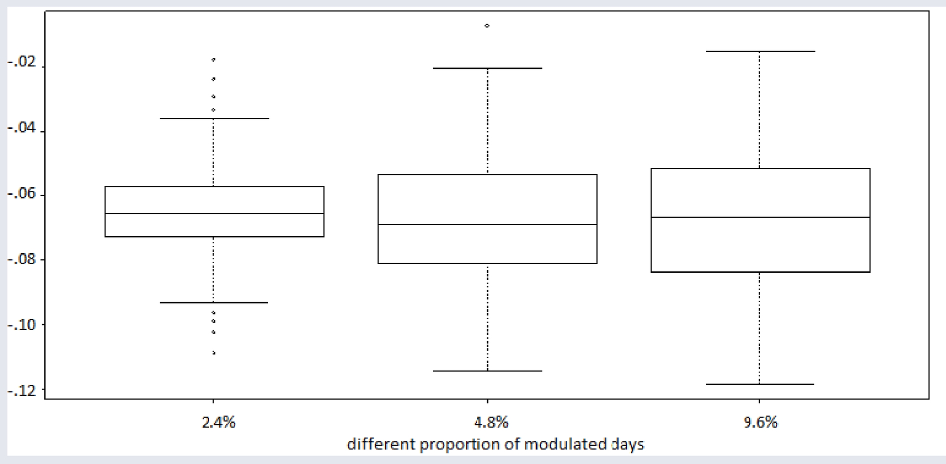
intercept, β_0



• past production, β_1

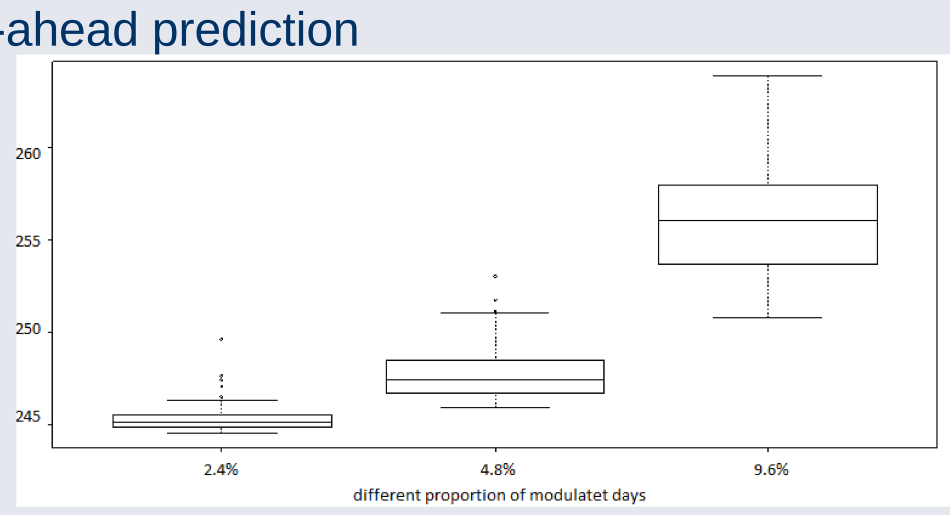


• past production, β_2

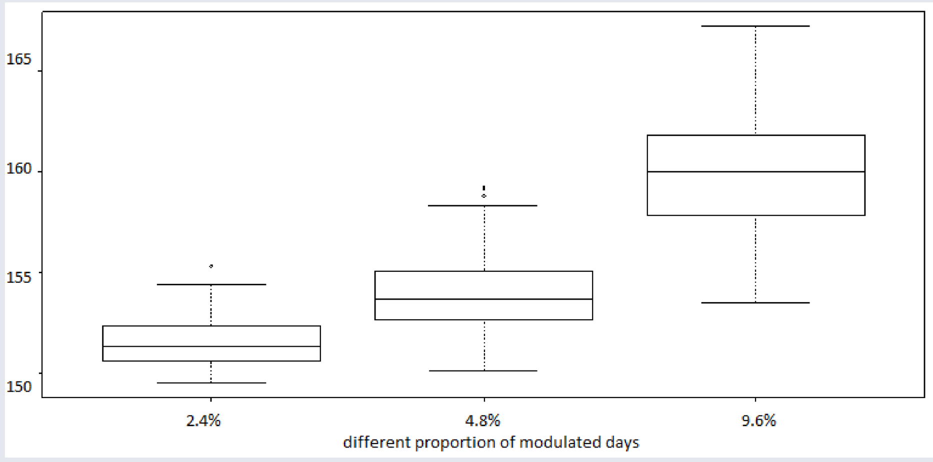


Influence of the switch-off upon the prediction performance of the model (2):

• RMSE of the 1-step-ahead prediction



• MAE of the 1-step-ahead prediction



Conclusions

Even a relatively small proportion of the undocumented switch-off can influence estimation and prediction performance of a forecasting model in a nontrivial and practically important way.

This is mainly because of the uneven influence the switch-off modulation has upon the bias of different regression coefficients of the model (2). The bias (and to a more limited extent also variability) influences mainly the most important β_1 (lag-1) production coefficient. The dynamics of clean and modulated data then differ, as do the predictions.

References

T. Hastie, R. Tibshirani, and J.H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed., Springer, Berlin, 2009.
G.E. Schwarz, *Estimating the dimension of a model*, Ann. Stat. 6 (1978), pp. 461-464.
S. N. Wood, *Generalized additive models*, Chapman and Hall/CRC Press, Boca Raton, 2006.

Acknowledgements

This work was partially supported by the project number LD12009 from Ministry of Education, Young and Sports, Czech Republic.

EWEA
THE EUROPEAN WIND ENERGY ASSOCIATION

EWEA Wind Power Forecasting Technology Workshop, Rotterdam, 3 – 4 December 2013

Call4Posters