PO.0103

COLORADO SCHOOL OF MINES
EARTH • ENERGY • ENVIRONMENT

Short-term Forecasting of Categorical Changes in Wind Power with Markov Chain Models

Amanda S. Hering (with M. Yoder, W. Navidi, and K. Larson)

Colorado School of Mines; Department of Applied Mathematics and Statistics

EWEA THE EUROPEAN WIND ENERGY ASSOCIATION

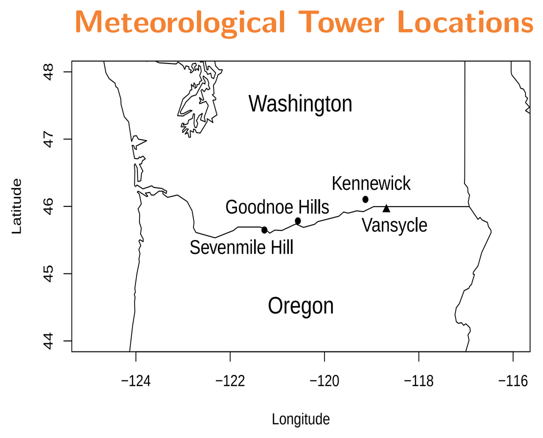
Introduction

- **Objective:**
  - Build a model to forecast a categorical change in wind power (increase, no change, decrease) one hour ahead with an associated measure of uncertainty.
- **Background:**
  - Balancing authorities must maintain
    - equilibrium within a given region (i.e., supply must equal demand)
    - stability such that the system is flexible enough to cover the power shortages or excesses due to unexpected events
  - Low forecast uncertainty ⇒ fewer rolling reserves and lower wind integration costs
  - High forecast uncertainty ⇒ greater rolling reserves and higher wind integration costs
  - No one single type of wind forecast, whether it be of mean wind speed or mean wind power for the upcoming hour or two, is likely to be sufficient.
- **Application:**
  - To answer the following operator question, "Will wind generated power stay the same for the next hour, or will changes in wind power require reallocating the generation mix?"
  - If wind power for a particular wind project will stay the same in the next hour with high probability, then the operator can quickly turn his attention to other decisions.
- **Prior Markov Chain Work in Wind:**
  - They have been popular for simulating long time series of wind speeds for synthetic system experiments.
  - They have been used to switch between regimes for wind power forecasting.
  - No one, to our knowledge, has used Markov models directly for wind forecasting, and no one has attempted to forecast changes in wind power.

Data

- Ten-minute observations collected over 2 years (2005 and 2006) at 4 met towers.
- Hourly averages of speed and directions are computed.
- A small amount of data was imputed, less than 2% for most variables/locations.
- The type of turbine originally installed at the Stateline Wind Project was Vestas V47-660, a 660 kW turbine whose hub height is 50 m above ground level.
- Wind speed is scaled to 50 meters agl using the power law.

Meteorological Tower Locations



Site Information of the Meteorological Towers

Tower	Latitude	Longitude	Elevation	Anemometer Height
Vansycle, OR	45°57′ N	118°41′ W	543 m	31 m
Kennewick, WA	46°06′ N	119°08′ W	671 m	26 m
Goodnoe Hills, WA	45°48′ N	120°34′ W	774 m	59 m
Sevenmile Hill, OR	45°39′ N	121°16′ W	573 m	30 m

See http://me.oregonstate.edu/ERRL/bpa.info.html for more information.

Derived Variables

1. **Quantitative Power:** We convert speed to power assuming a deterministic transformation and denote this variable by, for example, $V_{p,t}$.
2. **Categorical Power Change:** Quantitative wind power is converted to a change in wind power, $V_{pc,t}$, with with decreases coded a -1, no changes coded a 0, and increases coded a 1, as follows:

$$V_{pc,t} = \begin{cases} -1, & V_{p,t} - V_{p,t-1} < -\delta_1, \\ 0, & -\delta \le V_{p,t} - V_{p,t-1} \le \delta_1, \\ 1, & V_{p,t} - V_{p,t-1} > \delta_1, \end{cases} \quad (1)$$

where $\delta_1 > 0$ is some small change in wind power that would not impact the system significantly.

3. **Current State of Power:** The power output variable, $V_{po,t}$, denotes the current state of the power output:

$$V_{po,t} = \begin{cases} -1, & V_{p,t} \le \delta_2 \text{ kW}, \\ 0, & \delta_2 \text{ kW} < V_{p,t} < 660 - \delta_2 \text{ kW}, \\ 1, & V_{p,t} \ge 660 - \delta_2 \text{ kW}, \end{cases} \quad (2)$$

where $\delta_2 > 0$ is a small value of power, selected to ensure that this variable represents powers that are either very close to the maximum or very close to the minimum.

4. **Recent 20-min Power Trend:** The twenty minute trend variable, $V_{20,t}$, denotes whether a decrease, an increase, or no clear trend has occurred over the past 20 minutes.

$$V_{20,t} = \begin{cases} -1, & (V_{pc,50,t-1}, V_{pc,0,t}) = (-1,-1), \\ 0, & (V_{pc,50,t-1}, V_{pc,0,t}) \in \{(-1,0),(0,-1),(-1,1),(1,-1),(1,0),(0,1),(0,0)\}, \\ 1, & (V_{pc,50,t-1}, V_{pc,0,t}) = (1,1). \end{cases} \quad (3)$$

Summary of Variables Used in Models

Variable/Location	Vansycle	Kennewick	Goodnoe Hills	Sevenmile Hill
Wind Speed	$V_{s,t}$	$K_{s,t}$	$G_{s,t}$	$S_{s,t}$
Wind Direction	$V_{d,t}$	$K_{d,t}$	$G_{d,t}$	$S_{d,t}$
Sine of Direction	$V_{sin,t}$	$K_{sin,t}$	$G_{sin,t}$	$S_{sin,t}$
Cosine of Direction	$V_{cos,t}$	$K_{cos,t}$	$G_{cos,t}$	$S_{cos,t}$
Power	$V_{p,t}$	$K_{p,t}$	$G_{p,t}$	$S_{p,t}$
Power Change	$V_{pc,t}$	$K_{pc,t}$	$G_{pc,t}$	$S_{pc,t}$
Current Power	$V_{po,t}$	—	—	—
20-min Trend	$V_{20,t}$	—	—	—

Benchmark Models

We desire to forecast the following probabilities at Vansycle:

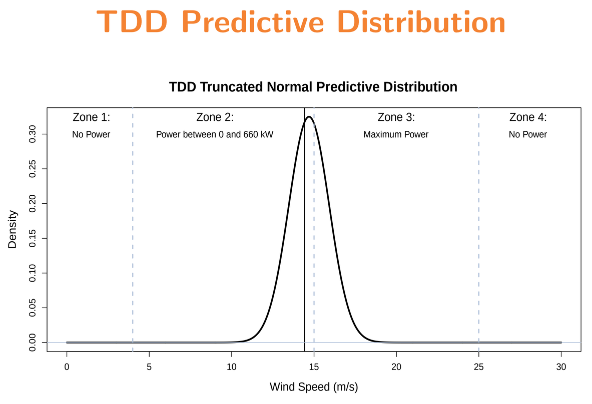
- decrease— $\pi_{-1,t+1} = P(V_{pc,t+1} = -1)$
- no change—$\pi_{0,t+1} = P(V_{pc,t+1} = 0)$
- increase—$\pi_{1,t+1} = P(V_{pc,t+1} = 1)$

The forecast, $\hat{V}_{pc,t+1}$, is the event corresponding to the maximum of the set $\{\hat{\pi}_{-1,t+1}, \hat{\pi}_{0,t+1}, \hat{\pi}_{1,t+1}\}$.

Persistence (PER): $\hat{V}_{pc,t+1} = V_{pc,t}$, the observed change in power between the prior and current hours.

Trigonometric Direction Diurnal (TDD): Hering and Genton (2010) model used for short-term wind speed forecasts is adjusted to make one hour ahead forecasts at Vansycle.

TDD Predictive Distribution



- The areas under the curve corresponding to a decrease, no change, or an increase for a given power curve are used to estimate the probabilities.
- This approach can be used for any forecasting model that produces a full predictive distribution and for any type of power curve.

Markov Chain Models

- **One-Step MC Model:** Assume that the response, $V_{pc,t+1}$, is random with discrete states $S = \{-1, 0, 1\}$.
- For a given sequence of time points, $1 \le 2 \le \cdots \le t+1$, the process should possess the following Markov property:

$$P[V_{pc,t+1} = j | X_1 = i_1, \ldots, X_{t-1} = i_{t-1}, X_t = i_t] = P[V_{pc,t+1} = j | X_t = i_t],$$

where $j, i_1, \ldots, i_t \in S$, and the conditioning variable, X_t, could be any one of $V_{pc,t}$, $V_{po,t}$, or $V_{20,t}$.

- The transition probabilities at time t can be organized in a transition matrix whose rows sum to 1 as

$$P_t^1 = \begin{bmatrix} \pi_{-1,-1} & \pi_{-1,0} & \pi_{-1,1} \\ \pi_{0,-1} & \pi_{0,0} & \pi_{0,1} \\ \pi_{1,-1} & \pi_{1,0} & \pi_{1,1} \end{bmatrix}$$

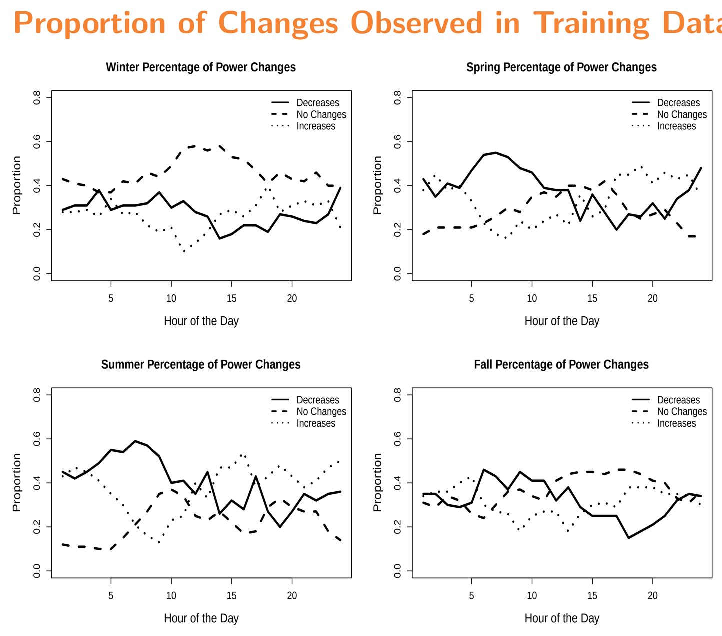
- Estimated transition probabilities are obtained by summing the frequencies of changes from state i to state j as follows:

$$\hat{\pi}_{ij} = \frac{\sum_{t=1}^{n-1} I(V_{pc,t+1} = j | X_t = i)}{\sum_{t=1}^{n-1} I(X_t = i)},$$

for $i, j \in S$, and where $I(\cdot)$ is an indicator function.

- Rolling windows of 180 days prior to the target forecast are used to estimate the transition matrices to account for seasonalities.

Proportion of Changes Observed in Training Data



- **Multi-Step MC Models:** The response states, $S_r = \{-1, 0, 1\}$, remain unchanged, but the the set of explanatory states, S_e, increases.
  - 2 Step Model $= 9 \times 3$ transition matrix
  - 3 Step Model $= 27 \times 3$ transition matrix
  - 4 Step Model $= 81 \times 3$ transition matrix
  - 5 Step Model $= 243 \times 3$ transition matrix
- Variables are selected with $V_{pc,t}$ always in the model since $V_{pc,t+1}$ is the response. Later time lags may enter the model only if the prior time lags at that location are present.
  The locations of interest are Vansycle, Kennewick, Goodnoe Hills, and Sevenmile Hill at time lags of t, $t-1$, and $t-2$.

Model Comparison

- For the MC1, MC2, and MC3 models, the addition of both $V_{po,t}$ and $V_{20,t}$ to the locations alone increases accuracy by $\sim 13\%$.
- The one, two, and three location MC models with the additional two variables perform very similarly. Thus, using the power change variable at one location alone is sufficient.
- The MC1+PO,20 model is 13.42% better than PER, which corresponds to over 1,000 forecasts and nearly **50 days.**
- The MC1+PO,20 model is 4.25% better than the TDD model, and this is roughly equivalent to 372 forecasts, or **15.5 days**, over the course of a year.

Overall Model Accuracy

Model	Overall Accuracy
PER	63.96%
TDD	73.13%
(1) MC1	63.96%
(4) MC1+PO,20	**77.38%**
(7) MC2$^{(K)}$	64.19%
(19) MC2$^{(K)}$+PO,20	**77.50%**
(26) MC3$^{(G,K)}$	63.88%
(56) MC3$^{(G,K)}$+PO,20	**76.98%**
(27) MC3$^{(G,S)}$	64.13%
(57) MC3$^{(G,S)}$+PO,20	**77.03%**

- The majority of errors occur when a decrease occurs but an increase is forecast or vice versa.
- MC2 model that includes $K_{pc,t}$ (19) has some of the lowest frequencies of forecasting an increase when a decrease occurs, indicating that this next closest location to Vansycle is provides some relevant information.
- If the MC models are used specifically to forecast a "no change,"
  - For MC1+PO,20, in only 5.37% (1.50% plus 3.87%) of cases is a no change forecast and then an increase or decrease actually occurs.
  - Increases or decreases are forecast in only 0.38% (0.13% plus 0.25%) of the hours when a no change occurs.
  - The total chance of misclassifying a no change with this model occurs in only **5.75%** of all forecasts.

Types of Errors in Model Forecasts

		True Change: $\mathbf{V_{pc,t+1}}$		
$\mathbf{\hat{V}_{pc,t+1}}$	**Model**	**Decrease**	**No Change**	**Increase**
Decrease	PER	18.57%	3.92%	11.42%
	TDD	22.50%	0.51%	10.47%
	(1) MC1	18.57%	3.92%	11.42%
	(4) MC1+PO,20	23.07%	**0.13%**	**7.53%**
	(7) MC2$^{(K)}$	24.32%	4.26%	16.94%
	(19) MC2$^{(K)}$+PO,20	**25.30%**	**0.14%**	9.70%
	(26) MC3$^{(G,K)}$	22.08%	4.04%	15.02%
	(56) MC3$^{(G,K)}$+PO,20	24.61%	0.30%	9.18%
	(27) MC3$^{(G,S)}$	22.13%	4.03%	14.82%
	(57) MC3$^{(G,S)}$+PO,20	23.66%	0.33%	8.28%
No Change	PER	1.51%	29.58%	3.89%
	TDD	**1.37%**	33.52%	**3.54%**
	(1) MC1	1.51%	29.58%	3.89%
	(4) MC1+PO,20	1.50%	**34.60%**	3.87%
	(7) MC2$^{(K)}$	1.51%	29.58%	3.89%
	(19) MC2$^{(K)}$+PO,20	1.53%	**34.63%**	3.85%
	(26) MC3$^{(G,K)}$	1.53%	29.60%	3.89%
	(56) MC3$^{(G,K)}$+PO,20	1.50%	34.34%	3.92%
	(27) MC3$^{(G,S)}$	1.60%	29.65%	3.95%
	(57) MC3$^{(G,S)}$+PO,20	1.53%	34.42%	3.89%
Increase	PER	13.82%	1.48%	15.81%
	TDD	10.03%	0.95%	17.11%
	(1) MC1	13.82%	1.48%	15.81%
	(4) MC1+PO,20	9.34%	**0.25%**	**19.71%**
	(7) MC2$^{(K)}$	8.08%	1.14%	10.29%
	(19) MC2$^{(K)}$+PO,20	**7.08%**	**0.21%**	17.57%
	(26) MC3$^{(G,K)}$	10.30%	1.34%	12.20%
	(56) MC3$^{(G,K)}$+PO,20	7.80%	0.34%	18.03%
	(27) MC3$^{(G,S)}$	10.17%	1.30%	12.35%
	(57) MC3$^{(G,S)}$+PO,20	8.71%	0.23%	18.95%

Multinomial Confidence Intervals

Compare the following two sets of forecasts:

$$\hat{\pi}_{-1} = 0.45, \ \hat{\pi}_0 = 0.03, \text{ and } \hat{\pi}_1 = 0.52$$

$$\hat{\pi}_{-1} = 0.12, \ \hat{\pi}_0 = 0.07, \text{ and } \hat{\pi}_1 = 0.81.$$

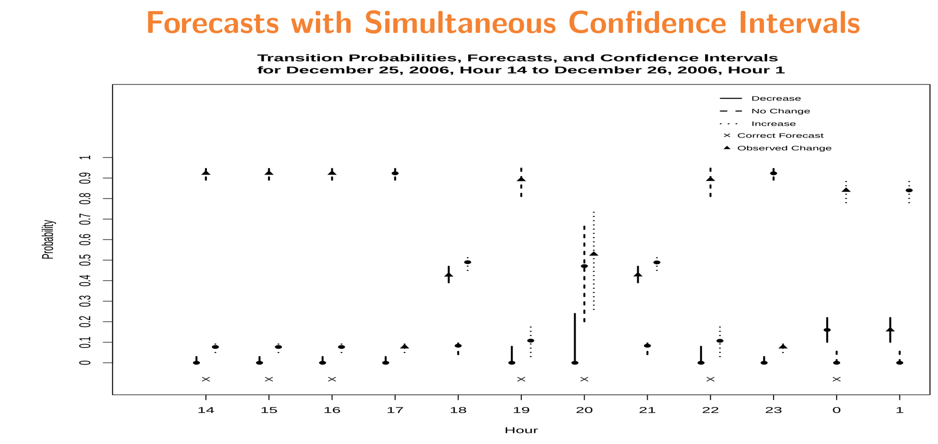
The second set indicates that the probability of an increase is more likely, but we also need to quantify the uncertainty in this estimate.

Fitzpatrick and Scott (1987) proposed a set of simultaneous multinomial confidence intervals. The $100(1-\alpha)\%$ simultaneous confidence interval for π_i is

$$\frac{n_i}{n} \pm \frac{z_{\alpha/4}}{2\sqrt{n}},$$

where $z_{\alpha/4}$ is the upper quantile from a standard normal distribution.

Forecasts with Simultaneous Confidence Intervals

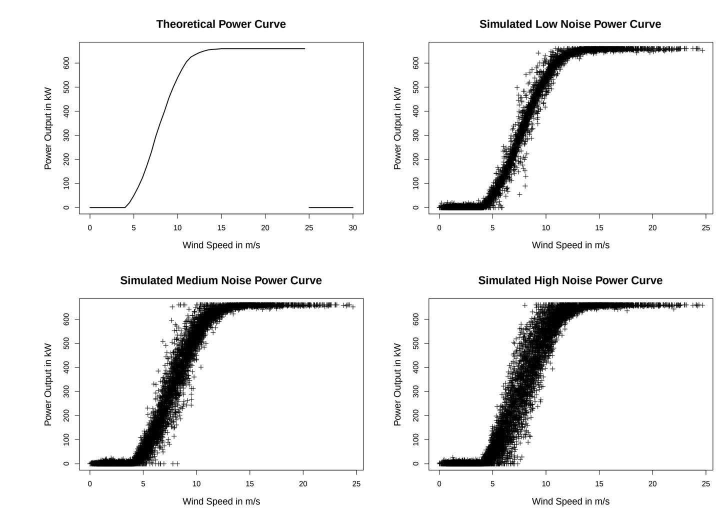


- For the MC1+PO,20 model, in 129 hours, the forecast was incorrect, but the forecast category had a confidence interval for its probability that overlapped with the confidence interval for the observed outcome's category.
  This gives a total of **78.86%** of hours in which the observed change in power was contained within the confidence intervals.

Profile and Power Curve Effects

- Adding noise to the wind power is expected to have some effect on the accuracies of the MC models, so the accuracy of the MC1+PO,20 model when applied to wind powers with different levels of contamination and for different choices of vertical profile is investigated.
- We use the Pinson et al. (2008) method for generating noise in a power curve.

Levels of Noise in Power Curve



To generate contaminated power data:

1. Scale the 10-minute wind speed observations to 50 m agl using one of six variations of scaling (2 profile laws)×(3 profile roughness coefficients).
2. Transform these scaled 10-minute wind speeds to power using the theoretical power curve.
3. Contaminate the 10-minute wind powers with low, medium, or high levels of noise.
4. Use the noisy 10-minute power values to create the values for $V_{20,t}$ from Equation (3).
5. Create noisy hourly averages by taking hourly averages of the noisy 10-minute power values.
6. Use the noisy hourly averages to create the values for $V_{pc,t}$ and $V_{po,t}$ from Equations (1) and (2), respectively, where $\delta_1 = 3.5$ and $\delta_2 = 5$.
7. Obtain the MC1+PO,20 forecasts, $\hat{V}_{pc,t+1}$ using $V_{pc,t}$, $V_{po,t}$, and $V_{20,t}$.
8. Compare $V_{pc,t+1}$ and $\hat{V}_{pc,t+1}$ to obtain a percentage accuracy.
9. Repeat steps 3 through 8 100 times for each combination of profile law, roughness coefficient, and noise level.

Average Accuracies for Varying Levels of Noise and Profile Laws

		Power Law			Log Law		
Noise	**Model**	0.13	0.14	0.15	0.01	0.02	0.03
Low	PER	39.13	39.10	39.05	39.55	39.51	39.47
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	MC	65.16	65.18	65.11	65.46	65.40	65.44
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Medium	PER	38.61	38.49	38.61	38.99	38.94	38.91
		(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
	MC	64.42	64.26	64.51	64.65	64.56	64.51
		(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
High	PER	38.09	38.03	38.09	38.41	38.38	38.32
		(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
	MC	63.55	63.52	63.60	63.65	63.67	63.57
		(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)

- The type of vertical scaling and roughness coefficient appear to have very little impact on the accuracy.
- MC1+PO,20 model's accuracy is reduced between 12% and 14%.
- Without noise in the data, the MC1+PO,20 model beats persistence by 13%, but with noise in the data, the MC1+PO,20 model improves upon persistence by approximately 26%.

Conclusions

- **The best overall model is the MC1+PO,20 model.**
- Models built to perform a particular function, such as forecasting wind speed, do not automatically produce the best forecast for a different type of response, such as a categorical change in power.
- Including the current status of the power output and the most recent trend in power changes dramatically improve the performance of the models.
- Including off-site locations did not significantly improve the forecast accuracy, perhaps due to the short forecast horizon.

References

- Fitzpatrick, S. and Scott, A. (1987) "Quick simultaneous confidence intervals for multinomial proportions," *Journal of the American Statistical Association*, 82: 875–878.
- Hering, A. S. and Genton, M. G. (2010) "Powering up with space-time wind forecasting." *Journal of the American Statistical Association*, 105: 92–104.
- Pinson, P., Nielsen, H. Aa., Madsen, H., and Nielsen, T. S. (2008) "Local linear regression with adaptive orthogonal fitting for the wind power application," *Statistics and Computing*, 18: 59–71.
- **Yoder, M., Hering, A. S., Navidi, W. C., and Larson, K. (2013) "Short-term forecasting of categorical changes in wind power with Markov chain models," Wind Energy, Now Online.**

For more information, email ahering@mines.edu.