Markov switching autoregressive modeling of wind power forecast errors

Roman Le Goff Latimier, E. Le Bouedec, V. Monbet

Markov Switching Autoregressive Modeling of Wind Power Forecast Errors

Le Goff Latimier R.
SATIE laboratory, ENS Rennes
Rennes, France
roman.legoff-latimier@ens-rennes.fr

Le Bouedec E.
LEGI, Univ. Grenoble
Grenoble, France
enzo.lebouedec@univ-grenoble.fr

Monbet V.
IRMAR, Univ. Rennes 1
Rennes, France
valerie.monbet@univ-rennes1.fr

***Abstract*—Forecast errors constitute the main hurdle to integrating variable renewable energies into electrical power systems. Errors are inherent to forecasting, although their magnitude varies significantly with respect to both the method adopted and the time horizon. Their dynamic and stochastic modeling is mandatory for power systems to efficiently balance out these errors. A Markov Switching Autoregressive – MS-AR – approach is proposed herein for wind power forecast errors. This particular model is able to identify weather regimes according to the forecast reliability. Such regimes are controlled by a Markov chain whose state – not directly observable – determines the AR model parameters. The statistical features of the data artificially generated by this model are very similar to those of the actual forecast error. This model is used to solve the optimal management of a storage associated with a wind farm. The resolution is performed by means of stochastic dynamic programming while comparing the proposed MS-AR approach with several other models. In this illustrative problem, a 15% reduction in operating costs is derived from a fine model of forecast errors.**

***Index Terms*—Forecast errors, Markov switching autoregressive, stochastic dynamic programming, wind power forecast**

I. Introduction

The integration of variable renewable energies into electrical systems is mainly hampered by the difficulty experienced in forecasting their electricity production [1]. This low predictability compels the power grid as a whole to compensate for their fluctuations in real time which may take the form of adjusting production [2] and consumption – Demand Side Management – or using storage [3]. This global problem, which involves all players within the electricity network, proves to be a particularly sensitive one when it entails planning for future operations on the basis of forecasts. Dynamically adapting the planned schedule becomes even more difficult when the deadline arrives and actual production is accessible.

Due to the burden inherent of forecast errors, a mojority share of the literature is devoted to forecasting renewable energy production. Several techniques are implemented and their complementarity makes it possible to refine the forecast gradually as more information becomes available as the deadline approaches [4], [5]. Such techniques include very short term forecasting, which can be carried out by means of imagery, satellite or fisheye camera. Statistical models can be introduced to extend forecasting horizons using time series [6], [7], [8] or neural networks [9]. Moreover global numerical weather prediction – NWP – models [10] provide a high-quality forecast with a several-day horizon.

However good they may be, these forecasts are definitely flawed by an irreducible error inherent in the weather forecasting task. Recent models provide information on the reliability of their forecast, in the form of an error range or ensemble forecast [5]. First, characterizing these errors [11] is important to the advancement of forecasting models. Moreover, this characterization is useful for electricity networks to anticipate sufficient operating reserves and infrastructure [12], [13].

Nonetheless, error modeling must extend beyond a statistical description [14]. The dynamic behavior of models is also critical: how does the error evolve over time, will it be prolonged or not [12]? Such information would be most helpful to deciding how the error should be counterbalanced. Calling upon storage resources is the easiest course of action should their capacity permit. But if starting up a backup power plant becomes mandatory, it would need to be anticipated. Moreover, it seems obvious that error modeling must indeed be stochastic [15], [16]. This demand is especially compelling in the case of NWPs, which only deliver their results every few hours.

Previous research on wind speed and wind power forecasting has extensively used Autoregressive Moving Average – ARMA – models [17], [12], [13]. Wind error forecasts requires other models since the dynamic is not the same: zero mean, varying volatility, less correlation in the long term for example. However, it may be considered obvious that the forecast error signal will most probably follow the same large scale structure as wind. Yet ARMA models do not capture the diversity of regimes that wind generation may encounter. Like any weather variable, wind generation is driven by weather types that can radically change its behavior. Several studies devoted to wind speed and wind power forecasts have applied this idea by introducing Markov Switching Autoregressive (MS-AR) models [7], [18], [8]. The main feature of these models is that their parameters are not unique but rather determined by a hidden state, whose evolution follows a Markov chain. Consequently, the signal characteristics can vary significantly from one time period to the next. This approach has proven its effectiveness in capturing different wind regimes. Weather types have also exerted an impact on forecast errors. Some

types of weather produce easily predictable wind conditions while others are much more chaotic and produce large errors.

This study focuses on the errors inherent in wind power prediction as regards their grid integration. Since the relation between wind speed and wind power is nonlinear, the model coefficients should be reevaluated before use on a wind speed forecast error. Nevertheless, the ability to accommodate regime changes – i.e. the main contribution of this article – would apply.

The objective of this study therefore is to describe wind power forecast errors by a Markov Switching Autoregressive model and highlight its relevance. Section II will be dedicated to the model presentation and its validation by ways of several statistical criteria. Other simpler models will also be introduced for comparison purposes. Section III will provide a representative application of such a forecast error model, namely optimal management of a storage associated with a wind power plant required to meet a generation commitment. This optimal management problem will be solved using stochastic dynamic programming. Solutions will be computed for various forecast error models in order to highlight the added value of a MS-AR model.

II. Modeling of wind power forecast errors

The time series of wind power forecast $\widetilde{P}$ and realization P are provided by [19]. Data are available from 2009 to 2018, with a 1 h time step. Let $P^\sharp$ denote the installed capacity, then the forecast error at time t is defined as

$$Y_t = \Delta P_t = \frac{P_t - \widetilde{P}_t}{P_t^\sharp} \tag{1}$$

An example of a trajectory is shown in Fig. 1 (top panel).

A. Model description

Markov Switching Autoregressive – MS-AR – models allow to describing a time series by a mixture of several autoregressive processes. These models were initially introduced by [20] to capture different regimes in market-related time series. Such models seem to be particularly relevant for wind power forecast errors which exhibit periods with low error (see Fig. 1 around 15th Nov.) and others with highly varying errors (see Fig. 1 end of Nov.). A latent variable π, called a hidden state or hidden regime, is introduced to capture transitions from one physical regime to another.

The dynamic of the latent variable π is driven by a Markov chain with M possible states. The transition probabilities are defined by an $M \times M$ matrix, denoted Γ. An element γ_{ij} represents the probability of switching from state i to state j. Moreover, at each time t, for a given state π_t, the forecast error is assumed to evolve like an AR(p) model whose coefficients depend on π_t. More precisely,

$$Y_t = a_0^{(\pi_t)} + a_1^{(\pi_t)} Y_{t-1} + ... + a_p^{(\pi_t)} Y_{t-p} + \sigma^{(\pi_t)} \epsilon_t \tag{2}$$

where ϵ is a Gaussian white noise.

Given a recorded time series and the MS-AR parameters, at each time step t, the probabilities of being affiliated with each hidden state – smoothing probabilities $\mathbb{P}(\pi_t | Y_1, \cdots, Y_T)$ – can be computed using a forward-backward (FB) algorithm [21]. The most likely sequence of latent states can therefore be deduced as illustrated in Fig. 1 for a three regime model. However when implemented in a real-time context, it is impossible to wait for future observations before estimating the most likely state. It then becomes necessary to use forward probabilities $\mathbb{P}(\pi_t | Y_1, \cdots, Y_t)$ instead of smoothing probabilities to predict the regime.

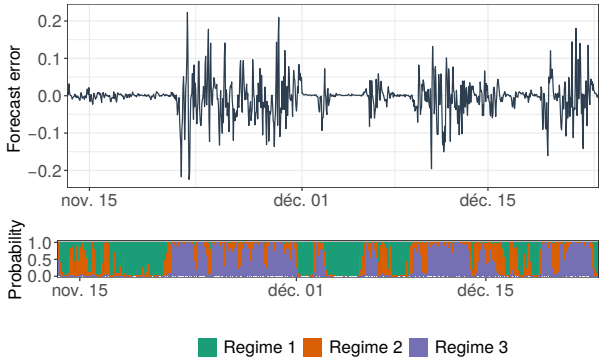


Fig. 1: Example of smoothing probabilities (bottom panel) of an MS(3)-AR(2) model for a given sequence of the time series of wind power production forecast errors (top panel).

The MS-AR model parameters – transition matrix Γ, AR coefficients $a_i^{(\pi)}$ and innovation standard deviations $\sigma^{(\pi_t)}$ – are calibrated using the Expectation-Maximization (EM) algorithm [22], which leads to a maximum of the likelihood of the model. The EM algorithm repeats two steps until convergence. The first step runs the FB algorithm for the current parameters values. The second step computes weighted empirical estimates of the AR coefficients. These weights are given by the smoothing probabilities.

B. Model selection and validation

Identifying a MS$(M)-$AR(p) model first requires selecting an order p for the AR models and the number M of hidden states. For models with hidden variables it is common practice to use the Bayes Information Criterion – BIC – because cross-validation is typically too expensive. BIC yields a compromise between model complexity and its likelihood. Fig. 2 reports – negatively oriented – BIC evolution for various models calibrated on the [19] wind power forecast error time series. Note that $M = 1$ corresponds to the AR model.

An initial effect shown in Fig. 2 is that the larger the value of M, the smaller the BIC value. Indeed, increasing the number of hidden states tends to more closely approximate the continuous variations of the actual meteorology. Nevertheless, the interpretability of results quickly becomes impractical when $M \geq 4$, in addition to a very small improvement of BIC for 3 to 4 regimes. Secondly, increasing the autoregressive order p offers a very small improvement. Since an order cannot be chosen with certainty by the BIC criterion, this study will be followed by the implementation into the optimal control of a

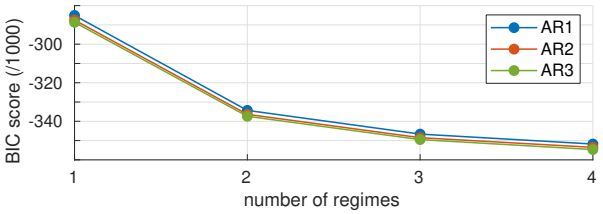


Fig. 2: Bayesian Information Criteria for various model parameters: number of states and order of the AR model.

representative system – see section III. The model complexity will then be bounded by the resolution feasibility. However, in an MS-AR model, increasing the AR order introduces as many new parameters as the number of possible hidden states. Moreover, a longer memory will increase resolution complexity when such a model is integrated into the optimal management of a power system. For this reason the remainder of the present section will consider the AR(2) and MS(3)-AR(2) models[1]. Estimated parameters are reported in Table I.

The AR parameters provide an information on the regularity of the time series inside the regimes: the higher the sum of a_1+a_2 the smoother the series. The observed time series inside the various regimes have a significantly different standard deviation σ. Although these coefficients were identified in an unsupervised manner using the EM algorithm, a physical interpretation of the results seems clear: hidden states are associated with weather types characterized by strong, moderate or weak predictability, respectively. These features are illustrated in Fig. 1, which shows a sample of the time series and its associated smoothing probabilities. The state 1 (green) is most likely when the errors have a very small amplitude. The state 3 (purple) appears to reflect high amplitude and low correlation errors. Intermediate situations are taken into account by a high probability of belonging to regime 2 (orange).

The transition matrix of the hidden Markov chain is diagonally dominant, which implies that the regimes are relatively stable. The measured mean duration of stay in regimes 1, 2, and 3 is respectively 10 hours, 7 hours 38 minutes and 16 hours 40 minutes. The weather conditions are in fact also relatively stable on an hourly scale.

Transitions between regime 1 (low and stable errors) and regime 3 (high and variable errors) are nearly impossible, thus suggesting no abrupt transition from a weather type where forecasts are reliable to another where they are highly uncertain. This finding seems to be consistent with an evolution in weather conditions and supports the notion that these regimes are well correlated with weather types.

For comparison, Table I reports the parameters of an AR(2) model fitted on the same time series; they are very similar to those of the third regime of the MS(3)-AR(2). Therefore, these two models may behave almost identically under some circumstances.

[1]The AR(1) and MS(3)-AR(1) versions will also be used in the following sections to compare the potential impact of different modeling approaches of various complexities on the final applications.

TABLE I: Fitted parameters of the MS(3)-AR(2) and AR(2)

MS(3) - AR(2)

	Transition matrix			AR parameters			
Reg.	1	2	3	a_0	a_1	a_2	σ
1	0.90	0.10	$8e^{-6}$	$5e^{-4}$	0.64	-0.08	$9e^{-6}$
2	0.04	0.87	0.08	$5e^{-4}$	0.73	-0.12	$3e^{-4}$
3	$5e^{-13}$	0.06	0.94	$-4e^{-3}$	0.67	-0.2	$3e^{-3}$

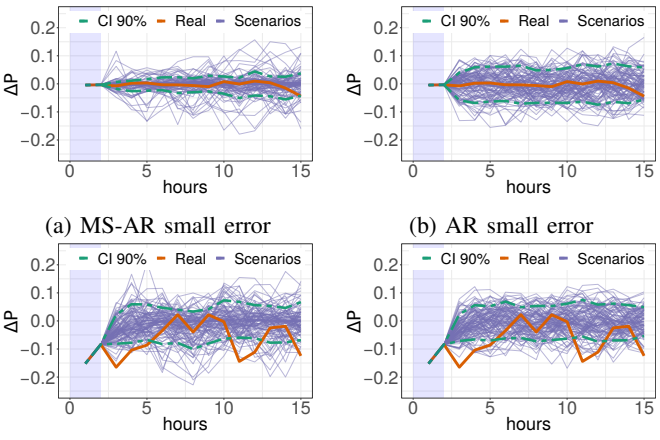
AR(2)

1	1.0	0.0	0.0	$-1.5e^{-3}$	0.68	-0.17	$4e^{-3}$

Fig. 3 presents some of the scenarios generated by the AR(2) and MS(3)-AR(2) models. Fig. 3a shows that the MS-AR model indeed acknowledges that it lies in a regime with small forecast errors, as illustrated by the 90% confidence interval (CI). In contrast, Fig. 3b points out that the AR model is incapable of taking this information into account, and the dispersion in scenarios remains the same all along. Fig. 3c shows that when initialized in a less predictable weather type, the MS(3)-AR(2) produces more variable scenarios and will occasionally reach more extreme values than those reached by the AR(2) (see Fig. 3d).

For both models, Table II reports the Mean Absolute Error (MAE), the model bias defined as the average error (BIAS), and the Root Mean Squared Error for forecast horizons of 12 hours and 24 hours. The MS-AR significantly improves both MAE and RMSE scores. The limited BIAS amplitude makes its rise insignificant. Overall performance on longer forecast horizons tends to decrease. Indeed, after a few forecasting steps, the model tends to generate a time series close to the mean of the stationary distribution. Therefore the positive impact of the initial probabilities belonging in a particular state has vanished.

Both the AR and MS-AR models facilitate the generation of synthetic data. This feature can be used to infer how some specific statistical characteristics are captured by comparing



(a) MS-AR small error (b) AR small error

(c) MS-AR large error (d) AR large error

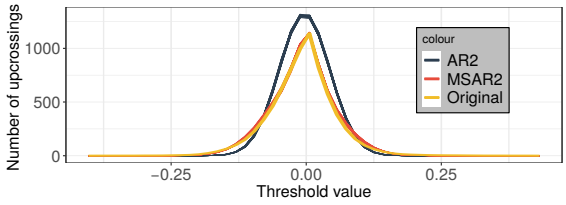
Fig. 3: Fifty 15-hour scenarios generated by the AR(2) and MS(3)-AR(2) models initialized on a test data sample and their associated 90% simulation interval

TABLE II: The range of scores per scenario obtained by the AR(2) and MS(3)-AR(2) models on the test dataset with 100 scenarios generated per time step

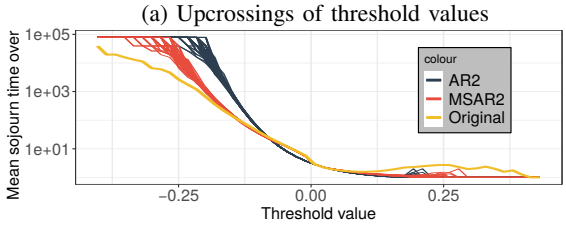
Metric	12H forecast horizon			24H forecast horizon		
	AR	MS-AR	variation	AR	MS-AR	variation
MAE	0.0450	0.0408	[9.3%]	0.0489	0.0452	[7.6%]
RMSE	0.0590	0.0544	[7.8%]	0.0632	0.0598	[5.4%]
BIAS	0.0049	0.0051	[-4.1%]	0.0054	0.0055	[-1.9%]

synthetic and original time series. Let's now focus on statistics that may be hard to reproduce with AR models. For instance, AR models generate trajectories exhibiting certain symmetries: the mean number of up-crossings for any value u – called threshold below – equals the number of up-crossings of $-u$. A set of 50 synthetic model-generated time series is compared to the original series in Fig. 4 through two metrics: mean duration of stay over a range of threshold values (bottom panel) and the number of up-crossings (top panel) for the same values. On both panels, the MS-AR model displays significant improvements compared to the AR model. Fig. 4a shows that the AR model overestimates variability for small errors, whereas the MS-AR overlays the original series' up-crossings. For both models, the 50 scenarios exhibit very little variation between one another. Fig. 4b indicates that both models fail to perfectly reproduce extreme wind power forecast errors – the reader's attention is drawn to the logarithmic scale. In this case, the distinction between scenarios is very clear, especially for extreme values. Such values are indeed rarely realized. Any slight variations are therefore more pronounced. However, the MS-AR provides significant improvements in both capturing the dynamic of small prediction errors and allowing for higher absolute errors to be reached.

This section has presented the MS-AR model, which is being suggested here to describe wind power forecast errors. After discussing the relevant orders for both the AR model and the number of hidden states, several statistical metrics were



(a) Upcrossings of threshold values



(b) Mean duration of stay over threshold values

Fig. 4: Statistical characterization of the real series as well as 50 synthetic series of the same length generated by the AR(2) and MS(3)-AR(2) models

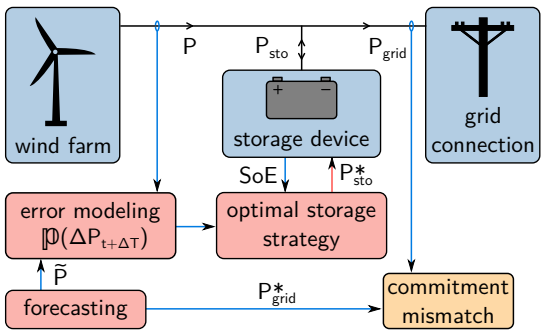


Fig. 5: Virtual power plant under commitment constraint being considered for this case study.

used to validate that these models are able to accurately reflect the behavior of forecast errors with a significant improvement compared to typical AR models. It must be highlighted however that the values of the model coefficients are very sensitive to the series on which they are identified. An illustration of this sensitivity is presented in Appendix A, where the same model is applied to another series [23]. Moreover, at the scale of a single wind farm, wind power forecast errors are likely to differ from errors aggregated over a vast geographic area.

III. Application to Optimal Storage Strategies

To extend beyond a statistical validation, the aim of this section will be to highlight how a better wind power forecast error model can improve final performance in a representative application. For this purpose a virtual power plant – VPP – consisting of the association of a wind farm and a storage unit is considered as in Fig. 5. A day ahead commitmnent constraint represents either a market context or a grid operation necessity. This commitment is assumed to be set equal to the forecast $\widetilde{P}$. The goal of this VPP then is to minimize the expected cost over time of the commitment gap penalties – assumed to be quadratic – and the losses – also quadratic[2]. The optimal management problem can therefore be stated in the following form:

$$\min_{P_{sto}(t)} \mathbb{E}_{\Delta P} \left\{ \sum_{\tau=t}^{\infty} \Big(\underbrace{\left(\Delta P^{\tau} - P^{\tau}_{sto}\right)^2}_{\text{commitment gap}} + \underbrace{a \cdot \left(P^{\tau}_{sto}\right)^2}_{\text{losses}} \Big) \right\} \tag{3a}$$

subject to $\forall t, \forall \tau$,

$$P^{\flat}_{sto} \leq P^{\tau}_{sto} \leq P^{\sharp}_{sto} \tag{3b}$$

$$0 \leq SoE \leq 1 \tag{3c}$$

$$SoE(t + \Delta T) = SoE(t) + \frac{\Delta T \cdot (P^{t}_{sto} \pm P^{t}_{loss})}{E_{sto}} \tag{3d}$$

where $a > 0$ is a loss coefficient and SoE denotes the state of energy of the storage unit. The sign of P_{loss} is the opposite of P_{sto}.

[2]These objective functions have deliberately been selected as basic for illustration purposes but can be replaced by any convex function.

Note: The capacity to shed producible is not considered herein. In a real-world situation, the plant operator could deoptimize conversion efficiency – via the wind turbine blade pitch – which would provide an additional decision variable. Although this set-up could be perfectly taken into account in the solving method presented, this possibility will not be addressed subsequently on because this study focuses on the impact of the forecast error model. Taking production shedding into account would indeed complicate the results interpretation step. First, it would introduce an asymmetry between the cost linked to positive and negative errors, and second, the commitment calculation would no longer equal the forecast expectation.

To support the stochastic and temporal coupling characteristics of the problem while minimizing the real-time computational cost, stochastic dynamic programming [24] is used to solve (3). The algorithm leads to an optimal strategy describing the best decision to make for any state vectir configuration. The result obtained therefore is not only the decision to be made in the current situation, but the optimal decisions for all possible configurations as well. The real-time control then consists of a simple interpolation of the matrix describing the optimal strategy. Stochastic dynamic programming solves the Bellman equation and it returns the costs associated with each state vector configuration when the optimal decision is applied. It is computed from the final system state at the T horizon and going backward in time.

$$V(T,X) = 0 \tag{4a}$$

$$\forall t < T, \forall x \in X,$$

$$V(t,X) = \min_{P_{sto}} \underbrace{f(x,P_{sto})}_{\text{instantaneous cost}} + \underbrace{\mathop{\mathbb{E}}_{\Delta P}\Big(V\big(t+\Delta T, f_{dyn}\left(x,P_{sto}\right)\big)\Big)}_{\text{expectation of the future cost}} \tag{4b}$$

The horizon T of the problem is not associated with any particular value. This final value is thus initialized to zero. However, no horizon value would be preferable in this context. Instead, it would be preferable to have an infinite horizon rather than become myopic beyond a given time frame. The resolution is therefore iterated back in time until the optimal strategy converges, so that it does not change from one iteration to the next. A strategy considering an infinite optimization horizon is then obtained.

f_{dyn} represents the dynamic system function; it links the current state and current control to the future system state:

$$x(t+\Delta T) = f_{dyn}\left(x\left(t\right), P_{sto}\left(t\right)\right) \tag{5}$$

In the present case, this dynamic function includes not only the deterministic component of (3d), but also a random component due to the evolution in forecast error that cannot be perfectly anticipated. All quantities involved in the forecast error model must therefore be included in the state vector in order to evaluate the anticipated forecast error at the next time step. Since this study is comparing several models, the state vector

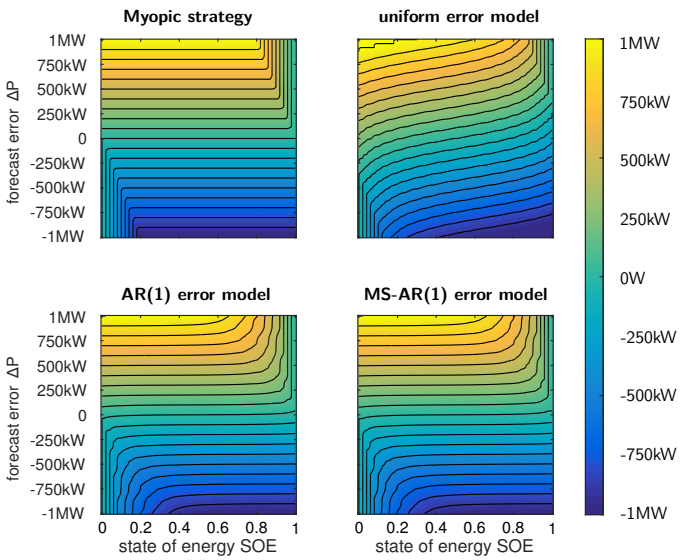


Fig. 6: Cross-sections of optimal storage strategies obtained with different forecast error models for $E_{sto} = 5\,\text{MWh}$. For the MS-AR model, the optimal strategy is three-dimensional. Only a cross section for $\pi = 3$ is represented.

composition will then differ from one resolution to another. The state of energy SoE and forecast error ΔP are mandatory. The hidden state π must be added in the case of MS-AR models.

$$x = \begin{pmatrix} SoE \\ \Delta P \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} SoE \\ \Delta P \\ \pi \end{pmatrix} \tag{6}$$

In the case of the resolutions involving a second order AR model, all possible forecast error configurations at the two previous time steps must be enumerated. ΔP then appears twice in the state vector, once for the current time step forecast error and after for the forecast error at the previous time step.

f denotes the convex instantaneous costs of the problem 3. This generic notation emphasizes that the resolution algorithm can handle any formulation depending on the system status x and its command P_{sto}. During the optimal strategy calculation, the cost f is evaluated for each configuration of the x state vector, hence the disappearance of the time index.

The Bellman equation (4) is solved using the various forecast error models. For each one, the result is a response surface associating every state vector configuration with the optimal storage power. Fig. 6 presents a number of cross-sectional views. The interpretation of these strategies is that if the system is in a configuration where the energy state is x on the abscissa and the forecast error is y on the ordinate, then the storage power described by the optimal surface must be applied. The first panel in this figure depicts the optimal strategy when no anticipation is made. The storage then seeks to perfectly compensate for the forecast error as long as SoE is sufficient. The iso-powers curves are therefore perfectly horizontal until the storage can no longer provide. On the

second panel, the forecast error is modeled by a uniform distribution. All error values are therefore equiprobable at the next instant, regardless of the current state. This approach entails a strong forward-looking behavior of the optimal strategy. Indeed, even in a state vector configuration with small error, it is necessary to anticipate that very large errors could occur during the next time steps. Consequently, the iso-powers curves are very steep, which means that the forecast error is never perfectly compensated, but merely attenuated. On the third panel, the optimal strategy is determined using an AR(1) model. Although rudimentary – this model uses only autocorrelation and a standard deviation – it allows for much better anticipation of future errors. More specifically, when the errors are of small amplitude, near total compensation is possible because it can be reliably anticipated that the error will remain of small amplitude during the next few time steps. On the last panel of the figure, the optimal strategy is determined using an MS(3)-AR(1) model. The overall behavior is therefore very similar to that of the strategy based on an AR(1). However, the optimal strategy based on an MS-AR model contains an additional dimension compared to the AR model. There is in fact an optimal strategy corresponding to each hidden state. We are only representing here a section corresponding to one of the three hidden states of the model – $\pi = 3$. Since the MS-AR model introduces the possibility to switch between hidden states, the corresponding strategy slightly differs from one hidden state to another because of the difference in standard deviations and error correlations. When used in real time, the probabilities of belonging to a hidden state are reconstructed according to the observations available up to the present time – forward probabilities. The storage power decision is then the weighted average of the decisions for each of the 3 hidden states.

Note: Although persistence is usually an excellent way to easily anticipate weather phenomena, this model leads to quite disappointing results. In the considered context, persistence would suggest that the currently observed forecast error would persist until the end of time. In such a case, any storage system will eventually be saturated regardless of its capacity. Therefore, the optimal storage strategy using such a model would be to do nothing.

These storage strategies have been applied over the last year of the time series of wind power forecast error [19], with the previous years being used to identify the models. At each time step, the state vector is formed, then the optimal strategy is interpolated. In Fig. 7, the temporal evolution of the stored power and corresponding energy state are displayed over a 700 hour sample. The forecast error is shown in blue. The storage power seeks to compensate it as much as possible. In red, the myopic strategy – without any anticipation – offers perfect compensation until saturation and then becomes basically useless. In yellow, the strategy based on a uniform noise model always seeks to maintain an energy state close to 0.5. The compensation for current errors is downgraded because of an overemphasis on possible future extreme errors. Autoregressive models – with or without a hidden state –

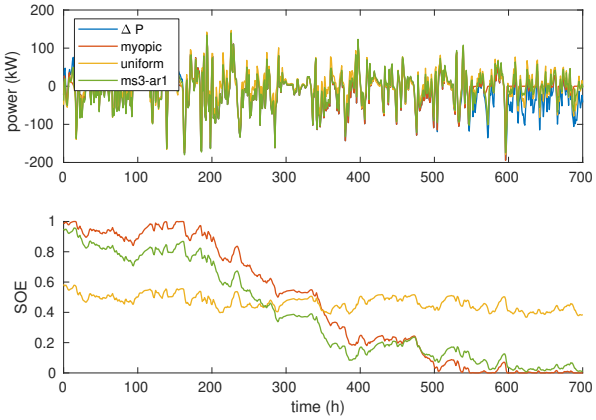


Fig. 7: Top: Time trajectories of the forecast error and associated stored power according to various strategies. Bottom: Corresponding state of energy of the storage device

exhibit an intermediate behavior that allows them to largely mitigate forecast errors while regulating the storage energy state, i.e. avoiding $SoE = 0.0$ or 1.0. Indeed, a few percent margin makes it possible to keep attenuating small amplitude variations.

However, battery capacity plays a crucial role in storage strategy performances: the best strategy is ineffective if it uses a too small storage. To dissociate the effects of battery capacity and management strategy, Fig. 8 shows – in blue, left axis – the evolution of total cost – associated with (3) – when storage is controlled without anticipatinf the future. Obviously, costs are reduced as capacity increases, even using such a rudimentary strategy. Yet, the addition of a simple forecast error model is sufficient to obtain a significant gain compared to such a short-sighted strategy. Using the latter as a reference, the other curves show the improvements obtained when other forecast error models are introduced. It should be noted that a gain is only possible if the storage is large enough, starting from a $1\,Wh/W$ ratio between storage capacity and wind rated power. It can be observed that a further improvement of up to 10% can be achieved when the autoregressive model order is

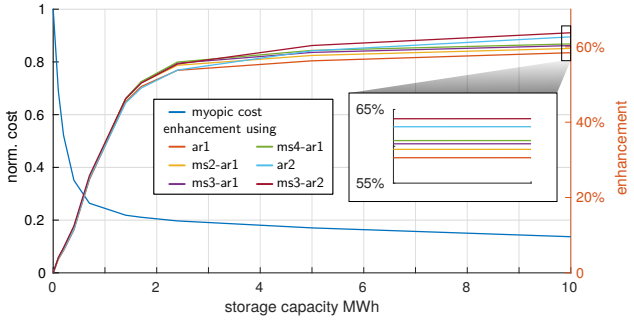


Fig. 8: Left axis: Normalized operating costs depending on battery size with myopic control – in blue. Right axis: Relative enhancement versus myopic control.

increased and hidden states are used. Performance improves when increasing the number of hidden states and the degree of the autoregressive model, with this latter factor being the more significant of the two, unlike what could be expected from the BIC criterion alone – see Fig. 2.

IV. Conclusion and further outlook

Since errors are inherent to any forecast, power systems must mitigate them in real time. A dedicated forecast error model of wind power has been proposed here, as it cannot be inferred from a wind power model. A Markov Switching Autoregressive model has been proposed and its relevance demonstrated on the basis of several statistical metrics. This MS-AR model was then used to solve a representative problem: managing a virtual power plant with a production commitment. The storage control strategy was optimized by using stochastic dynamic programming on the basis of several forecast error models. The contribution of effective error modeling was highlighted, with the management strategy based on an MS-AR model significantly improving overall performance.

One avenue of future work calls for adding covariates (temperature, pressure, etc.) into the model to better specify the meteorological conditions that could undermine forecast reliability. This would lead to non-homogeneous transition probabilities within the MS-AR model. In addition, using this model together with other predictive models (energy prices, solar production) may reveal couplings. Furthermore, taking spatial correlations into account for multisite forecast error models would offer an interesting development given that geographic effects are crucial for the power grid [25], [26].

Appendix

A. Comparison with a model identified on another time series

TABLE III: Parameters of the MS(3)-AR(2) identified on [23]

MS(3) - AR(2)

	Transition matrix			AR parameters			
Reg.	1	2	3	a_0	a_1	a_2	σ
1	0.93	0.07	$2e^{-4}$	$5e^{-4}$	1.34	-0.41	$3e^{-5}$
2	0.03	0.92	0.05	$4e^{-4}$	1.35	-0.42	$2e^{-4}$
3	$9e^{-4}$	0.27	0.73	$-2e^{-3}$	1.21	-0.34	$2e^{-3}$

The AR coefficients are quite dissimilar. However, the range of variances σ show that the regimes are differentiated by their predictability. The transition matrix is once again diagonal dominant. This results in persistent hidden states whose mean time of duration are measured at 14 hr 16 min, 12 hr 30 min and 3 hr 42 min.

References

[1] A. Fabbri, T. G. San Roman, J. R. Abbad, and V. H. M. Quezada, "Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market," *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1440–1446, 2005.

[2] C.-K. Woo, I. Horowitz, J. Moore, and A. Pacheco, "The impact of wind generation on the electricity spot-market price level and variance: The texas experience," *Energy Policy*, vol. 39, no. 7, pp. 3939–3944, 2011.

[3] N. Gast, J.-Y. Le Boudec, A. Proutière, and D.-C. Tomozei, "Impact of storage on the efficiency and prices in real-time electricity markets," in *4th international conference on Future energy systems*, 2013, pp. 15–26.

[4] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos, "Improvements in wind speed forecasts for wind power prediction purposes using kalman filtering," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 12, pp. 2348–2362, 2008.

[5] Y. Ren, P. N. Suganthan, and N. Srikanth, "Ensemble methods for wind and solar power forecasting—a state-of-the-art review," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 82–91, 2015.

[6] B. G. Brown, R. W. Katz, and A. H. Murphy, "Time series models to simulate and forecast wind speed and wind power," *Journal of climate and applied meteorology*, vol. 23, no. 8, pp. 1184–1195, 1984.

[7] P. Ailliot and V. Monbet, "Markov-switching autoregressive models for wind time series," *Environmental Modelling & Software*, vol. 30, pp. 92–101, 2012.

[8] P. Pinson and H. Madsen, "Adaptive modelling and forecasting of offshore wind power fluctuations with markov-switching autoregressive models," *Journal of forecasting*, vol. 31, no. 4, pp. 281–313, 2012.

[9] G. Kariniotakis, G. Stavrakakis, and E. Nogaret, "Wind power forecasting using advanced neural networks models," *IEEE transactions on Energy conversion*, vol. 11, no. 4, pp. 762–767, 1996.

[10] Y. Seity, P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, "The arome-france convective-scale operational model," *Monthly Weather Review*, vol. 139, no. 3, pp. 976–991, 2011.

[11] H. Bludszuweit, J. A. Domínguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 983–991, 2008.

[12] P. Haessig, B. Multon, H. Ben Ahmed, S. Lascaud, and P. Bondon, "Energy storage sizing for wind power: impact of the autocorrelation of day-ahead forecast errors," *Wind Energy*, vol. 18, no. 1, pp. 43–57, 2015.

[13] C. Wang, Z. Liang, J. Liang, Q. Teng, X. Dong, and Z. Wang, "Modeling the temporal correlation of hourly day-ahead short-term wind power forecast error for optimal sizing energy storage system," *International Journal of Electrical Power & Energy Systems*, vol. 98, pp. 373–381, 2018.

[14] L. Soder, "Simulation of wind speed forecast errors for operation planning of multiarea power systems," in *2004 International Conference on Probabilistic Methods Applied to Power Systems*, 2004, pp. 723–728.

[15] M. Lange, "On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors," *Journal of Solar Energy Engineering*, vol. 127, no. 2, pp. 177–184, 2005.

[16] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol. 12, no. 1, pp. 51–62, 2009.

[17] P. E. de Mello, N. Lu, and Y. Makarov, "An optimized autoregressive forecast error generator for wind and load uncertainty study," *Wind Energy*, vol. 14, no. 8, pp. 967–976, 2011.

[18] M. Yoder, A. S. Hering, W. C. Navidi, and K. Larson, "Short-term forecasting of categorical changes in wind power with markov chain models," *Wind energy*, vol. 17, no. 9, pp. 1425–1439, 2014.

[19] "BPA: Balancing authority load & total wind generation." [Online]. Available: https://transmission.bpa.gov/Business/Operations/Wind/

[20] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, pp. 357–384, 1989.

[21] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–22, 1977.

[23] "Actual and forecast wind energy feed-in - TenneT." [Online]. Available: https://www.tennet.eu/electricity-market/transparency-pages/transparency-germany/network-figures/actual-and-forecast-wind-energy-feed-in/

[24] D. P. Bertsekas, *Dynamic programming and optimal control*. Athena scientific Belmont, MA, 1995, vol. 1, no. 2.

[25] T. Qijun, W. Chengfu, L. Jun, and L. Zhengtang, "Research on modeling spatiotemporal correlation of wind power forecast error on multiple wind farms based on copula theory," in *2017 2nd International Conference on Power and Renewable Energy (ICPRE)*. IEEE, 2017, pp. 447–450.

[26] A. A. Ezzat, M. Jun, and Y. Ding, "Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting," *IEEE transactions on sustainable energy*, vol. 9, no. 3, pp. 1437–1447, 2018.