

# Improving short-term forecasting during ramp events by means of Regime-Switching Artificial Neural Networks

C. Gallego<sup>1</sup>, A. Costa<sup>1</sup>, and A. Cuerva<sup>2</sup>

<sup>1</sup>Wind Energy Forecasting Group, CIEMAT. Avd. Complutense 22, 28040 Madrid, Spain

<sup>2</sup>IDR/UPM, E.T.S.I. Aeronáuticos, Universidad Politécnica de Madrid, Pza. de Cardenal Cisneros 3, 28040 Madrid, Spain

Received: 21 October 2010 – Revised: 4 March 2011 – Accepted: 8 March 2011 – Published: 21 March 2011

**Abstract.** Ramp events are large rapid variations within wind power time series. Ramp forecasting can benefit from specific strategies so as to particularly take into account these shifts in the wind power output dynamic. In the short-term context (characterized by prediction horizons from minutes to a few days), a Regime-Switching (RS) model based on Artificial Neural Nets (ANN) is proposed. The objective is to identify three regimes in the wind power time series: Ramp-up, Ramp-down and No-ramp regime. An on-line regime assessment methodology is also proposed, based on a local gradient criterion. The RS-ANN model is compared to a single-ANN model (without regime discrimination), concluding that the regime-switching strategy leads to significant improvements for one-hour ahead forecasts, mainly due to the improvements obtained during ramp-up events. Including other explanatory variables (NWP outputs, local measurements) during the regime assessment could eventually improve forecasts for further horizons.

## 1 Introduction

Since wind energy cannot be neither scheduled nor large-scale stored, wind power forecasting is required in order to minimize the impact of the variability of the wind. In particular, short-term forecasting is currently required by energy producers (in a daily electricity market context) and the transmission system operators (TSOs) in order to keep the balance of an electrical system. Within the short-term context, time-series based models (i.e. statistical models) have shown a better performance than Numerical Weather Prediction (NWP) models for horizons up to few hours (Giebel, 2003; Costa, 2005). These models try to learn and replicate the dynamic shown by a certain time series, for instance the power output time series of a wind farm.

Wind power ramp events are characterized by large gradients in the time series during relatively short time periods (few hours). Ramp events occur not very often, but they may have a real impact on an electrical system due to the unexpected variation in the generation side. Additionally, energy traders incur penalties in energy markets due to deviations from the scheduled energy (see Greaves et al.,

2009, and references therein). These impacts could be reduced with forecasting models focused on ramp events (Potter et al., 2009). Ramp events are usually motivated by specific meteorological processes (crossing fronts, fast changes in the local wind direction) involving several scales (synoptic, mesoscale, microscale). In other cases, wind power generation may experience drops related to other causes, such as voltage dips. A ramp event can be then considered as an “unexpected/atypical” dynamic due to a change in the underlying causes of the wind power conversion process. Consequently, traditional statistical models considering only one dynamic for the whole wind power time series may be inadequate. Section 2 describes the implementation of statistical models based on Artificial Neural Networks (ANNs) for the case of wind power short-term forecasting. In particular, Sects. 2.1 and 2.2 deal with a non-regime and a regime-switching strategy respectively. Main results of the study are gathered and discussed in Sect. 3.



Correspondence to: C. Gallego  
(cristobalj.gallego@ciemat.es)

## 2 Methodology

Let  $\{y_t\}$ ,  $t = 1, \dots, N$  be a discrete time series with  $N$  observations of hourly averaged wind power. Time series based models traditionally states that  $\{y_t\}$  follows a stochastic process like (Peña, 1987; Madsen, 2007):

$$y_{t+k} = f(y_{t-d+1}, \dots, y_t, \Theta) + \varepsilon_{t+k} \quad (1)$$

where  $k$  is the prediction horizon.  $f$  represents the deterministic component of  $y_{t+k}$  as a function of a set of  $d$  previous values  $[y_{t-d+1}, \dots, y_t]$  and a certain set of parameters  $\Theta$ .  $\{\varepsilon_t\}$  is the noise of the stochastic process, assumed to follow a white noise process  $\{\varepsilon_t\} \sim \mathcal{N}(0, \sigma^2)$ .

The purpose of time series models is to estimate the unknown function  $f(\cdot, \Theta)$  in order to provide accurate forecasts. In this way, an appropriate window size  $d$  and an optimized set of parameters  $\Theta$  have to be found from historical observations. A criterion based on the minimization of the Root Mean Squared Error (RMSE) is usually employed.

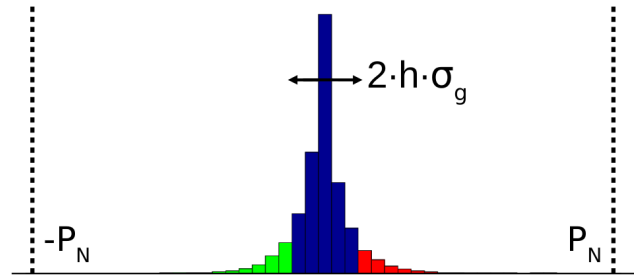
### 2.1 The single-ANN model

Non-linear models are usually required in order to estimate the unknown function  $f(\cdot, \Theta)$  introduced in Eq. (1). It is mainly due to the fact that there are many non-linear effects involved in the wind power conversion process, such as the behaviour of the atmosphere and the power curve of wind turbines. Multilayer Feedforward Networks, a kind of ANNs, represent a powerful tool to approximate any function (Hornik et al., 1989). In particular, we employ Multilayer Perceptrons to estimate  $f(\cdot, \Theta)$  based on the backpropagation learning algorithm (Lippmann, 1987).

The implementation of an optimal ANN requires an appropriate choice of:

- The ANN architecture, which comprises the number of layers and the number of neurons per layer. The set of parameters  $\Theta$  gathers weights and bias of the connections between neurons from one layer to the next one.
- The window size  $d$  (Kil et al., 1997).

In this work, the selection of the optimal ANN is based on a cross-validation criterion. First, several ANNs with different architectures (from 1 to 3 layers) and window size (from 1 to 10 lags) are trained over a historic dataset (training-set). The mentioned range of values are based on empirical experience. Then, the ANN with the lowest RMSE over the validation dataset is considered to provide the best approach of  $f(\cdot, \Theta)$ , being the validation dataset different from that dataset used for training the ANN. The performance of the selected ANN is evaluated over a third dataset (test-set), different from the previous ones.



**Figure 1.** Histogram of local gradient time series  $g_t$  (case for  $h = 1$ ).  $P_N$  stands for the rated power of the wind farm. Local gradients triggering Ramp-up and Ramp-down regimes are in red and green colour respectively.

### 2.2 A Regime-Switching model based on ANNs

Regime-Switching (RS) models consider that a certain time series evolves shifting between a number  $r$  of different dynamics. In this case, the stochastic process can be written as follows:

$$y_{t+k} = \begin{cases} f^{(1)}(y_{t-d(1)+1}, \dots, y_t, \Theta^{(1)}) + \varepsilon_{t+k}^{(1)}, & \text{if } s_t = 1; \\ f^{(2)}(y_{t-d(2)+1}, \dots, y_t, \Theta^{(2)}) + \varepsilon_{t+k}^{(2)}, & \text{if } s_t = 2; \\ \dots & \\ f^{(r)}(y_{t-d(r)+1}, \dots, y_t, \Theta^{(r)}) + \varepsilon_{t+k}^{(r)}, & \text{if } s_t = r; \end{cases} \quad (2)$$

where  $\{s_t\}$ ,  $s_t \in \{1, 2, \dots, r\}$  provides the current regime at time  $t$ . This approach permits to consider different generating processes  $f^{(r)}$  with different window sizes  $d^{(r)}$  and white noise processes  $\{\varepsilon_t^{(r)}\} \sim \mathcal{N}(0, \sigma_r^2)$ , which is consistent with the idea that the underlying causes of each regime are different. This work considers three different regimes: Ramp-up, Ramp-down and No-ramp regime. Each regime is modelled by means of an ANN.

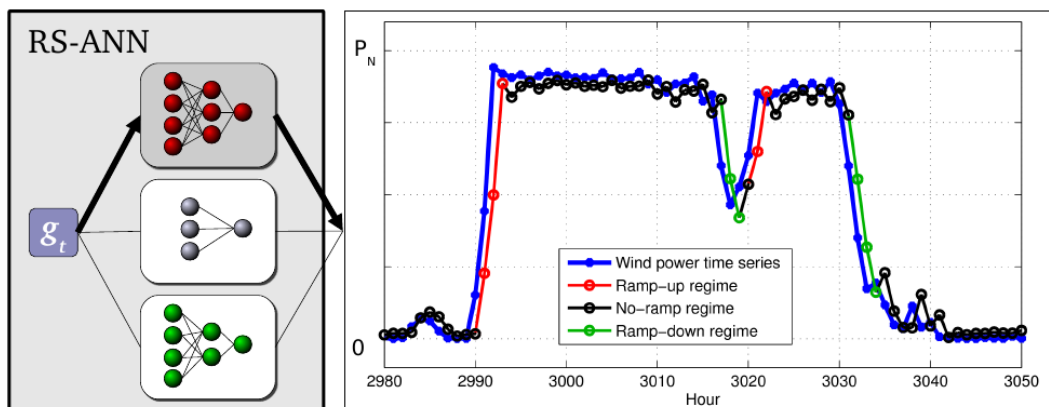
Additionally, a criterion have to be defined so as to determine  $s_t$  at each time step. Assuming that ramp events are motivated by specific conditions at certain meteorological scales (Cutler et al., 2007), meteorological expertise could provide  $s_t$  as a function of local observations, NWP outputs, etc. This work proposes an on-line regime assessment based on the observed wind power local gradient. The gradient time series  $\{g_t\}$  can be defined as:

$$g_t = y_t - y_{t-1} \quad (3)$$

Then, the regime assessment is based on the following proposed criterion:

$$\begin{aligned} g_t > h \cdot \hat{\sigma}_g &\implies s_t = 1 && \text{(Ramp-up)} \\ |g_t| \leq h \cdot \hat{\sigma}_g &\implies s_t = 2 && \text{(No-Ramp)} \\ g_t < -h \cdot \hat{\sigma}_g &\implies s_t = 3 && \text{(Ramp-down)} \end{aligned} \quad (4)$$

where  $\hat{\sigma}_g$  is an estimate of the standard deviation of  $\{g_t\}$  obtained from historical data.  $h$  is a scalar factor related to the ramp event definition. For example, higher values of  $h$  provide few but large ramp events (see Fig. 1). A suitable value for  $h$  has to be found during the training process.



**Figure 2.** RS-ANN model (left). One-hour ahead forecasting (right). The architecture of each ANN was optimized for each prediction horizon considered.

### 3 Results and discussion

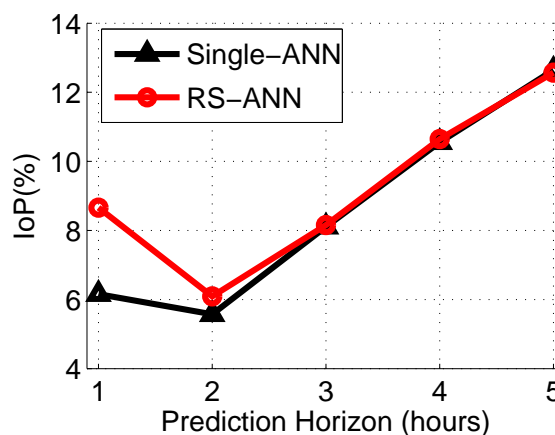
The case of a wind farm located in the complex terrain of Alaiz (north of Spain) has been considered. Three years of available power output data with an hourly resolution have been divided into three sets, the length of each set being one year: training-set (2001), validation-set (2002) and test-set (2003). In this sense, considering that the underlying dynamics in the wind power time series is strongly conditioned by atmospheric processes, it is important to remark that the one-year period was selected for each set so as to take into account monthly seasonality in the wind behaviour. Additionally, the three mentioned periods showed a similar relative frequency of ramp events (around the 5–10% of the time).

For a given prediction horizon  $k$ , a single-ANN model was performed following the methodology described in Sect. 2.1. Then, the RS-ANN (Eq. 2) was implemented by combining three ANNs with the regime assessment criterion given by Eq. (4). The optimal ANN architecture and the window size was tracked for each regime. Several values for  $h$  were also considered, varying from 0.2 to 2 in 0.2 steps. Figure 2 illustrates the proposed RS-ANN model.

The performance of the models have been obtained based on the Improvement over Persistence (IoP%). Persistence is the most common reference framework for time series forecasting models (Giebel, 2003), which states that the forecasting at time  $t+k$  is the last observation at time  $t$ . The IoP% of a certain model is defined as follows:

$$\text{IoP\%} = 100 \cdot \frac{\text{RMSE}_{\text{persistence}} - \text{RMSE}_{\text{model}}}{\text{RMSE}_{\text{persistence}}} \quad (5)$$

Figure 3 illustrates the results obtained for different prediction horizons. The improvement over Persistence tend to be modest since Persistence is difficult to improve on for few-hours ahead (Madsen et al., 2005). It can be seen that RS-ANN improved the single-ANN only for the case of one-hour ahead. It may be due to the fact that the on-line regime



**Figure 3.** Improvement over Persistence for different prediction horizons (case for  $h = 1$ ).

assessment was based on previous observations of the wind power time series. Hence, the ramp regime is triggered when the observed local gradient is already significant enough. We speculate that off-site measurements and NWP outputs could lead to relevant improvements also for further horizons. Another open question is how different time series resolution (for example, 10-min averaged data) may modify the conclusion here attained.

The case of one-hour ahead was analysed in depth. It was found that the optimal thresholds between regimes were given by a scalar factor  $h = 1$  (the estimated standard deviation of  $\{g_t\}$  during the training-set is  $\hat{\sigma}_g = 10.08\%$  of the rated power  $P_N$ ). However, the performance of the RS-ANN models did not show a smooth trend with respect to  $h$ . This may be due to the fact that the training process involves a numerical learning algorithm. This algorithm may evolve in a different manner depending on several factors, for instance, the initial conditions of the parameters.

**Table 1.** IoP% for the case of one-hour ahead forecasting.

Model	Globally	Ramp-up	Ramp-down
Single-ANN	6.16%	17.60%	18.55%
RS-ANN	8.66%	22.25%	19.55%

Additionally, a locally-weighted evaluation was carried out to obtain the IoP% during ramp-up and ramp-down events for both models (see Table 1). It was found that the improvement during ramp-up events was clearly higher than during ramp-down events. This fact suggest that ramp-up patterns are more regular (hence, better captured by the model) than ramp-down patterns. It could be due to the fact that ramp-down events may be motivated by a broader number of causes, including yaw-misalignments and wind turbine shut-down due to high wind speeds. If further research reinforces this effect, special attention should be paid to understand the underlying processes during ramp-down events in order to get more reliable predictive models.

Edited by: M. Sprenger

Reviewed by: T. Pagano and another anonymous referee

**sc | nat**  The publication of this article is sponsored by the Swiss Academy of Sciences.

## References

- Costa, A.: Mathematical/Statistical and Physical/Meteorological Models for Short-term Prediction of Wind Farms Output, Ph.D. thesis, Escuela Técnica Superior de Ingenieros Industriales (Universidad Politécnica de Madrid), 2005.
- Cutler, N., Kay, M., Jacka, K., and Nielsen, T. S.: Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT, *Wind Energy*, 10, 453–470, 2007.
- Giebel, G.: The state of the art in short-term prediction of wind power – A literature overview, Tech. rep., ANEMOS EU project, 2003.
- Greaves, B., Collins, J., Parkes, J., and Tindal, A.: Temporal Forecast Uncertainty for Ramp Events, *Wind Engineering*, 33, 309–320, 2009.
- Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, *Neural Networks*, 2, 359–366, 1989.
- Kil, R. M., Park, S. H., and Kim, S.: Optimum window size for time series prediction, *Annual International Conference of the IEEE Engineering in Medicine and Biology – Proceedings*, 4, 1421–1424, 1997.
- Lippmann, R. P.: An introduction to computing with neural nets, *IEEE ASSP magazine*, 4, 4–22, 1987.
- Madsen, H.: *Time series analysis*, Chapman & Hall, CRC, 2007.
- Madsen, H., Pinson, P., Kariniotakis, G., Nielsen, H. A., and Nielsen, T. S.: Standardizing the performance evaluation of short-term wind power prediction models, *Wind Engineering*, 29, 475–489, 2005.
- Peña, D.: *Estadística Modelos y métodos*, Vol. 2, Alianza Editorial, 2nd Edn., 1987.
- Potter, C. W., Gritmit, E., and Nijssen, B.: Potential Benefits of a Dedicated Probabilistic Rapid Ramp Event Forecast Tool, in: 2009 IEEE/PES Power Systems Conference and Exposition, Seattle, Vol. 1–3, 409–413, 2009.