Modeling of the wind power forecast errors and associated optimal storage strategy

Abstract— **Production forecast errors are the main hurdle to integrate variable renewable energies into electrical power systems. Regardless of the technique, these errors are inherent in the forecast exercise, although their magnitude significantly vary depending on the method and the horizon. As power systems have to balance out these errors, their dynamic and stochastic modeling is valuable for the real time operation. This study proposes a Markov Switching Auto Regressive – MS-AR – approach. The strength of such a model is to be able to identify weather types according to the reliability of the forecast. These types are captured with a hidden state whose evolution is controlled by a transition matrix. The autocorrelation and variance parameters of the AR models are then different from one state to another. After having validated its statistical relevance, this model is used to solve the problem of the optimal management of a storage associated with a wind power plant when this virtual power plant must respect a production commitment. The resolution is carried out by stochastic dynamic programming while comparing the proposed MS-AR with several other models of forecast errors. This illustrative problem highlights the improvements made by a fine modeling of forecast errors.**

Keywords : forecast errors, wind energy, Markov Switching Auto Regressive, Stochastic Dynamic Programming

NOMENCLATURE

ΔP	forecast error of the wind power	W
ΔT	time step	h
$\widetilde{P}$	Wind power generation forecast	W
E_{sto}	rated capacity of the storage device	Wh
f	cost function of the storage management problem	–
f_{dyn}	dynamic equation of the system	–
P	actual wind power generation	W
P_{loss}	power of the losses int the storage device	W
P_{sto}	power exchanged with the storage device	W
SoE	State of Energy of the storage device	–
x, X	state variable of the system, set of the discretised values of the state variable	
AR	Auto Regressive	
MS-AR	Markov Switching Auto Regressive	

I. INTRODUCTION

The integration of variable renewable energies into electrical systems is mainly hampered by the difficulty of forecasting their electricity production [8], [30]. This low predictability compels the power grid as a whole to compensate for their fluctuations in real time. This may involve adjusting production [33] and consumption – Demand Side Management [28]– or using storage [9]. This global problem, which involves all the players in the electricity networks, is already sensitive when it comes to planning operations in advance on the basis of forecasts. Dynamically adapting the planned schedule is therefore all the more difficult when the deadline arrives and actual production is accessible.

Because of the burden of forecast errors, a major share of the literature is therefore devoted to forecasting renewable energy production. Several techniques are implemented and their complementarities make it possible to refine the forecast progressively when more and more information becomes available while approaching the deadline [18], [25]. These include global numerical weather prediction – NWP – models [26] that provide a quality forecast with a horizon of several days. Statistical models allow this forecast to be improved using time series [6] or neural networks [15], [27]. Finally, very short term forecasting can be done by means of imagery, satellite or by fisheye camera.

Nevertheless and however good they may be, these forecasts are necessarily tainted by an irreducible error. This error is inherent in the weather forecasting task. Recent models therefore provide information on the reliability of their forecast, in the form of an error range or ensemble forecast [25]. First of all the characterization of these errors [5] is important to help the progress of forecasting models. Moreover, this characterization is useful for electricity networks in order to anticipate sufficient operating reserves and infrastructure [10], [31].

However, error modeling must go beyond a statistical description [29]. Indeed, their dynamic behavior is also crucial : how does the error evolve over time, will it be prolonged or not [10] ? If such information are available, this will be most helpful to decide how the error should be counterbalanced. Calling upon storage resources is the easiest bet as long as their capacity permits but starting up a backup power plant when it is mandatory needs to be anticipated. Moreover, it seems obvious that error modeling must be stochastic modeling [16], [23].

This study focuses on wind energy forecast errors. Previous research on this issue has extensively used Auto Regressive Moving Average models – ARMA [20], [10], [31]. Indeed, one may consider as natural that the forecast error signal will most probably follow the same dynamics as the signal to be predicted. It is therefore relevant to rely on ARMA models which are widely used to predict wind power. However, these models do not capture the diversity of regimes that wind generation may encounter. Like any weather variable, it is driven by weather types that can radically change the behavior of this variable. Several studies devoted to wind speed forecasts have made use of this idea by introducing Markov Switching Auto Regressive MS-AR models [3], [34], [22]. The main idea of these models is that the parameters of the AR model are not unique but are determined by a hidden state whose evolution follows a Markov chain. The characteristics of the signal can consequently be very different from one time step to the next one. This approach has proven its effectiveness in capturing different wind regimes. However, weather types have also an impact on forecast errors. Some types of weather produce easily predictable wind conditions – hence small forecast errors – while others are much more chaotic and produce large errors.

The objective of this study is therefore to propose a modeling of wind production forecast errors by the mean of a Markov Switching Auto Regressive model, as well as to highlight its relevance. Section II will be dedicated to the presentation of these models and their validation by several statistical criteria. Other simpler models will also be introduced so as to allow comparison. Section III presents a representative application of such a forecast error model : the optimal management of a storage associated with a wind power plant that must meet a generation commitment. This optimal management problem will be solved using stochastic dynamic programming. As this algorithm relies on a stochastic modeling of the future, the resolution will be carried out several time on the basis of various forecast errors models in order to validate the added value of an MS-AR model. This resolution method will be carried out on the case study presented section IV. The section V will finally outline the results of this case study and how a better modeling of the forecast errors may improve the overall performance.

II. Modeling of the wind power forecast errors

A. Data description

Studying the forecast errors requires datasets with informations on both wind power actual realization and its associated forecast. While actual production can be directly measured, the production forecast first takes the form of a wind speed forecast delivered by meteorologists weather models. This forecast must then be transformed into a production forecast. An open dataset of the production forecast – provided by the wind generators to the local grid operator – and the realization is here provided by the Bonneville Power Administration [2]. The time series consists of 9 years of data, from 2009 to the present, with one input every 5 minutes. Years from 2009 to 2017 have been used as a train dataset and the period from January 1^{st} 2018 to May 2^{nd} 2018 as a test dataset. It includes aggregated forecast and realization over BPA balancing authority area. In order to improve the generality, this time series has been normalized according to the installed capacity of wind power generation [2] and reduced to an hourly time step by averaging. The convention used to define prediction errors is

$$\Delta P = P - \widetilde{P}$$

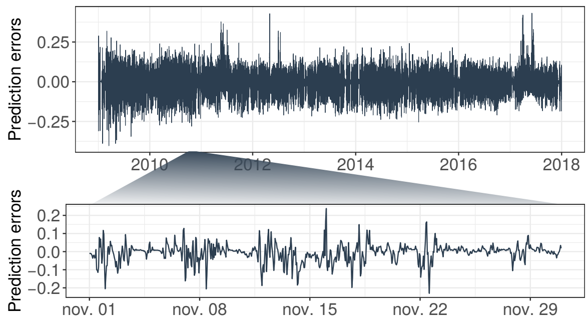


Fig. 1: Normalized prediction errors for 2009 - 2017 and a zoom on November 2010

where $\widetilde{P}$ stands for the forecast and P for the actual realization. The time series is shown in Figure 1.

B. Models descriptions

In order to model forecast errors, MS-AR models are here defined. The autoregressive AR models are also recalled, on the one hand because they are at the genesis of MS-AR, on the other hand because they will be used thereafter to provide a benchmark for modeling performance. Along this section, the generic notation Y_t is used to denote the observation at time t in any model. All of the models here introduced are then identified on the previously described time series of prediction errors (ΔP_t).

a) Auto-Regressive (AR): $AR(p)$ processes model the next time step as a linear combination (1) of the p previous time steps plus a white noise ϵ with unit variance and an intercept a_0. ϵ is also called *innovation*.

$$Y_t = a_0 + a_1 Y_{t-1} + ... + a_p Y_{t-p} + \sigma \epsilon_t \tag{1}$$

Conventional statistical models such as the AR and their extensions to SARIMA are commonly used in wind speed and power forecasting. However, with successive improvements in forecasts accuracy and as stated by [14], these models are now more considered as reference models.

b) Markov Switching Auto Regressive (MS-AR): MS-AR models allow to model a time series by a mixture of several AR processes. These models have been introduced initially by [11] to capture markedly different regimes in economy related time series. As weather related time series also exhibit weather regimes, typically cyclonic and anti-cyclonic (e.g. Figure 1) such models seems particularly adapted for wind power forecast errors. In particular, these unsupervised models demonstrated their ability to associate weather types with their latent state, by definition unobservable. The switches between the different regimes are described by a latent variable π whose dynamic is driven by a Markov chain. Therefore the Markov fundamental property (see (2)) is respected.

$$\mathbb{P}\left(\pi_{t+1} | \pi_t, \pi_{t-1}, ..., \pi_1\right) = \mathbb{P}\left(\pi_{t+1} | \pi_t\right) \tag{2}$$

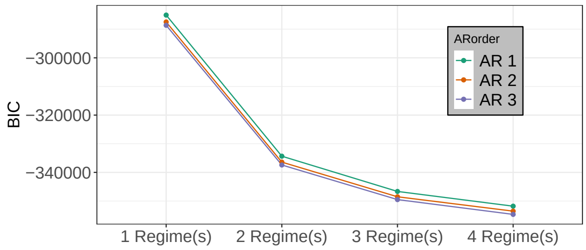


Fig. 2: Bayesian Information Criteria for different model parameters

Let us denote M is the number of possible states of the latent variable. The transition probabilities are defined by a $M \times M$ matrix denoted Γ. An element γ_{ij} represent the probability to switch from π_i to state π_j. Another fundamental property of the MS-AR model is that the observations are conditionally independent regarding to the current latent state and the p previous observations. In other words, at time t, given the latent state π_t,

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

In practice, given a time series, the MS-AR model is calibrated by maximum likelihood. Furthermore, the Viterbi algorithm allow to estimate for all t the smoothing probability $\mathbb{P}(\pi_t | Y_1, \cdots, Y_T)$ and deduce the most likely sequences of latent states. Usually, the number of regimes M and the order of the AR processes p are chosen according to the Bayes Information Criterion (BIC). This criterion computes a compromise between a model with a high likelihood and low number of parameters. In practice, the model with the smallest BIC should be retained.

C. *Results*

Identifying a MS-AR model first requires to decide an order p for the AR models and the number of hidden states M. Table I reports BIC values for MS-AR models with $p = 1$ to 3 and M varying from 1 to 4. Remark that $M = 1$ corresponds to the AR model.

TABLE I: BIC Values obtained for different MS-AR parameters

	AR(1)	AR(2)	AR(3)
M = 1	-285080	-287490	-288646
M = 2	-334339	-336405	-337408
M = 3	-346642	**-348507**	-349488
M = 4	-351787	-353521	-354654

A first effect shown in table I is that the larger M the smaller the BIC value. The improvement brought by the number of possible hidden states can be contextualized by the notion of weather type. This is a meteorological phenomenon that cannot but poorly be described in terms of a finite number of states. Nevertheless, the interpretability of the results quickly becomes impractical when $M \geq 4$. For this reason the 3 state hidden model is chosen later, despite the slight improvement made by the MS(4). RLGL : quasi plus d'amélioration avec l'AR(3) et le MS(4) Secondly the decrease of BIC while increasing the auto regressive order p brings a much smaller improvement. In a MS-AR model, increasing the AR order introduces a significant number of new parameters because of the several hidden states. Moreover a longer memory can be computationally expensive in some practical applications. As will be discussed in the following sections, these models are likely to be used to solve stochastic optimization problems. Some of the algorithms used may be sensitive to the dimension of the problem. One can think in particular of the case of dynamic programming and its curse of dimensionality. As adding an extra memory therefore implies adding a new dimension, an AR(2) model will therefore be preferred for the remainder of the study. The previous studies dedicated to the modeling of wind power forecast errors be the mean of AR models [10] – which did not take into account hidden states – were first order models to the best of our knowledge.

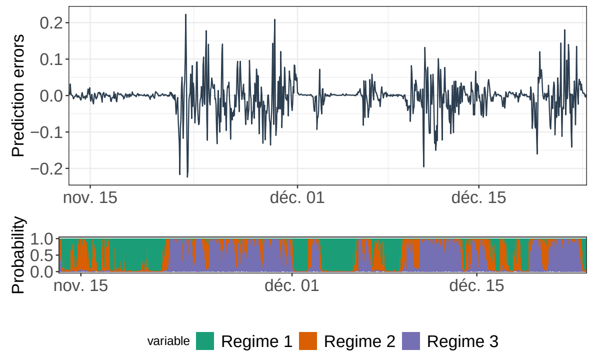


Fig. 3: Exemple of forward probabilities for a given sequence of the time series.

The remainder of this section will therefore consider the AR(2) and MS(3)-AR(2) models. The AR(1) and MS-AR(1) versions will also be used in the following sections to compare the impact that different modeling of various complexities may have on the final applications. Model parameters are shown in table II. The transition matrix is diagonally dominant which entails that the regimes are relatively stable. The mean duration of sojourn in the regimes 1, 2, and 3 are respectively 10 hours, 7 hours 38 minutes and 16 hours 40 minutes. Indeed, weather conditions are also relatively stable on a hourly scale. To have almost impossible transitions between states 1 and 3 is to say that one does not abruptly move from one weather type where predictions are reliable to one where they would be very uncertain. This seems consistent with an evolution in weather conditions and supports the idea that these states are associated with weather types. The observed time series inside the various regimes have different innovation variances. That means that they are associated with weather types characterized by their predictability (strongly, moderately or weakly in this case). This feature is illustrated in figure 3 where a sample of the time series and its associated forward probabilities are shown. One also notices that the parameters of the AR(2) are

TABLE II: 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	$-4e^{-3}$

very similar to those of the third regime of the MS(3)-AR(2). Therefore these two models may in some circumstances behave almost identically.

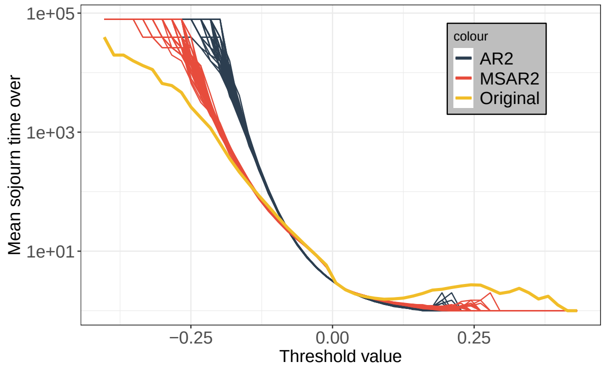
Both AR and MS-AR models allow to easily generate synthetic data. This feature can be used in order to infer how some specific statistics characteristics are captured by comparing synthetic and original time series. Figure 4 show the mean sojourn time over threshold values and the number of up-crossings for synthetic data generated by the previous models and the original series. On both the MS-AR model show significant improvements compared to the AR model. Figure 4b shows that the AR model overestimate variability for small errors and whereas the MS-AR overlays the original series's up-crossings. One can note that, for both models, the 50 scenarios have very little variation between one another. Figure 4a shows that both models fail to reproduce extreme prediction error. In this case, the distinction between scenarios is very clear, especially for extreme values. This is due to the fact that very slight variations in the maxima reached will by definition have a very strong impact on average sojourn time. However the MS-AR brings significant improvements in both capturing the dynamic of small prediction errors and allowing for higher absolute errors to be reached.

According to [19] and [13] a classical method to estimate the reliability of forecast errors is to compare the following metrics : Mean Absolute Error (MAE), the model bias defined as the average error (BIAS) and the Root Mean Squared Error (RMSE). More specifically, [19] propose to use the improvement score to compare two models which is defined in (3) where EC stands for one of the metrics mentioned before and new and ref point out respectively to the new model and the reference model.

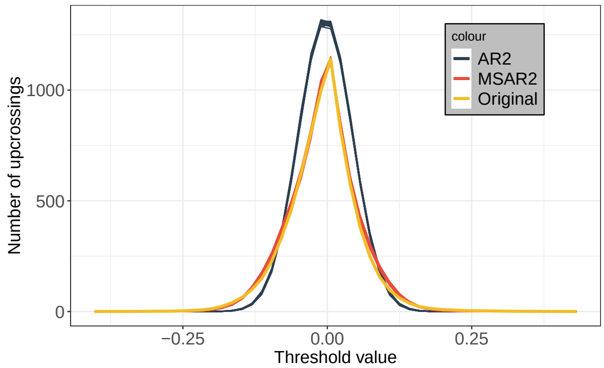
$$ISC(new, ref) = \frac{EC_{ref} - EC_{new}}{EC_{ref}} \quad (3)$$

These results, calculated on the test dataset, are presented in table III. One can notice that the MS-AR improves significantly on both MAE and RMSE scores while enhancing BIAS. However BIAS magnitude shows more that both models do perform well on this metric. After a few forecasting steps, the Markov chain tends to its stationary distribution and therefore the positive impact of the initial probabilities to belong in a particular state has vanished. This explains why overall performances on longer forecast horizons tends to decrease.

Figure 5 presents some sample scenarios generated by the AR(2) and MS(3)-AR(2) models. In the MS(3)-AR(2) case, the first hidden state of every scenarios are determined through random draws from a multinomial distribution characterized by the forward probability $\mathbb{P}(\pi_t|Y_1...Y_t)$ of the last known observation. Transitions from one hidden state to another at every time steps are then fully characterized by the transition matrix Γ. For both models, observations are generated through an auto regressive process. Figures 5d shows that, through the forward probabilities, the MS-AR model acknowledges that it is in a regime with small prediction errors (as illustrated by the 90% simulation interval). On the other hand, Figure 5c points out that the AR model is not able to take this information into account and the dispersion of scenarios remains the same all along. Figure 5b shows that when initialized in a less predictable weather type, the MS(3)-AR(2) produces scenarios that are more variable



(a) Mean sojourn time over threshold values

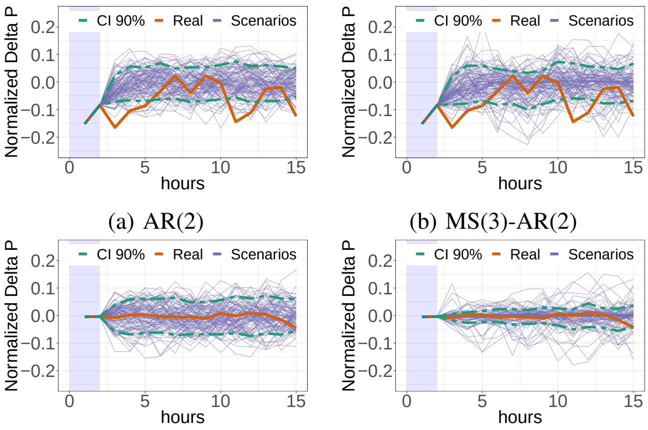


(b) Upcrossings of threshold values

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

TABLE III: Different scores per scenario obtained by the AR(2) and MS(3)-AR(2) on the test dataset with 100 scenarios generated per time steps

Metric	12H forecast horizon			24H forecast horizon		
	AR	MS-AR	ISC	AR	MS-AR	ISC
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%]



(a) AR(2) (b) MS(3)-AR(2)

(c) AR(2) (d) MS(3)-AR(2)

Fig. 5: 50 15-hour scenarios generated by the AR(2) and MS(3)-AR(2) models initialized on a sample of the test data and their associated 90% simulation interval

and will occasionally reach more extreme values than those reached by the AR(2) (see Figure 5a).

This section presented the MS-AR model that is here suggested to describe wind generation forecast errors. After having discussed the relevant orders for the AR model and for the number of hidden states, several statistical scores were used to validate that these models allow to reflect accurately the behaviour of forecast errors with a significant improvent in comparison with usual AR models. However the reader's attention should be drawn to the fact that the value of the model coefficients are very largely sensitive to the series on which it is identified. An illustration of this sensitivity is presented in Appendix A, where the same model is applied to another series [1]. In order to go beyond the validation of the statistical relevance of the proposed model, the following sections propose to use them to solve a canonical problem in order to evaluate the improvement on the final performance that can be brought about by the modeling of forecast errors.

III. Optimal storage strategies

Because of the forecast errors of wind power generation – as well as generally speaking variable renewable energies – their management within power systems is made challenging in many different ways. Examples include their integration into energy markets [21], [12], compliance with grid constraints for voltage and frequency [17], supply of reserves [32], *et cetera*. As all of these contexts come under the hazard of forecast errors, using models to anticipate them can therefore being useful in many different ways.

In many cases, combining the renewable power plant with a storage unit is a particularly relevant solution to overcome the intrinsic restrictions of variable renewable energies [35], [36]. The issue of optimal management of this storage – how much power must be exchanged with the storage unit at any one time – is then a problem that has two characteristics that require the use of a forecast error model.

- The problem is dynamic : using a storage unit introduces a temporal coupling between successive moments due to the integrating behaviour of the storage. This behaviour can be described by a dynamic equation – which is considered deterministic – over a time step ΔT :

$$SoE(t+\Delta T) = SoE(t) + \frac{\Delta T \cdot (P_{sto} - P_{loss})}{E_{sto}} \quad (4)$$

where SoE designates the state of energy of the storage unit – which is bounded between 0 and 1, P_{sto} stands for the setpoint of the power exchanged with the storage, P_{loss} indicates the internal losses of the device and E_{sto} refers to the storage capacity, *i.e.* the maximum energy that can be stored into it. This temporal coupling will therefore make it necessary to anticipate future moments, hence the need for a model describing the evolution of the error.
- The problem is stochastic : since forecast errors are inherently imperfectly controlled, the model used to anticipate them must be a stochastic model. Only then will it be possible for each agent to determine what could be the best decision to anticipate an uncertain future, depending of his own risk policy.

A. Optimal management of a storage

The scope of this study aims to highlight how a better modelling of forecast errors can improve the final performance of the renewable plant. In this section, the problem of managing a storage associated with a wind farm will therefore be used for this purpose. This problem is at this stage formulated in a generic way and can be adapted to as many situations as possible. The resolution method used is also chosen to require few assumptions and to be easily adaptable. In the next section, this method will be applied to a basic and representative case study to quantify the contributions of the forecast error model.

A wind power plant associated with a storage unit is therefore considered. The goal of this virtual power plant is to minimize a cost over time. This cost function is the sum over time of instant costs. The optimal management problem can therefore be stated in the following form :

$$\min_{P_{sto}(t)} \mathop{\mathbb{E}}_{\Delta P} \left\{ \sum_{\tau=t}^{\infty} \Big(f\left(x\left(\tau\right), P_{sto}\left(\tau\right), \Delta P\left(\tau\right)\right) \Big) \right\} \quad (5a)$$

such that, $\forall t, \forall \tau$,

$$P_{sto}^{\flat} \leq P_{sto} \leq P_{sto}^{\sharp} \quad (5b)$$

$$0 \leq SoE \leq 1 \quad (5c)$$

where f refers to the cost function which allows to evaluate the instantaneous cost at each moment. This depends on the system state x, the command P_{sto} and the forecast error with random behavior ΔP.

Note : the capacity to shed productible is not taken into account in this case. In an actual situation, the renewable plant operator can deoptimize the conversion efficiency – via the wind turbine blade pitch or the PV panel operating point – which would provide an additional decision variable. Although this may well be taken into account in the resolution method that is presented, this possibility will not be exploited later on because this study focuses on the impact of the forecast error model. Indeed, taking load shedding into account would have several consequences which would make more difficult the interpretation of the results. First it would introduce an asymmetry between the cost linked to positive and negative errors. Secondly the calculation of the commitment would no longer be equal to the expectation of the forecast.

B. Resolution using Stochastic Dynamic Programming

Several methods may be considered to overcome such a problem. Their required properties are to be able to support the stochastic and multitemporal characters of the problem. Methods like Model Predictive Control could be considered. Nevertheless, such methods would require that the calculation of the optimal decision should be carried out at each moment according to the present situation, which would potentially require a substantial real-time calculation burden.

In order to support the stochastic and temporal coupling characteristics of the problem while minimizing the real-time computation cost, the resolution of the problem (5) is performed here by the use of stochastic dynamic programming [4]. This algorithm allows to establish an optimal strategy which describes the best decision to take for any configuration of the state vector. The result obtained is therefore not only the decision to be taken in the current situation, but the optimal decisions for all possible configurations. The real-time use of this strategy then consists of a simple interpolation of the matrix describing the optimal strategy.

Stochastic dynamic programming is based on the resolution of the Bellman equation. This allows the calculation of the costs associated with each configuration of the state vector when the optimal decision is applied. It is calculated from the final state of the system at the T horizon and going back in time.

$$
\begin{aligned}
V(T,X) &= 0 && \text{(6a)}\\
\forall t < T, \forall x \in X, &&&\\
V(t,X) &= \min_{P_{sto}} \underbrace{f(x,P_{sto})}_{\text{instantaneous cost}} + && \text{(6b)}\\
&\quad \underbrace{\mathop{\mathbb{E}}_{\Delta P}\Big(V\big(t+\Delta T, f_{dyn}\left(x,P_{sto}\right)\big)\Big)}_{\text{espectation of the future cost}} &&
\end{aligned}
$$

where

- The horizon T of the problem is not associated with any particular value. This final value is therefore initialized to a zero. However, no horizon value could be preferable in the context of managing storage associated with a wind power plant. Instead it would be desirable for the problem to have an infinite horizon rather than become myopic beyond a given time frame. The resolution is thus iterated back in time until the optimal strategy converges, so that it does not change from one iteration to the next. We then obtain a strategy considering an infinite horizon of optimization.
- f_{dyn} represents the dynamic function of the system. This links the current state and the current control to the future state of the system :

$$x(t+\Delta T) = f_{dyn}\left(x\left(t\right), P_{sto}\left(t\right)\right) \tag{7}$$

In the present case, this dynamic function includes not only the deterministic component of the equation 4, but also a random component due to the evolution of the forecast error which cannot be perfectly anticipated. Therefore all quantities that are involved in the forecast error model must be included in the state vector in order to be able to evaluate the expectation of the forecast error at the next time step. As this study compares several models, the composition of the state vector will then differ from one resolution to another. Prediction error models described as uniform noise, persistence or a first-order AR model will have the state vector :

$$x = \begin{pmatrix} SoE \\ \Delta P \end{pmatrix} \tag{8}$$

In the case of the resolutions involving a second order AR model, all possible configurations of the forecast error at the two previous time steps must be enumerated. The forecast error ΔP then appears twice in their state vector, a first time for the current time step forecast error and a second time for the forecast error at the previous time step. Finally resolutions based on MS-AR models add the hidden state into their state vector :

$$x = \begin{pmatrix} SoE \\ \Delta P \\ \pi \end{pmatrix} \tag{9}$$

- the f notation represents the instant costs of the problem. This generic notation emphasizes that the method which is here used to solve the optimal management problem is almost fully independent from the formulation adopted to estimate instantaneous costs. Any formulation depending on the system status x and its command P_{sto} can be adopted. The section IV will present the example of formulations adopted for the application case. 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.

As described during section III, the resolution of this problem is based on a model evaluation the expectancy of the future cost, depending on the stochastic variables of the problem. To compare the added value of a fine forecast error model, the resolution will therefore be carried out several times using the following models.

- a uniform distribution of the errors : all possible values are supposed to be equiprobable.

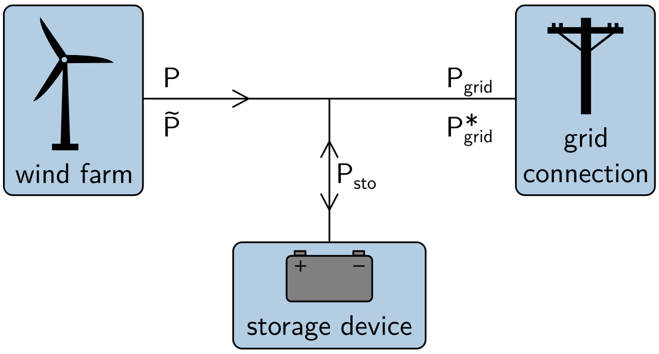


Fig. 6: Situation being considered for the case study.

- a persistence of the previous error : the error at the previous time step is supposed to be equal to the current value.
- AR(1) and AR(2) models.
- MS(3)-AR(1) and MS(3)-AR(2) models.

Finally a last resolution is carried out where the future is not anticipated. This resolution is blinded and does not take into account the future forecast errors.

IV. Case study

In order to present an application of the optimal strategy described in the previous section, a case study is introduced here. The objective of this case study is to present as straightforwardly as possible the impact that the forecast error model can have on the decisions to be made. A situation as simple as it can is therefore under consideration. A wind power plant is considered in association with a storage unit. This virtual power plant is supposed to be compelled to a production commitment. It can arise from a participation in an energy market, a constraint linked to the electricity network, a regulatory obligation, *et cetera*. In the case where the cost functions of the problem are symmetrical and the possibility of load shedding is not taken into account, the optimal commitment of the plant is then equal to the expectation of the forecast :

$$P^*_{grid} = \widetilde{P} \tag{10}$$

The objective function to be minimized encompass two terms :

- losses within the storage system. Many technologies of components can be considered in association with a renewable power plant depending on the reactivity and amount of stored energy required. Nevertheless, losses are inherent in each exchange of energy while it is charged or discharged. The case of lithium-ion batteries is considered here, as this technology is used in a wide variety of applications. A quadratic model is adopted to describe the losses :

$$P_{loss} = a_{loss} \cdot P^2_{sto} \tag{11}$$

Here it is necessary to decide how to quantify this lost power in an objective function. In fine this choice depends on the case study and the bussiness model considered. We choose here to adopt a quantification in an equivalent primary energy in order to keep as much as possible the generality of this study. The wind power plant is therefore considered to be characterized by a yield over life cycle η_{wind}, i.e. the total energy it can produce – depending on site weather conditions – divided by the primary energy that was required for its construction. The losses are therefore a degradation of this efficiency and an associated loss of primary energy :

$$C_{loss} = \frac{\Delta T \cdot P_{loss}}{\eta_{wind}} \tag{12}$$

- the penalty for deviations from the commitment. In the various situations where the production commitment obligation is present, a penalty is associated with non-compliance. In a similar way to the quantification of the cost of losses, the actual quantification of a deviation from the commitment will depend on the final application case. In this study, therefore, we only consider the case of a quadratic penalty for the mismatch :

$$C_{mis} = a_{mis} \cdot (\Delta P - P_{sto})^2 \tag{13}$$

in the case where the commitment is set as equal to the forecast expectancy.

TABLE IV: Values of the parameters and coefficients used within the case study RLGL : ATTENTION il faut le mettre à jour

error model	myopic	uniform	AR(1)
$\mathrm{kWh_p}/day$	32.6	20.4	12.5
error model	AR(2)	MS-AR(1)	MS-AR(2)
$\mathrm{kWh_p}/day$	12.4	12.2	12.1

V. Results and discussion

The problem that is described in the section IV is solved several times using the Bellman equation (6) using the different forecast error models described in the section II. This section will first present the observable consequences on the optimal strategies thus obtained. In a second step these optimal strategies will be applied to the time series of forecast errors in order to compare temporal behaviors and calculate the obtained performances.

A. Optimal storage strategies

The result of the resolution of the Bellman equation is a response surface associating the optimal storage power with every configuration of the state vector – discretized on a grid fine enough to guarantee a satisfactory interpolation between two grid points. These response surfaces are therefore of the same dimension as the state vector. The Figure 7 represents some of the calculated strategies. The interpretation of these illustrations is that if the system is in a configuration where the energy state is x on the abscissa and the forecast error is

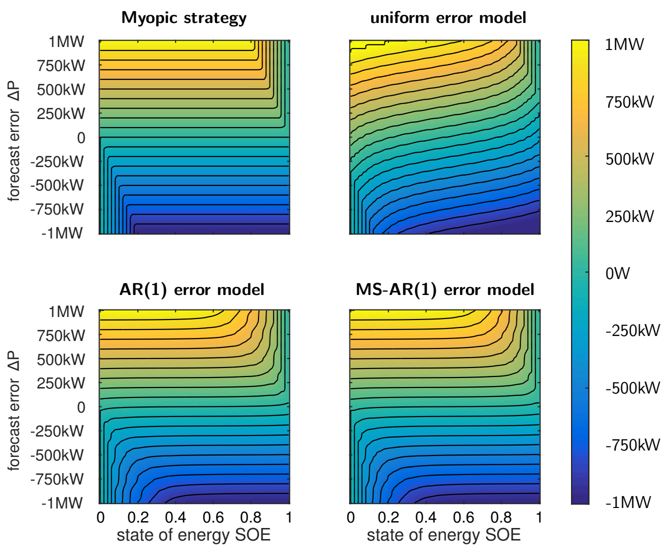


Fig. 7: Cross-sections of optimal storage strategies obtained with different forecast error models for $E_{sto} = 5\,\text{MWh}$. In the case of the *Markov Switching Auto Regressive* model, the cross section is in the $\pi = 3$ hidden state.

y on the ordinate, then the storage power described by the optimal surface must be applied.

- The first panel in this figure represents the optimal strategy when no anticipation is made. The storage then tries to perfectly compensate for the forecast error until it is too much or too little charged. The iso-powers curves are therefore perfectly horizontal, until the storage can no longer provide.
- On the second panel, the forecast error is modelled by uniform noise. All error values are therefore equi-probable at the next instant, regardless of the current state. This entails a strong forward-looking behaviour of the optimal strategy. Indeed even in a configuration of the state vector where the error is low, it is nevertheless necessary to anticipate that very strong errors can occur in the next time steps. As a result, the isopowers curves are very steep, which means that the forecast error is never perfectly compensated, but only 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 much better anticipation of future errors. Especially when the errors are of small amplitude, almost total compensation is possible because it can be reliably anticipated that the error will remain of small amplitude during the next few time steps.
- Finally, on the last panel of the figure, the optimal strategy is determined using a MS-AR(1) model. The overall behaviour is therefore very similar to that of the strategy based on an AR(1). However, this is a cut for only one of the three hidden states of the model. Indeed, since the MS-AR model introduces the possibility to switch from one hidden state to another, the strategies corresponding to these three states differ finely according to the standard deviation and the correlation of errors from one hidden state to another. When being used in real time, the probabilities of belonging to a hidden state are reconstructed according to the observations available up to now – for example by a Viterbi algorithm. 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 of anticipating weather phenomena with ease, this model here leads to dreadful results. Indeed, within the context of the resolution here presented, the persistence suggests that the value of the forecast error that is currently observed would persist until the end of time. In such a case, any storage system regardless of its capacity will eventually be saturated. The optimal storage strategy when this model is used is therefore to do nothing. This case is mentioned here to draw the reader's attention to this adverse effect.

B. Application to a forecast error time serie

The previously described storage strategies are applied to the time series of the wind power forecast error which has been presented during section II to identify the various models. The initial state of energy of the storage is set to $SoE(t = 0) = 0.5$. At each time step of this simulation, the state vector is formed, then the optimal strategy is interpolated to determine the storage power. The dynamic equation of the system is then applied to progress to the next time step. In the case of MS-AR models, the decision being applied is the average of the optimal decisions for each hidden state, weighted by the probabilities of belonging to each state. These forward probabilities are established on the basis of the observations available so far. The Figure 3 represents an example of temporal evolution of these probabilities. The succession of regimes is clearly observable with the naked eye : periods of very small errors are followed by periods when errors are large and chaotic. This succession is translated by the evolution of the probabilities of belonging to each hidden state, represented in the second panel of the figure. The state 1 – in red – is most likely when the errors have a very small amplitudes. The state 3 – in blue – appears to reflect high amplitude and low correlation errors. Intermediate situations are taken into account by a high probability of belonging to regime 2 – in green.

On the Figure 8, the temporal evolution of the stored power and the corresponding energy state are represented over a 400 hour sample. The forecast error is shown in blue. The stored power tries to compensate it as much as possible.

- The basic strategy – in red – does not anticipate the future. Therefore it offers perfect compensation but quickly causes storage saturation and then becomes almost useless.
- The second strategy is based on a uniform noise model of the forecast error. It always seeks to maintain an energy state close to 0.5. This is due to the overem-

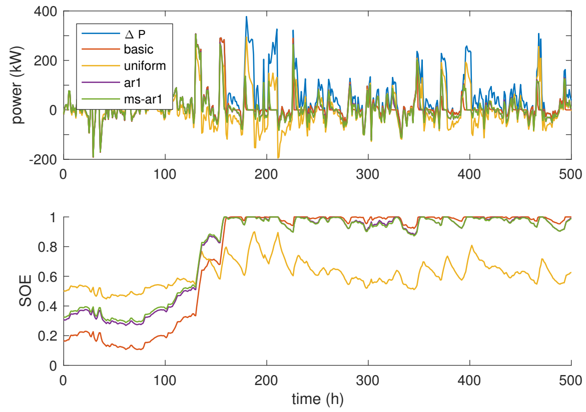


Fig. 8: Top : time trajectories of the forecast error and associated stored power according to various strategies. Bottom : corresponding state of energy of the storage device

TABLE V: Comparison of the total costs obtained by the different management strategies, expressed in primary energy per day

error model	myopic	uniform	AR(1)
$\mathrm{kWh_p}/day$	32.6	20.4	12.5
error model	AR(2)	MS-AR(1)	MS-AR(2)
$\mathrm{kWh_p}/day$	12.4	12.2	12.1

phasis this strategy places on extreme errors. The compensation for current errors is therefore downgraded because of this exaggerated anticipation. It can thus be seen that the stored power deviates significantly from the forecast error to unload the storage.

- Autoregressive models – with or without hidden state – exhibit an intermediate behavior that allows them to largely compensate for forecast errors while controlling changes in the energy state of the storage. The differences between models with and without hidden state are very fine. They are notable when total costs are calculated over the course of a year.

The table V summarizes the total cost evolution according to the different optimal storage strategies. This total cost corresponds to the problem 5 computed throughout the forecast error series. These results first show the reduction of the total cost with each refinement of the model used to describe forecast errors. This highlights that the storage strategy can be significantly improved when it is based on a relevant model. However, it should be noted that the most significant improvement in the results appears as soon as an AR(1) model is used. The transition to an AR(2) and then to MS-AR does continue to improve the final result, but in a much more marginal way. The performance of such a simple model can therefore be underlined here.

VI. Conclusion and further works

The scope of this paper was the modeling of wind power forecast errors. Such errors are inherent in any forecast, regardless of the method used to generate it. Therefore a modeling of forecast errors is necessary in any system that relies on a forecast to establish its planning, in order to compensate them in real time. In particular, such a model must support the temporal evolution of errors – dynamic modeling – and the intrinsic uncertainty in the forecast error signal – stochastic modeling.

The temporal autocorrelation of wind generation forecast errors was first pointed out. Moreover, the succession of regimes where errors are of small or large amplitude has been observed. This motivated the identification of a Markov Switching AutoRegressive model for forecast errors. The coefficients of such a model depend on a hidden state that is driven by a Markov chain. The relevance of this model was initially shown on the basis of several statistical scores.

In a second part, this MS-AR model was used to solve an optimal storage management problem within the context of an elementary and representative problem : the management of a battery associated with a wind power plant subject to a production commitment. As the storage must mitigate the forecast errors as best it can, the forecast error modeling is integrated into the resolution using stochastic dynamic programming algorithm. Several error models were also compared to the MS-AR models : autoregressive, uniform and myopic. The contribution of a good error modeling was highlighted, the management strategy based on an MS-AR model having improved the overall performance.

As an outlook to this study, significant improvements remain open in the MS-AR modeling of forecast errors. For instance using different models for different periods of the year seems a natural way to deal with the seasonality of wind energy forecast errors. However, this seasonality could also be integrated into a single model using non-homogeneous transition probabilities. That is, the probabilities of switching from one state to another would not be constant but would depend on an observable variable, such as the calendar date for example. It would then be possible to create states that only appear at specific seasons of the year.

Furthermore taking spatial correlations into account in forecast errors would be a particularly interesting development for electricity network management. Indeed, geographical effects are crucial to cope with the limitations of the power grid within which many wind farms are connected. The grid operator must then ensure that a forecast error on the production injected at one location can only be compensated by another remote equipment within the operating limits of the network. These spatial effects are starting to be taken into account when predicting wind conditions [24], [7].

Finally, the study presented here on the optimal storage strategy should be extended by a study of the optimal sizing of the battery. Sizing and management strategy are indeed mutually dependent : in similar situations, the optimal decision cannot be the same for two batteries having different capacities. The present paper can therefore be a prerequisite for the introduction of forecast error models within a co-optimization approach over life cycle of the design and management, integrating the effects of aging of the storage device.

REFERENCES

[1] Actual and forecast wind energy feed-in - TenneT (Available at https ://www.tennet.eu/electricity-market/transparency-pages/transparency-germany/network-figures/actual-and-forecast-wind-energy-feed-in/).

[2] BPA : Balancing Authority Load & Total Wind Generation. Available at https ://transmission.bpa.gov/Business/Operations/Wind/.

[3] Pierre Ailliot and Valérie Monbet. Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*, 30 :92–101, April 2012.

[4] Dimitri Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena scientific Belmont, MA, 1995.

[5] H. Bludszuweit, J.A. Dominguez-Navarro, and A. Llombart. Statistical Analysis of Wind Power Forecast Error. *IEEE Transactions on Power Systems*, 23(3) :983–991, August 2008.

[6] Barbara G Brown, Richard W Katz, and Allan H Murphy. Time series models to simulate and forecast wind speed and wind power. *Journal of climate and applied meteorology*, 23(8) :1184–1195, 1984.

[7] Ahmed Aziz Ezzat, Mikyoung Jun, and Yu Ding. Spatio-Temporal Asymmetry of Local Wind Fields and Its Impact on Short-Term Wind Forecasting. *IEEE Transactions on Sustainable Energy*, 9(3) :1437–1447, July 2018.

[8] A. Fabbri, T. GomezSanRoman, J. RivierAbbad, and V.H. Mendez-Quezada. Assessment of the Cost Associated With Wind Generation Prediction Errors in a Liberalized Electricity Market. *IEEE Transactions on Power Systems*, 20(3) :1440–1446, August 2005.

[9] Nicolas Gast, Jean-Yves Le Boudec, Alexandre Proutière, and Dan-Cristian Tomozei. Impact of storage on the efficiency and prices in real-time electricity markets. page 15. ACM Press, 2013.

[10] Pierre Haessig, Bernard Multon, Hamid Ben Ahmed, Stéphane Lascaud, and Pascal Bondon. Energy storage sizing for wind power : impact of the autocorrelation of day-ahead forecast errors : Energy storage sizing for wind power. *Wind Energy*, pages n/a–n/a, October 2013.

[11] James D. Hamilton. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2) :357, March 1989.

[12] Li Han, Carlos E Romero, Xuesong Wang, and Liping Shi. Economic dispatch considering the wind power forecast error. *IET Generation, Transmission & Distribution*, 12(12) :2861–2870, July 2018.

[13] Ian T. Jolliffe and David B. Stephenson, editors. *Forecast Verification : A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons, Ltd, Chichester, UK, December 2011.

[14] Jaesung Jung and Robert P. Broadwater. Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews*, 31 :762–777, March 2014.

[15] Georges Kariniotakis, G. Stavrakakis, and E Nogaret. Wind power forecasting using advanced neural networks models. *IEEE Transactions on Energy Conversion*, 11(4) :762 – 767, 1996.

[16] Matthias Lange. On the Uncertainty of Wind Power Predictions—Analysis of the Forecast Accuracy and Statistical Distribution of Errors. *Journal of Solar Energy Engineering*, 127(2) :177, 2005.

[17] J.A. Peças Lopes, N. Hatziargyriou, J. Mutale, P. Djapic, and N. Jenkins. Integrating distributed generation into electric power systems : A review of drivers, challenges and opportunities. *Electric Power Systems Research*, 77(9) :1189–1203, July 2007.

[18] 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*, 96(12) :2348–2362, December 2008.

[19] Henrik Madsen, Pierre Pinson, Georges Kariniotakis, Henrik Aa Nielsen, and Torben Skov Nielsen. A Protocol for Standardazing the performance evaluation of short term wind power prediction models. page 18.

[20] Phillip E. Mello, Ning Lu, and Yuri Makarov. An optimized autoregressive forecast error generator for wind and load uncertainty study : An optimized autoregressive forecast error generator. *Wind Energy*, 14(8) :967–976, November 2011.

[21] Pierre Pinson, Christophe Chevallier, and George N. Kariniotakis. Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power. *IEEE Transactions on Power Systems*, 22(3) :1148–1156, August 2007.

[22] Pierre Pinson and Henrik Madsen. Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models. *Journal of Forecasting*, 31(4) :281–313, July 2012.

[23] Pierre Pinson, Henrik Madsen, Henrik Aa. Nielsen, George Papaefthymiou, and Bernd Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1) :51–62, January 2009.

[24] Teng Qijun, Wang Chengfu, Liang Jun, and Liang Zhengtang. Research on modeling spatiotemporal correlation of wind power forecast error on multiple wind farms based on Copula theory. pages 447–450. IEEE, September 2017.

[25] Ye 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*, 50 :82–91, October 2015.

[26] 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*, 139(3) :976–991, March 2011.

[27] Zhichao Shi, Hao Liang, and Venkata Dinavahi. Direct Interval Forecast of Uncertain Wind Power Based on Recurrent Neural Networks. *IEEE Transactions on Sustainable Energy*, 9(3) :1177–1187, July 2018.

[28] Goran Strbac. Demand side management : Benefits and challenges. *Energy Policy*, 36(12) :4419–4426, December 2008.

[29] Lennart Söder. Simulation of Wind Speed Forecast Errors for Operation Planning of Multi-Area Power Systems. *8th International Conference on Probabilistic Methods Applied to Power Systems*, page 7, 2004.

[30] Falko Ueckerdt, Lion Hirth, Gunnar Luderer, and Ottmar Edenhofer. System LCOE : What are the costs of variable renewables ? *Energy*, 63 :61–75, December 2013.

[31] Chengfu Wang, Zhengtang Liang, Jun Liang, Qijun Teng, Xiaoming Dong, and Zhaoqing 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*, 98 :373–381, June 2018.

[32] Fengyu Wang and Kory W. Hedman. Dynamic Reserve Zones for Day-Ahead Unit Commitment With Renewable Resources. *IEEE Transactions on Power Systems*, 30(2) :612–620, March 2015.

[33] 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*, 39(7) :3939–3944, July 2011.

[34] Megan Yoder, Amanda S. Hering, William C. Navidi, and Kristin Larson. Short-term forecasting of categorical changes in wind power with Markov chain models : Forecasting categorical changes in wind power. *Wind Energy*, pages n/a–n/a, June 2013.

[35] Haoran Zhao, Qiuwei Wu, Shuju Hu, Honghua Xu, and Claus Nygaard Rasmussen. Review of energy storage system for wind power integration support. *Applied Energy*, 137 :545–553, January 2015.

[36] Yu Zheng, Junhua Zhao, Yue Song, Fengji Luo, Ke Meng, Jing Qiu, and David John Hill. Optimal Operation of Battery Energy Storage System Considering Distribution System Uncertainty. *IEEE Transactions on Sustainable Energy*, 9(3) :1051–1060, July 2018.

APPENDIX

IDENTIFICATION OF THE MS-AR MODEL ON ANOTHER TIME SERIES

As it has been emphasized in section II, the parameters of both the AR and MS-AR processes are highly dependent on the considered time series. To highlight this, table VI shows the parameters obtained on another time series of prediction errors provided by Tennet [1].

One can notice by looking at the variances σ that the regimes seems again to match with weather type characterized by their predictability. Moreover the transition matrix

TABLE VI: Fitted parameters of the MS(3)-AR(2) on another time series provided by Tennet

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	$2e^{-3}$

show that the hidden states are persistent with respectively a mean sojourn time of 14 hours 16 minutes, 12 hours 30 minutes and 3 hours 42 minutes. However the autoregressive coefficients are nothing alike. This means that the parameters shown in this paper are not meant to be transposed as such to another time series. Regardless of the fact that one needs to fit the model himself to use it for its own purposes, it can be seen as a benefit. Indeed, this allow the model to take into account specificities of the considered time series.