A Two-Level Markov Model for Replicating Cross-Time Distributions for Simulations of Renewables in Power Systems

Joe Durante, *Member, IEEE*, Raj Patel, and Warren B. Powell

***Abstract*—A robust energy system with high penetration of renewables must account the possibility that energy from wind or solar falls below forecasts for a period of time. Classical statistical models such as ARIMA and GARCH can do a good job of modeling forecast error and autocorrelation, but we have found they do a poor job of capturing the intertemporal behavior that can produce outages when energy from wind or solar falls below the forecast for a long period. We extend prior work in statistical modeling by focusing on capturing the distribution of the amount time that actual power output is over or under forecast. We propose a nonparametric, two-level Markov model that accurately reproduces not only the error distribution at a point in time, but the length of time that the actual power output is above or below the forecast, known as the crossing time. The method is shown to accurately reproduce cross-time distributions for forecasts of wind power. It is also extended to a very different type of stochastic process, locational marginal prices (LMPs), where we are also able to capture the relationship with forecasted temperature.**

I. Introduction

WHEN developing policies for controlling a system under uncertainty, it is essential they perform across a realistic population of scenarios. In the context of power system control, uncertainty in power generation from renewable power sources provokes a need for simulation methods which accurately replicate characteristics of a renewable source's power output time series. In general, the power output series tends to follow the forecasted output, plus or minus some error; however the error distribution is non-stationary and thus a more sophisticated approach is required to model the time series. To account for the intertemporal correlation of these errors, we focus on directly modeling the series of forecast errors. Various time series models are able to produce sample paths that replicate certain characteristics of observed forecast error series, such as the autocorrelation, partial autocorrelation, and the distribution of errors. However, one characteristic that is often overlooked, yet significant, especially in power system applications, is the distribution of cross-times - consecutive periods of time for which the actual power produced is above or below the forecast. A model has captured key features of the time series if both cross-time distributions (up-crossings and down-crossings) and the error distribution match observed distributions. Not only will sample paths properly reflect the renewable source's tendencies to produce more or less power than expected, but the periods of time for which the source is outperforming or underperforming expectations are carefully modeled as well, implying the model accurately approximates surpluses or deficits of energy (areas between the actual output and forecast). Models that do not capture these behaviors are at risk of producing simulations which do not reflect the intertemporal behaviors of renewable power outputs that may lead to power outages, such as tendencies to fall below forecasts for long periods of time. Other methods, such as in [1], model intertemporal correlation of forecast errors but do not directly model cross-time behavior.

In this paper, a two-level Markov model is constructed with an emphasis on its ability to simulate forecast error time series with distributions - up-crossing time, down-crossing time, and error - that match observed data. The model is nonparametric and assumes no prior knowledge about these distributions. Instead, the model utilizes empirical distributions for sample path generation so that it is effective in a variety of cases, regardless of the form of the empirical distributions. Though intended to model the series of forecast errors in wind power generation, the model can be extended to simulate other stochastic processes with different properties. For example, the model is also used to generate simulations of locational marginal prices (LMPs), which exhibit heavily skewed price distributions and are highly correlated with temperature.

II. Time Series Models

Let Y_t be a point in the realized path at time t and let $\hat{y}_t$ be the forecast at that time. Define $W_t = Y_t - \hat{y}_t$ to be the error between the sample path and the forecast at time t. A common approach is to model the error time series as an Autoregressive Integrated Moving Average process of order p,d,q (an ARIMA(p,d,q) process). As an initial modeling step, differencing can be used to form a stationary time series out of a non-stationary one. A first order differencing forms the series $W'_t = W_t - W_{t-1}$, while a second order differencing forms the series $W''_t = W'_t - W'_{t-1}$. This pattern is repeated up until order d. Consider $\{W_t\}$ the series we are working with, differenced or not. It can be modeled as an ARMA(p,q) process as follows:

$$W_t = c + \sum_{i=1}^{p} \alpha_i W_{t-i} + E_t + \sum_{j=1}^{q} \beta_j E_{t-j},$$

J. Durante: Princeton University, Department of Electrical Engineering, email: jdurante@princeton.edu; R. Patel and W.B. Powell: Princeton University, Department of Operations Research and Financial Engineering, emails: rmpatel@princeton.edu, powell@princeton.edu

where $E_t \sim \mathcal{N}(0, \sigma^2)$. Determining the order - the p and q - of the ARMA model can be done through comparisons to the theoretical autocorrelation function (ACF) and partial autocorrelation function (PACF) for specific p and q combinations [2]. The autocorrelation at lag-ℓ is defined as $Corr(W_t, W_{t+\ell})$, while the lag-ℓ partial autocorellation is the correlation between W_t and $W_{t+\ell}$ with the linear dependence on intermediate random variables $W_{t+1}, W_{t+2}, ...W_{t+\ell-1}$ removed. As an example of how theoretical distributions can guide model choice, the theoretical ACF of a stationary ARMA(p,0) process will slowly decay towards 0 for increasing lags, while only terms up to lag-p in the theoretical PACF may be nonzero. If the ACF of the observed series exhibits this behavior, while coefficients of terms past lag-p in the PACF are not statistically significant, then an ARMA(p,0) model would be an appropriate model choice [2]. Once the order of the model is chosen, the parameters $c, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, \sigma$ are chosen to maximize the likelihood of the data.

Note in the ARMA model, the volatility σ is constant throughout the process as we assume we have a stationary time series. However, non-stationarity is common, and volatility may tend to change over time. To allow for this, a Generalized Autoregressive Conditional Heteroskedastic (GARCH) process of order p', q' described in [3], an extension of the ARCH(q') model presented in [4], can be used to model the conditional volatility at time t as follows:

$$\sigma_t^2 = k + \sum_{n=1}^{p'} \phi_n E_{t-n}^2 + \sum_{m=1}^{q'} \theta_m \sigma_{t-m}^2.$$

This can then be merged with the ARMA model by letting σ_t replace σ to produce a hybrid ARMA-GARCH model. Again, parameters are chosen to maximize likelihood.

In wind power applications, allowing volatility to evolve over time based on recent volatility history would seem to make sense. To see why, consider for example a time period in which there are violent storms with intermittent calmer periods. In a month-ahead forecast, the power generated by this weather pattern may be tough to predict. Thus, during these times forecast errors will likely be much more volatile than during calm periods. Having a general model to handle these volatility changes without needing additional explanatory variables can be useful, and indeed ARMA-GARCH hybrid approaches have been used in wind speed forecasting applications such as in [5].

If instead, we want to allow more parameters of our non-stationary model to evolve over time we can use a state space method [6]. Another approach is to assume there exists distinct changes in regime over the course of the time series that cause the parameters to change suddenly; this idea is described in [7]. Using this method, one could use a discrete state Markov chain to model changes in regime, while fitting a unique ARIMA, GARCH, or other type of model to the data occurring in each regime.

III. Two-Level Markov Model

For the remainder of the paper let F_X be the empirical distribution of the random variable X.

Our error model is a two-level Markov model with two state variables that evolve on different time scales. The complete state variable, which is responsible for carrying a minimally dimensioned function of history up to time t necessary to determine the distribution of W_{t+1} [8], evolves at every point in time, while the underlying crossing state of the system, which contains information about whether $W_t > 0$ or $W_t < 0$ in addition to the duration of time the sample path is above or below the forecast, only makes transitions at crossing points at which errors switch signs. If S_t^C is the crossing state of the system, a second level Markov model conditioned on S_t^C dictates the distribution of the error W_{t+1} given W_t, thus the crossing state is a part of the complete state variable. However, it evolves according to the separate first level Markov chain giving $P(S_{t+1}^C | S_t^C)$.

A. First Level Crossing State Markov Model

In order to fully describe the crossing state variable S_t^C, we must first define a crossing time. Let the set of all indices such that errors crossed over from the negative to positive regime be:

$$\mathcal{C}^U = \{t | W_{t-1} \leq 0 \wedge W_t > 0\}.$$

Likewise, let the set of all indices such that errors crossed over from the positive to negative regime be:

$$\mathcal{C}^D = \{t | W_{t-1} \geq 0 \wedge W_t < 0\}.$$

An up-crossing time of duration d starting at time $t' \in \mathcal{C}^{\mathcal{U}}$ is then defined as:

$$T_{t'}^U = d \text{ if } \begin{cases} W_{t'+k} > 0 & \forall k \in \{0, 1, ..., d-1\} \\ W_{t'+d} \leq 0 \end{cases}.$$

Similarly, a down-crossing time of duration d starting at time $t' \in \mathcal{C}^D$ is defined as:

$$T_{t'}^D = d \text{ if } \begin{cases} W_{t'+k} < 0 & \forall k \in \{0, 1, ..., d-1\} \\ W_{t'+d} \geq 0 \end{cases}.$$

Examples of points in time belonging to $\mathcal{C}^U$ and $\mathcal{C}^D$, as well as an up and down crossing time are shown, approximately, in Figure 1.

For both the up-crossing and down-crossing times, there exists distributions F_{T^U} and F_{T^D} respectively. Up-crossing time distributions are quantized by partitioning into Q bins, splitting at the quantile points such that q_i^U is the i^{th} quantile for $i \in \{1, 2, ..., Q\}$ (where q_Q^U represents the maximum value in the distribution). Note, this implies the bins will likely not be equally sized. If we also let $q_0^U = 0$, an up-crossing time $T_{t'}^U$ would then belong to bin b_i^U if $q_{i-1}^U \geq T_{t'}^U < q_i^U$, with the exception being that the maximum value in the distribution belongs to bin b_Q^U. If, for example, Q is chosen to be 3, the lower third of up-crossing times belonging to bin b_1^U could be interpreted as "short runs above the forecast," while the middle third belonging to bin b_2^U would be interpreted as "medium length runs above the forecast." The same procedure is carried out for down-crossing times.

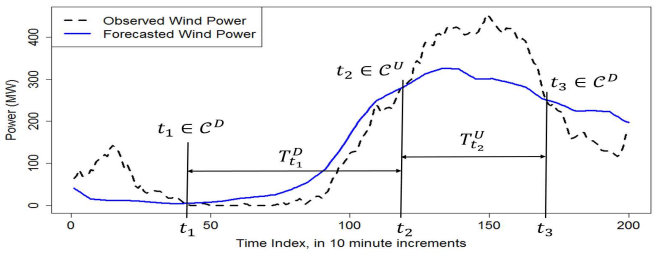

Fig. 1. Examples of points in time belonging to $\mathcal{C}^U$ and $\mathcal{C}^D$ and both an up and down crossing time. Note the crossing times portrayed are approximate, and extend one time index too long as shown in the figure since they are defined to end before errors switch signs.

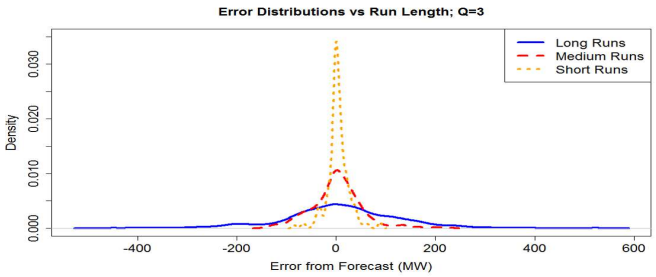

Fig. 2. Error distributions conditioned on B^C_t, the cross-time quantile, with $Q = 3$. Both positive and negative errors for equal values of B^C_t are combined to form the distributions. Long cross-time states tend to produce errors that deviate far from zero, while errors occurring during short crossing states tend to be more concentrated around zero. Therefore, conditioning error generation on run length is appropriate in this case.

Our crossing state variable S^C_t is then defined as the pair of variables describing whether or not the path is above the forecast and what cross-time duration bin the system is currently in. We can express this as the pair $S^C_t \equiv (X^C_t, B^C_t)$, where:

$$X^C_t = \begin{cases} U \text{ if } W_t \geq 0 \\ D \text{ if } W_t < 0 \end{cases},$$

and, after identifying $\underset{t' \in \mathcal{C}^{X^C_t}}{\operatorname{argmax}}(t' - t)$ such that $t' - t \leq 0$ (finding the last point at which the errors switched signs),

$$B^C_t = i \text{ if } T_{t'} \in b^{X^C_t}_i .$$

When training is finished, there exists a distribution of cross-times $F_{T^{S^C_t}}$ for each crossing state; this will be important for simulation.

Note if $t+1 \notin (\mathcal{C}^U \cup \mathcal{C}^D)$, $P(S^C_{t+1} = S^C_t | S^C_t) = 1$ as the crossing state remains constant until errors switch signs. If, however, $t+1 \in (\mathcal{C}^U \cup \mathcal{C}^D)$, the crossing state will evolve according to a nontrivial distribution $F_{S^C_{t+1}|S^C_t}$, which is computed from data as follows: considering only points in time such that $t+1 \in (\mathcal{C}^U \cup \mathcal{C}^D)$, let $n(S^C_{t+1} = j | S^C_t = i)$ be the count of the transitions from state i to state j occurring for each pair of crossing states (i, j) and let $n(S^C_t = i)$ be the number of times $S^C_t = i$ for each crossing state i; then the empirical transition probability from crossing state i to j is given by:

$$P(S^C_{t+1} = j | S^C_t = i) = \frac{n(S^C_{t+1} = j | S^C_t = i)}{n(S^C_t = i)}.$$

B. Second Level Markov Model for Generation of Sample Points

The error distributions $F_{W_t|S^C_t}$ are not identical across all possible crossing states S^C_t; in fact they are likely to be quite different, such as in the case where the error distribution is asymmetric. Furthermore, error distributions are likely to vary with run length as well. For example, consider the distribution of errors from forecast in a short run compared to a long run (of either type). It is often more likely that a short run narrowly rises above/drops below the forecast before switching crossing states, leading to a larger concentration of errors near 0. On the other hand, a long run may deviate farther from the forecast during its duration and, furthermore, part of the reason long runs occur is because the forecast turned out to be quite inaccurate for some period of time. This behavior is seen in Figure 2, which shows error densities for each type of run length for the data set analyzed in section IV-A with $Q = 3$. Thus, to better capture the behavior of the error process, the second level Markov model is conditioned on the crossing state S^C_t.

In addition to errors being crossing state dependent, they are dependent on error history as well; a first order Markov chain is used to model this behavior. As errors may be continuous or take on many discrete values, we must first quantize the distribution to form a manageable number of error states. Similar to how the cross-time distributions are partitioned, each error distribution $F_{W_t|S^C_t}$ is split into R bins: $b^{S^C_t}_1, b^{S^C_t}_2, ..., b^{S^C_t}_R$, splitting at the quantile points such that $q^{S^C_t}_i$ is the i^{th} quantile for $i \in \{1, 2, ..., R\}$ (where $q^{S^C_t}_R$ represents the maximum value in the distribution). Letting $q^{S^C_t}_0$ be the minimum value in the distribution, $W_t \in b^{S^C_t}_i$ if $q^{S^C_t}_{i-1} \leq W_t < q^{S^C_t}_i$ (with the exception that the maximum value in the distribution belongs to bin $b^{S^C_t}_R$) and the crossing state at time t is S^C_t. Then, given $W_t \in b^{S^C_t}_i$, we have a conditional distribution for the error at time $t+1$, $F_{W_{t+1}|S^C_t, W_t \in b^{S^C_t}_i}$. Therefore, the complete state variable is defined as $S_t \equiv (S^C_t, B^W_t) \equiv (X^C_t, B^C_t, B^W_t)$ where:

$$B^W_t = i \text{ if } W_t \in b^{S^C_t}_i .$$

The dependence of W_{t+1} on B^W_t, the state of the current error, is illustrated in Figure 3 in which conditional distributions for W_{t+1} are plotted for a fixed crossing state, but varying states B^W_t. In the example, the data set is the one used in section IV-A, $Q = 3$, and $R = 5$.

There are a few practicalities to consider when using the model to generate a sample path. To produce a sample path of errors from forecast of length L, use Algorithm 1 written in pseudocode in which vectors are indexed starting at 0.

C. Extending the Model: Addition of Explanatory Variables

If there exists other variables that influence the behavior of the time series, they can be integrated into the model as a

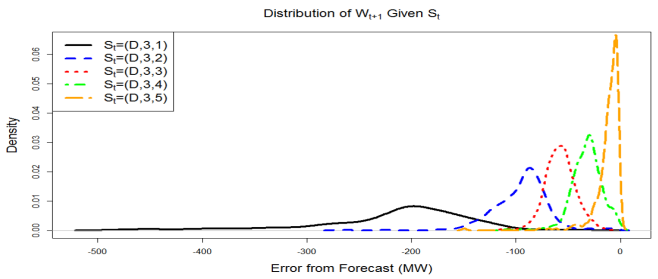

Fig. 3. Example of conditional distributions for W_{t+1} given a fixed crossing state, $S_t^C = (D, 3)$, but varying which bin, $b_i^{S_t^C}$, that W_t belongs to. It is clear from the figure that the magnitude of the next error is largely dependent on the magnitude of the current error.

Algorithm 1 Simulation Algorithm

Choose Q and R.
From observed data, form and store empirical distributions $F_{S_{t+1}^C|S_t^C}$,$F_{T^{S_t^C}}$, $F_{W_t|S_t^C}$, $F_{W_{t+1}|S_t}$, and $F_{S_t^C}$
Initialize $Sim = [\,]$
Initialize $t \leftarrow 0$
Sample an initial crossing state S_0^C from $F_{S_t^C}$.
while $t < L - 1$ **do**
  Sample a run length (cross-time) r from $F_{T^{S_t^C}}$
  $RunVec = zeros(r)$
  for $t' = 0, 1, 2, ..., r - 1$ **do**
    if $t' = 0$ **then**
      $RunVec[t']$ sampled from $F_{W_t|S_t^C}$
    else
      Determine B_t^W from the error $RunVec[t' - 1]$ and S_t^C; B_t^W combined with S_t^C forms S_t
      $RunVec[t']$ sampled from $F_{W_{t+1}|S_t}$
    end if
  end for
  $Sim \leftarrow concat(Sim, RunVec)$
  $t \leftarrow t + r - 1$
  Change crossing states by sampling from $F_{S_{t+1}^C|S_t^C}$
  $t \leftarrow t + 1$
end while
return $Sim[0 : L - 1]$

conditioning variable in either Markov model as applicable. To express this mathematically, if there are N explanatory variables $V_{1,t}^C, V_{2,t}^C, ..., V_{N,t}^C$ that affect the crossing state behavior, we can write our crossing state Markov model as:

$$P(S_{t+1}^C | S_t^C, V_{1,t}^C, V_{2,t}^C, ..., V_{N,t}^C).$$

Additionally, if there are M explanatory variables $V_{1,t}, V_{2,t}, ..., V_{M,t}$ that affect the behavior of the individual errors, we can write this Markov model as:

$$P(W_{t+1} | S_t^C, W_t \in b_i^{S_t^C}, V_{1,t}, V_{2,t}, ..., V_{M,t}).$$

Note it is possible, and often likely, that the crossing state explanatory variables will be the same as the ones affecting the error model.

One should be aware that with the addition of an explanatory variable which can take on k values, the amount of possible states for the Markov model increases by a factor of k. Thus if the amount of training data is limited, one should restrict the addition of explanatory variables to those with the most impact. Furthermore, to keep these low dimensional, combine possible values which the variable may take on into groups to form aggregated states (as was done with the crossing state and the state of the current error).

D. Relation to Other Markov Models

The concept of using a two-level Markov model to generate time series is not new; for example the model used in [9] is a two-level Markov model used to simulate rain events. A first level chain transitions between "rain" and "inter-rain" states by remaining in each state for a duration sampled from the appropriate empirical distribution for each state and then switching. Then, given the state is currently "rain," the second level model gives the distribution of the rainfall rate at time $t+1$ given the rate at time t. This is a similar model to the one used here and fits under the category of explicit or variable duration Markov Models [10]–[12]. To use that framework for our model, we would simply reformulate the crossing state Markov chain such that our state space is $S_t^C \in \{U, D\}$ with $P(U|D) = 1$ and $P(D|U) = 1$ and duration for S_t^C originating from the appropriate cross-time distribution F_{T^U} or F_{T^D}. This method accurately replicates empirical cross-time distributions as well.

Instead, our first level crossing state Markov chain incorporates duration into the state variable, splitting both the above and below forecast states into sub-states based on duration ranges, and in doing so, increasing the number of states by a factor of Q. As a result, each state has a minimum and maximum duration, and thus this structure is related to the bounded state duration Markov model described in [13]. Using this structure captures intertemporal characteristics of the data such as the likelihood of a short run below the forecast following a short run above the forecast may be significantly higher than that of a long run below the forecast following a short run above the forecast. Thus, crossing state transitions are duration dependent, but not as in the classic model presented in [14] where transition probabilities between states are direct functions of duration; rather the states themselves are partially defined by duration and transition probabilities between states are constant.

The model also falls under the larger category of regime switching Markov processes in which the parameters for a general second level model are determined by a first level Markov chain. For example, one model that is also tested in this paper is the model presented in [7] in which the parameters for an ARIMA process are determined by the underlying state which evolves according to a Markov chain. However, the two-level Markov model is more closely related to the switching Markov models described in [15] as the second level model is also a discrete state Markov chain.

IV. Results

The two-level Markov model is first used to produce simulations of wind farm power generation and results are

compared to more common time series modeling techniques. Then, testing the method in a different setting with a very different stochastic process, the model is used to generate simulations of LMP paths.

A. Wind Farm Power Generation

We will be modeling the wind power generation from a single wind farm in the United States' Great Plains region in April 2013. Observed data points were collected over the course of the month in ten minute intervals. Along with the observed time series, a month-ahead forecast of power output is available at each ten minute increment as well. Due to space constraints, we are only able to display results from one test case. However, similar results were seen in tests of other wind farms during different months of the year.

1) Models Tested: From Figure 7, we see the magnitudes of only the first three lags in the observed partial autocorrelation function are statistically significant, and the autocorrelation function trends slowly towards 0; thus a properly fitted ARIMA(3,0,0) model is an appropriate model for the error sequence [2], referred to as an AR(3) model for the remainder of the paper, as no differencing is used and only the autoregressive terms are retained. Also tested is an AR(3)-GARCH(1,1) model as this model can account for any changes in forecast error volatility. In addition, we consider a regime switching autoregressive model in which an AR(3) model is fit to each series of errors occurring in each crossing state S_t^C. The crossing state Markov model is then used as in the two-level Markov model to switch regimes, and state durations are sampled in the same manner as well. Based on S_t^C, the appropriate AR(3) model is used to simulate for the duration of the run before switching to a new crossing state. To ensure the model does not produce sample points inconsistent with the crossing state - an inconsistent point would be, for example, if the model produces a negative error while we are in an above-the-forecast state - if an inconsistency does occur, the sample is rejected and re-sampled. We will call this the CSAR(3) (crossing state AR(3)) model. Finally, the two-level Markov model is tested with parameters $Q = 3$, $R = 6$, and no explanatory variables.

Each model produces 100 forecast error sample paths. The forecast series is then added to each simulation to form simulations of wind power output. These must be non-negative and are clipped if they fall below 0 MW; they are also capped at the maximum power the wind farm can produce. The resulting distributions are plotted for all simulations produced by a single model combined. In addition, for models whose distributions match observed data best, we examine characteristics of individual sample paths.

2) Comparing Results - Distributions: Figure 4 shows the resulting error distributions for each model versus observed, while Figure 5 displays cross-time distributions. From these, we see that the AR(3)-GARCH(1,1) model does not replicate any distribution well and is not a strong model choice for our purposes. The standard AR(3) model does match the error distribution well, but note without the crossing state variable to guide transitions between error regimes, cross-

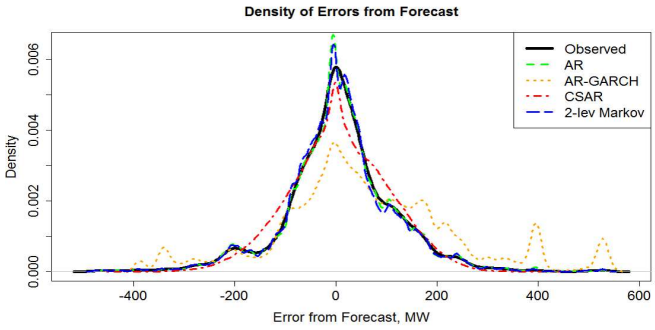

Fig. 4. Error distributions over 100 simulations using each model versus the observed distribution. The standard AR(3) and two-level Markov models fit the observed distribution best, fitting the tails of the distribution as well. Conversely, the AR(3)-GARCH(1,1) model and CSAR(3) models produce errors too frequently in some regions and not often enough in others.

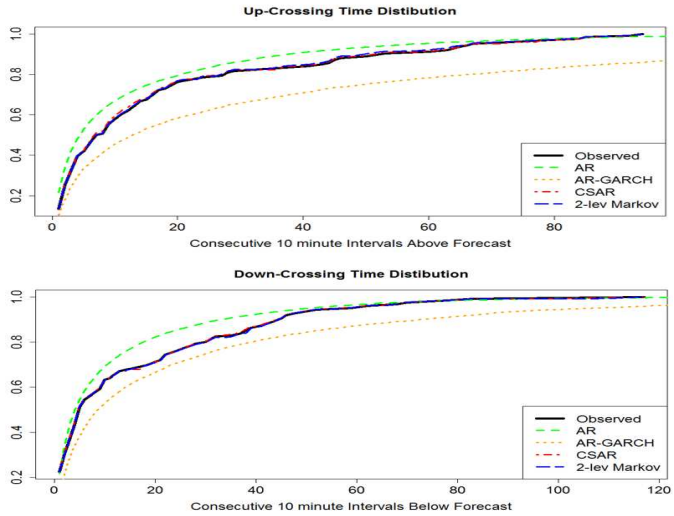

Fig. 5. Comparing Up (top) and Down (bottom) crossing time CDF's for 100 simulations using each model versus the observed distributions, which are given by the solid blue lines. The two models that utilize the crossing state replicate the observed distribution almost exactly, while other time series models either overestimate or underestimate the amount of time spent above and below forecast.

times are generally shorter than in the observed data. This is an undesirable quality of sample paths as, for example, consistently underestimating the amount of time that wind power generation will be lower than expected could lead to power dispatch policies that hold up under simulations, yet fail in real world scenarios. The CSAR(3) model does replicate the cross-time distributions well, but the error distribution is degraded in quality from the standard AR(3) model. Finally, the two-level Markov model reproduces all three distributions well.

Figure 6 examines another measure of model effectiveness: the distributions of areas, above and below, between the paths and the forecast. Reproducing these distributions well would indicate not only are we replicating how long the paths deviate from expected power output, but also by how much during those periods, giving the total error in energy produced during these periods. The models best matching

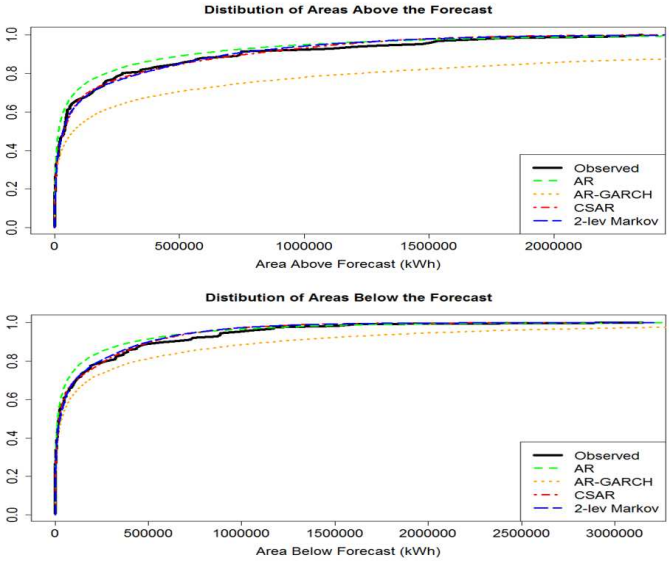

Fig. 6. Comparing the CDF's of areas above forecast (top) and areas below forecast (bottom) in kWh for 100 simulations using each model versus the observed distributions, which are given by the solid lines. The two models that incorporate the crossing state replicate the observed distribution closest.

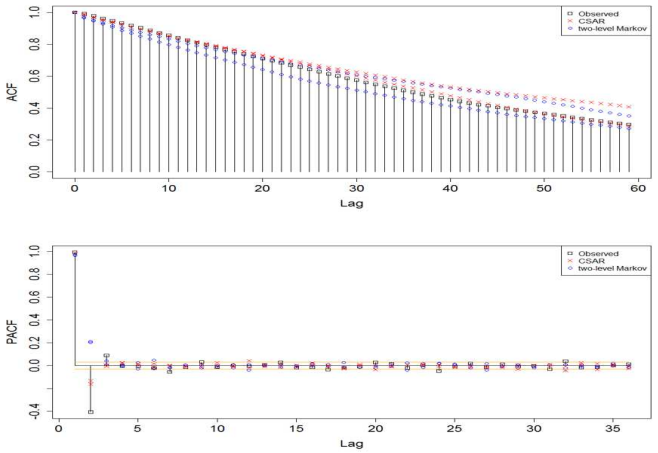

Fig. 7. The autocorrelation (top) and partial autocorrelation (bottom) functions for two sample paths generated from both models utilizing the crossing state vs observed data. Horizontal lines on the PACF plot indicate the threshold for statistical significance. A downside to the two-level Markov model is shown here: individual sample paths seem to behave slightly differently than observed data as the lag-2 term in the PACF from simulated paths of the opposite sign, though the CSAR(3) model doesn't match well either.

both observed distributions incorporate the crossing state into the model. However, none of these models have a technique devoted to controlling the distribution of areas above and below the forecast similar to the crossing state Markov model for cross-time distributions. Thus, the quality of simulated area distributions is more variable than that of simulated cross-time distributions, and in other test cases area distributions do not match as well. This is a problem for future research to address.

3) Comparing Results - Individual Sample Path Properties: Figure 7 shows the autocorrelation function and partial autocorrelation function for two sample paths produced by both

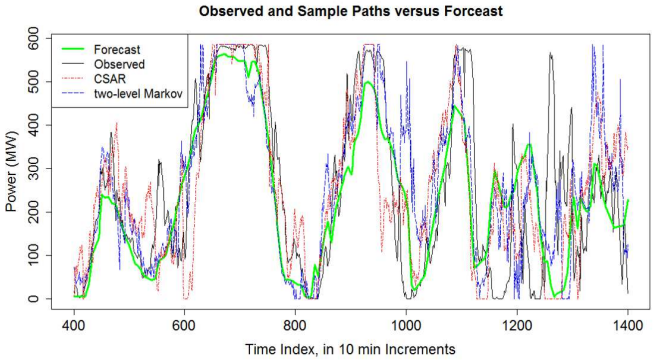

Fig. 8. Simulated sample paths of wind power from both crossing state models vs observed power for the a week during April 2013. Also shown is the power forecast for the time period.

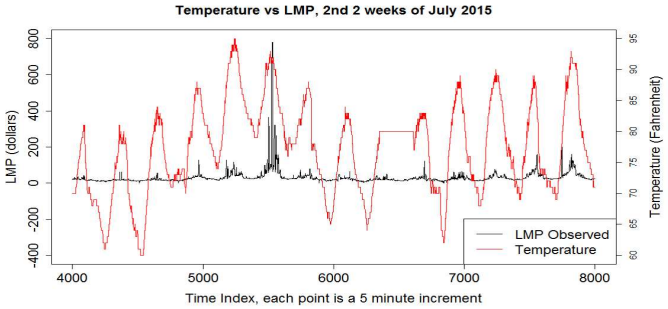

Fig. 9. LMP path and temperature data for the Princeton, NJ, area during 2 weeks in July 2015; this is only a portion of the training data, as all of July and August 2015 are used for training. LMP's spike with daily peaks and exhibit larger spikes when the peaks are higher.

the CSAR(3) and two-level Markov models. Note that while both models produce sample paths with similar autocorrelation functions to the observed path, the lag-2 terms in the PACF's for the sample paths produced by the two-level Markov model are positive rather than negative. This error is not present in the CSAR(3) model, however the lag-2 coefficient is still smaller in magnitude. Thus, when choosing a model there is a trade off between having better error distributions (two-level Markov model) and having sample paths with better matching PACF's (CSAR(3) model). For both models, sample paths do pass a qualitative "eye test" as they seem to behave similarly to observed paths as in Figure 8.

B. Electricity Prices

Electricity prices have been modeled using regime switching models previously, as in [16] and [17] for example. Though, if one has an application in which it is desirable to replicate both cross-time distributions (with respect to some reference) and price distributions, our model for forecast errors can extended to simulate LMP time series as well. As mentioned earlier, this stochastic process is quite different from wind power generation, yet a similar methodology can be used effectively in this setting.

1) The Data and Threshold Level: The time series we consider is the LMP path in the Princeton, New Jersey area

during the months of July and August, 2015. Also available is the observed temperature in the area during these months. A plot of a two week portion of these two series is shown in Figure 9. Notice there exists a seasonal (daily) component to the LMP path; it goes through a daily cycle of low at night to slightly higher during the day. Before attempting to use the model for training and simulation, we remove this seasonality component. This is done by computing averages for all sets of points in the time series exactly 1 day apart, and subtracting the resulting averages from the appropriate points in the series. Following modeling and simulation, the seasonal component is then added back to the simulations. This allows us to focus on modeling tendencies of the data that cannot be easily modeled by a simple periodic function.

Observe from Figure 9 that LMP's often spike during the high points of temperature each day. Additionally, the higher the peak temperature, the higher the spikes and price level in general. If we want our simulated paths to emulate this behavior it would make sense to add temperature as an explanatory variable to the model. While in practice we will not have observed temperature data available for periods in the future, a temperature forecast can be substituted as it is accurate enough to capture the behavior of temperature in the near future.

Note that we are also missing a critical element to the two-level Markov model - the forecast for LMP's. However, a true forecast is not necessary to use the model; only a reference point at each point in time is required, forming a series of reference points. This reference series can be chosen to be a series of interest, such as a threshold over which the policy is to sell energy to the grid. The series should be chosen, if possible, such that the time series crosses the reference series often so that plenty of data is available for training the Markov models below and above the reference series. In the wind power application, the forecast is a reference series which varies with time. For this application we use the mean of the previous month's (June 2015) prices, from which the daily component is also first removed, as a fixed threshold level to serve as our reference series. Thus, if c is the mean of the altered LMP series from the previous month, $\hat{y}_t = c \ \forall t$.

2) Addition of Temperature as an Explanatory Variable to Model: We will condition the second level Markov model for generation of sample points on temperature. As previously mentioned, LMP's spike during temperature peaks, and the magnitude of the spikes is also dependent on the magnitude of the temperature peak. One method of conditioning would be to condition directly on the temperature series. However this will produce unwanted behavior as some of the troughs in the temperature series for hotter days are in the same range as the peak values for cooler days. Thus, when simulating, smaller peaks are liable to appear during temperature troughs as well. In order to circumvent this issue, we first isolate two components of the temperature series - the seasonal component and the trend component (identified by applying a length 50 MA filter to the temperature series). Then, the Markov model is conditioned on both of these variables. As both series take on continuous values, they must be aggregated into bins first. To keep the Markov model low dimensional, we use only 2 bins for the seasonal component and 3 bins for the trend component. The explanatory variables $V_{1,t}$ and $V_{2,t}$ are defined as follows:

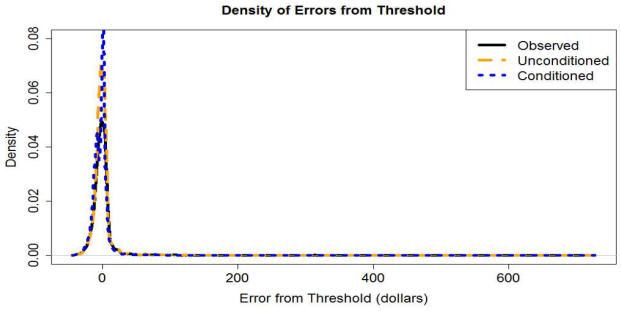

Fig. 10. Error distributions for observed and simulated LMP paths from both the conditioned and unconditioned two-level Markov models. The error distributions from the two models are slightly different, but both capture the heavy positive skew seen in the data.

Let τ_t^S be the seasonal component of the temperature series at time t. Let $\tau_{max}^S = \max_t \tau_t^S$. The variable $V_{1,t}$ may take on two values at time t:

$$V_{1,t} = \begin{cases} 2 \text{ if } \tau_t^S \geq 0.75 \times \tau_{max}^S \\ 1 \text{ if } \tau_t^S < 0.75 \times \tau_{max}^S \end{cases}.$$

This division is intended to separate generation of sample points when the temperature is peaking from all other times of the day (similar behavior can also be obtained by conditioning on time of day instead).

Let τ_t^T be the trend component of the temperature series at time t. Let $\tau_{max}^T = \max_t \tau_t^T$. The variable $V_{2,t}$ may take on three values at time t:

$$V_{2,t} = \begin{cases} 3 \text{ if } \tau_t^T \geq 0.8 \times \tau_{max}^T \\ 2 \text{ if } 0.3 \times \tau_{max}^T \leq \tau_t^T < 0.8 \times \tau_{max}^T \\ 1 \text{ if } \tau_t^T < 0.3 \times \tau_{max}^T \end{cases}.$$

This division is intended to separate generation of sample points between cool, average, and hot days.

Finally, in the interest of being able to form more complete conditional distributions for W_{t+1} in each state from a limited amount of data, the generation of sample points is not conditioned on the run length variable B_t^C (but it remains relevant for crossing state transitions). Thus, our conditioned model will maintain the crossing state Markov chain:

$$P(S_{t+1}^C | S_t^C = (X_t^C, B_t^C)).$$

However, sample points are generated using:

$$P(W_{t+1} | X_t^C, B_t^W, V_{1,t}, V_{2,t}).$$

3) Resulting Distributions and Sample Paths: Using the temperature-conditioned two-level Markov model, we run 100 simulations of the LMP time series. 100 sample paths are generated using the unaltered two-level Markov model as well. The resulting error distributions and cross-time distributions are shown in Figure 10 and Figure 11 respectively.

We see that both models have near identical performance when it comes to replicating these distributions. Figure 10

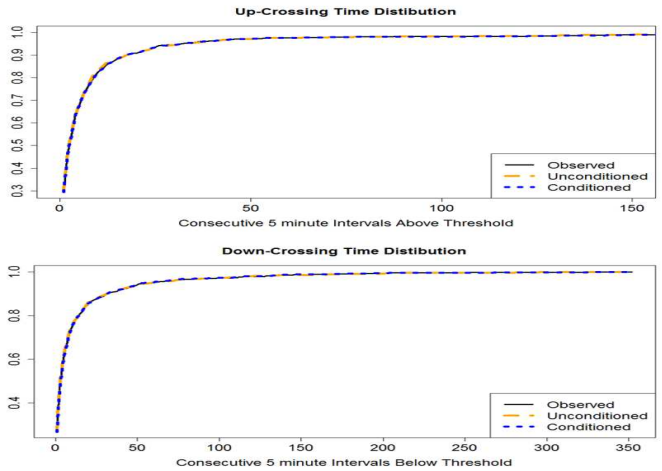

Fig. 11. Up (top) and Down (bottom) crossing time CDF's for observed and simulated LMP paths from both the conditioned and unconditioned two-level Markov models. As the altered model uses conditioning only for sample point generation, the simulated cross-time distributions are nearly identical and both match observed distributions.

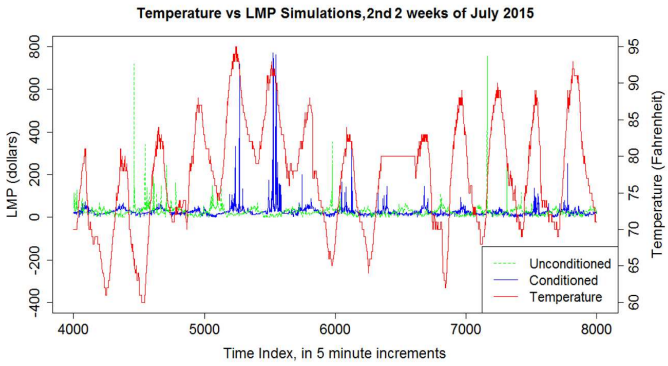

Fig. 12. LMP simulations plotted along with temperature for 2 weeks in July 2015. The simulated series using the conditioned model exhibits spiking behavior during temperature peaks and higher peaks on hotter days, as in the observed data. On the other hand, the unconditioned model's sample paths may exhibit spiking at any time.

highlights another general strength of the two-level Markov model - its ability to replicate a variety of error distributions. Deviations from the threshold are positively skewed, with quite a large tail, due to the spiking behavior of LMP paths. Both models accurately capture this behavior.

The impact of conditioning is displayed in Figure 12, which shows a simulated price path from each model. While spikes are equally likely to occur at any point in the series using the unconditioned model, the conditioned model's sample paths will exhibit spiking almost exclusively when the temperature is at a daily peak, a characteristic seen in the observed LMP series. Larger peaks are more likely to occur when temperature is higher on average as well.

V. Conclusion

The nonparametric two-level Markov model presented is tailored for producing sample paths of wind power generation at individual wind farms. The first level crossing state Markov chain is utilized to keep simulated cross-time distributions consistent with observed distributions, while the second level Markov model is trained to replicate various error distributions effectively. Using this method, areas between forecasted and actual power, and thus the surpluses or deficits of energy produced versus expected, are also replicated well. Therefore, intertemporal behaviors of wind power outputs which may lead to power outages if unaccounted for, such as their tendency to underperform expectations for extended periods of time, are reflected in the simulations. Though developed for wind power applications, the model can be extended to effectively produce simulations of other stochastic processes with different properties, such as locational marginal prices. Through conditioning on additional explanatory variables, we are able capture the dependence of the process on temperature.

Acknowledgment

The research was supported in part by NSF grant CCF-1521675.

References

[1] 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.

[2] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[3] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[4] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007, 1982.

[5] H. Liu, E. Erdem, and J. Shi, "Comprehensive evaluation of ARMA–GARCH (-M) approaches for modeling the mean and volatility of wind speed," *Applied Energy*, vol. 88, no. 3, pp. 724–732, 2011.

[6] C.-J. Kim, C. R. Nelson *et al.*, *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. MIT press Cambridge, MA, 1999, vol. 2.

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

[8] W. B. Powell, "A unified framework for optimization under uncertainty," *Tutorials in Operations Research*, 2016, in review.

[9] C. AIasseur, L. Husson, and F. Pérez-Fontán, "Simulation of rain events time series with markov model," in *Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004. 15th IEEE International Symposium on*, vol. 4. IEEE, 2004, pp. 2801–2805.

[10] S.-Z. Yu, "Hidden semi-Markov models," *Artificial Intelligence*, vol. 174, no. 2, pp. 215–243, 2010.

[11] 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.

[12] J. D. Ferguson, "Variable duration models for speech," in *Proceedings of the Symposium on the Application of HMMs to Text and Speech*, 1980, pp. 143–179.

[13] H.-Y. Gu, C.-Y. Tseng, and L.-S. Lee, "Isolated-utterance speech recognition using hidden Markov models with bounded state durations," *IEEE Transactions on Signal Processing*, vol. 39, no. 8, pp. 1743–1752, 1991.

[14] S. Vaseghi, "State duration modelling in hidden Markov models," *Signal processing*, vol. 41, no. 1, pp. 31–41, 1995.

[15] T. V. Duong, H. H. Bui, D. Q. Phung, and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-markov model," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 838–845.

[16] R. Weron, M. Bierbrauer, and S. Trück, "Modeling electricity prices: jump diffusion and regime switching," *Physica A: Statistical Mechanics and its Applications*, vol. 336, no. 1, pp. 39–48, 2004.

[17] R. Huisman and R. Mahieu, "Regime jumps in electricity prices," *Energy economics*, vol. 25, no. 5, pp. 425–434, 2003.