Wind Power Forecasting Algorithms and Application

Statistics Seminar – Toulouse School of Economics

Ricardo Bessa (rbessa@inescporto.pt)
Talk Overview

• Introduction to the wind power forecasting problem

• Information Theoretic Learning for Wind Power Point Forecast

• Time-adaptive Quantile-Copula for Wind Power Uncertainty Forecast

• Application: Setting the operating reserve (very brief presentation)

• Avenues for future research
Introduction

• Wind power forecasting is vital for two groups of end-users
  – system operators → managing the power system
  – wind power producers → bidding in the electricity market

• Synergy between three research areas
  – Meteorology (physical): numerical weather predictions
  – Statistics (mathematical): time series and data mining models
  – Power systems (forecast value): decision-making problems with forecasts as input

• It is a business
  – several companies sell wind power forecasts as a service
    • Prewind Lda, spin-off company from three institutes in Portugal
  – other companies sell forecasting systems
Exhibit A: Nonlinear Conversion of Wind Speed to Power

The non-linear part amplifies the wind speed forecast error.
Exhibit B: Noisy Data

Power curve with forecasted wind speed
Exhibit C: Evolving Structure of Data

• Continuous stream of measured data from the wind farms
• Changes in the data (or concept drift)
  – limitation of the maximum produced energy
  – new wind farm in the same region
  – maintenance of wind farms
  – loss in performance
  – changes in the wind speed prediction model
  – etc...

• Time-adaptive models are needed!
Statistical and Physical Forecasting Models

Physical

- Wind Farm and terrain characteristics
- Numerical Weather Prediction (NWP) Atmospheric Variables
- Transformation to Hub Height
  - Spatial Refinement
    - Local Roughness, Orography, Atmospheric Stability
- Wind Turbine Power Curve
  - Conversion to power
- Model Output Statistics (MOS)
  - Systematic Error Correction
- Wind Generation Forecast

Statistical

- Numerical Weather Prediction (NWP) Atmospheric Variables
- SCADA Data
- Statistical Model
  - Wind Generation Forecast

Hybrid: Physical + Statistical
Very Short-term Wind Power Forecasting (~6 hrs ahead)

SCADA: supervisory control and data acquisition

<table>
<thead>
<tr>
<th>Wind Speed Forecasting</th>
<th>Wind Power Forecasting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalman Filter</td>
<td>Adaptive Linear Models</td>
</tr>
<tr>
<td>Grey Predictor</td>
<td>Fuzzy Time Series</td>
</tr>
<tr>
<td>Takagi-Sugeno</td>
<td>Self-exciting Threshold Autoregressive</td>
</tr>
<tr>
<td>Discrete Hilbert Transform</td>
<td>Smooth Transition Autoregressive</td>
</tr>
<tr>
<td>Abductive Networks (GMDH)</td>
<td>Markov-switching Autoregressive</td>
</tr>
<tr>
<td></td>
<td>Adaptive Fuzzy Logic Models</td>
</tr>
<tr>
<td></td>
<td>Adaptive Linear Models</td>
</tr>
<tr>
<td></td>
<td>ARIMA time series models</td>
</tr>
<tr>
<td></td>
<td>Neural Networks</td>
</tr>
<tr>
<td></td>
<td>Adaptive Neural Fuzzy Inference System</td>
</tr>
</tbody>
</table>
Short-term Wind Power Forecasting (~72hrs ahead)

NWP: Numerical Weather Predictions (e.g. Meteo France)

**Methods**
- Neural Networks
- Support Vector Machines
- Regression Trees with Bagging
- Random Forests
- Adaptive Neural Fuzzy System
- Mixture of Experts
- Nearest Neighbor Search
- Autoregressive with Exogenous input (ARX)
- Locally Recurrent Neural Networks
- Local Polynomial Regression
- Takagi-Sugeno FIS
- Fuzzy Neural Networks
- Bayesian Clustering by Dynamics (BCD)
Model Chain for Forecasting

NWP Point Forecasts → Wind Power Point Forecast Model → Probabilistic Model → Probabilistic Forecasts

Information Theoretic Learning for Wind Power Forecast


• Wind power forecast errors are non-Gaussian
  – Mean Square Error (MSE) is only optimal under Gaussian distribution
• The Information Theoretic Learning (ITL) idea...
  – The ideal case is when the pdf of errors $\epsilon$ is a Dirac function - all errors equal (of the same value)

$$\epsilon = T - O$$

- with all errors equal, one could have perfect matching between output $O$ and target $T$, by adding a bias term
Correntropy as a Cost Function

- Correntropy is a generalized similarity measure between Target (T) and Output (O)

\[ J_{\sigma}(T, O) = \frac{1}{N} \sum_{i=1}^{N} G(t_i - o_i, \sigma) \]

G: Gaussian kernel

- MCC (Maximum Correntropy Criterion) \( \rightarrow \) \( \max J_{\sigma} \rightarrow \) maximizing the pdf of errors at the origin
Correntropy Induced Metric (CIM)

- Contours of CIM (X,0) - distance to the origin in a 2d space (for $\sigma = 1$)
- Small errors: Euclidian (behaves like MSE)
- Medium size errors: Manhattan

Large errors: indifference
Results for a Real Wind Farm with a Neural Network

![Graph showing normalized error and frequency distribution](image)

- **Higher frequency of errors close to zero**
- **Not Gaussian**

**Statistics Seminar | 13 Dec. 2011 | Toulouse School of Economics**
ILT based Training: Remarks

• ITL criteria (correntropy in particular) presented better results than MSE, in terms of
  – higher frequency of errors close to zero
  – insensitivity to outliers
  – best fit of the predictions to the actual values verified

• ITL criteria cannot be ignored when building robust wind power prediction models

• The ITL criteria were used in neural networks, however it can be applied in other statistical and machine learning methods
Why Kernel Density Forecast?

• The information provided by point forecasts is unsatisfactory for some decision-making problems

• Modeling of wind power uncertainty is “distribution free”

• Uncertainty characterized by the full distribution

• Fast method for uncertainty modeling

• High flexibility to represent uncertainty (e.g. density func., quantiles, mass func.)

• Captures multimodal pdf
Kernel Density Estimation (KDE)

\[
\hat{f}_X(x) = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{1}{h_i} \cdot K \left( \frac{x - X_i}{h} \right)
\]

\[
K(x) = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x-X_i)^2}{2\cdot h^2}}
\]

Conditional KDE: estimating the density of a r.v. \(Y\), knowing that the explanatory r. v. \(X\) is equal to \(x\)

Joint or multivariate density function of \(X\) and \(Y\)

\[
f(y|X = x) = \frac{f_{XY}(x, y)}{f_X(x)}
\]

Marginal density of \(X\)
Quantile-Copula Estimator

Copula Definition

\[ F_{XY}(x, y) = C(F_X(x), F_Y(y)) \]

multivariate distribution function separated in:

• marginal functions
• dependency structure between the marginals, modeled by the copula

\[ f(x, y) = \frac{\partial^2}{\partial u \cdot \partial v} \cdot C(u, v) = f_X(x) \cdot f_Y(y) \cdot c(u, v) \]

\[ f(y|X = x) = \frac{f_X(x) \cdot f_Y(y) \cdot c(u, v)}{f_X(x)} = f_Y(y) \cdot c(u, v) \]

KDE ESTIMATOR

\[ \hat{f}_Y(y) = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{1}{h_y} \cdot K \left( \frac{y - Y_i}{h_y} \right) \]

\[ \hat{c}(u, v) = \frac{1}{N} \sum_{i=1}^{N} K \left( \frac{u - U_i}{h_u} \right) \cdot K \left( \frac{v - V_i}{h_v} \right) \]

\[ U_i = F_X^e(X_i) \text{ and } V_i = F_Y^e(Y_i) \]

\[ F^e: \text{empirical cum. dist.} \]
Quantile-Copula Estimator

\[
\hat{f}(y|X = x) = \frac{1}{N \cdot h_y} \cdot \sum_{i=1}^{N} K_y \left( \frac{y - Y_i}{h_y} \right) \cdot \frac{1}{N} \cdot \sum_{i=1}^{N} K_u \left( \frac{F_X^e(u) - F_X^e(U_i)}{h_u} \right) \cdot K_v \left( \frac{F_X^e(v) - F_X^e(V_i)}{h_v} \right)
\]

- Advantages
  - methods based on the Nadaraya-Watson are numerically unstable when the denominator is close to zero
  - for a problem with several explanatory variables, this method has only one kernel product, instead of two
  - purely based on density estimation methods
  - gives the full pdf for the wind power uncertainty

Time-adaptive Estimator

Recursive Estimator

\[ \hat{f}_n(x) = \frac{n - 1}{n} \cdot \hat{f}_{n-1}(x) + \frac{1}{n \cdot h_i} \cdot K \left( \frac{x - X_i}{h_i} \right) \]

for stationary data streams, and \(1/n\) aprox. 0 when \(n \to \infty\)

Exponential Smoothing

\[ \hat{f}_n(x) = \lambda \cdot \hat{f}_{n-1}(x) + \frac{(1 - \lambda)}{h_i} \cdot K \left( \frac{x - X_i}{h_i} \right) \]

for nonstationary data streams

forgetting factor

\[ \lambda = \frac{n}{n + 1} \]
Time-adaptive Quantile-Copula Estimator

**Time-adaptive empirical cumulative distribution function**

\[
F^e(x)_t = \lambda \cdot F^e(x)_{t-1} + (1 - \lambda) \cdot I(x_i \leq x)
\]

**Time-adaptive conditional KDE**

\[
\hat{f}(y|x = X)_t = \hat{f}_t(y) \cdot \hat{c}_t(u, v)
\]

\[
\hat{f}_t(y) = \lambda \cdot \hat{f}_{t-1}(y) + (1 - \lambda) \cdot K_h \left( \frac{y - Y_i}{h} \right)
\]

\[
\hat{c}_t(u, v) = \lambda \cdot \hat{c}_{t-1}(u, v) + (1 - \lambda) \cdot \left[ K_1 \left( \frac{F^e_X(u) - F^e_X(U_i)}{h_q} \right) \cdot K_2 \left( \frac{F^e_X(v) - F^e_X(V_i)}{h_q} \right) \right]
\]
Formulation of the Wind Power Forecast Problem

Forecast the wind power pdf at time step $t$ for each look-ahead time step $t+k$ of a given time-horizon knowing a set of explanatory variables (NWP forecasts, wind power measured values, hour of the day)

$$
\hat{f}_P(p_{t+k} | X = x_{t+k} | t) = \frac{f_{P,X}(p_{t+k}, x_{t+k} | t)}{f_X(x_{t+k} | t)}
$$
Quantile-Copula Estimator vs Classical Multivariate KDE

Wind Speed (m/s) vs Power (p.u.):

Joint Density Function

copula density $c(u,z)$

Nadaraya-Watson Approach

Quantile-copula
Kernel Choice

- Depends on the variable data and type
- In this problem we have two different types
  - Variables with range $[0,1]$: wind power and quantiles transforms $u$ and $v$
    - beta kernels
  - circular variables: hour of the day and the wind direction
    - Von Mises distribution

\begin{align*}
\text{Beta Kernel} & \\
\text{Gaussian Kernel, for var. with } ]-\infty,+\infty[ & \\
\text{Von Mises dist.} &
\end{align*}
Macro-beta Kernel

• The integrals computed from the beta kernels may
  – lead to distributions that do not have an integral equal to 1
  – the kernel is also inconsistent for distributions that are point mass at 0% and 100%

• This is due to lack of normalization, and the idea is a modified beta kernel estimator (named “macro-beta”)

\[ \hat{f}'(x) = \frac{\hat{f}(x)}{\int_0^1 \hat{f}(x)dx} \]

normalization is employed over the conditional function
Case-Study and Evaluation Metrics

- NREL dataset: wind power generation of 15 sites
- U.S. Midwest wind farm
  - 42 hrs ahead forecasts
  - NWP produced by a GFS + WRF model chain

- Benchmark algorithms
  - Splines quantile regression (or additive quantile regression)
  - Nadaraya-Watson KDF estimator

- Evaluation metrics
  - **Calibration**: deviation between forecasted and estimated probabilities, e.g. the 85% quantile should contain 85% of obs. values lower or equal to its value
    - Primary requirement measure by the deviation to perfect calibration
  - **Sharpnes**: mean size of the forecast intervals (target: narrow intervals)
  - **Skill score**: combined information about the predictor’s performance in a single measure (less negative the better, 0 for perfect probabilistic forecasts)
Proof of Concept – Data with Artificial Change

NREL dataset: test dataset Oct-Dec, connection of 2 sites after Oct

Nominal proportion rate [%]
Deviation [%]
5 15 25 35 45 55 65 75 85 95
-10 -5 0 5 10 Offline
Lambda=0.999
Lambda=0.995
Lambda=0.99

quantiles overestimation

quantiles underestimation
Evaluation Results – Offline Version

**Sharpness**

- Nominal coverage rate [%]
- Intervals mean length [% of rated power]

**Skill Score**

- Look-ahead Time [hours]
- QC, NW, SplinesQR
Evaluation Results – Time-adaptive Version

Time-adaptive
Nadaraya-Watson estimator

Quantile-copula
Nadaraya-Watson estimator

Nominal proportion rate [%]
Deviation [%]

Offline
Time-adaptive (n=2738)
Time-adaptive (n=1000)
Time-adaptive (n=200)

Nominal proportion rate [%]
Deviation [%]

Offline
Time-adaptive (n=2738)
Time-adaptive (n=1000)
Time-adaptive (n=200)
Evaluation Results – Time-adaptive Version

**Sharpness**

- Offline
- Time-adaptive (n=2738)
- Time-adaptive (n=1000)
- Time-adaptive (n=200)

**Skill Score**

- Offline
- Time-adaptive (n=2738)
- Time-adaptive (n=1000)
- Time-adaptive (n=200)
Kernel Density Forecast: Remarks

• The time-adaptive QC copes with evolving data and non Gaussian data

• Provides probabilistic forecasts that can be used in several decision-making problems

• Quantile-copula have a tendency to present a better performance in terms of calibration

• the quantile regression methods have a tendency to present a better performance in terms of sharpness

• the skill score of quantile-copula and quantile regression is rather similar

• the time-adaptive approach changes the bias of the probabilistic forecasts (calibration), while changing slightly the sharpness and resolution
Application: Setting the Power System Operating Reserve

• The operation of Electric Energy Systems must take into account possible unbalances between generation and load
  – Failures of the generating units
  – Load forecast errors
  – Wind, Solar, etc, forecast errors

• This leads to the need of an Operational Reserve able to respond to possible problems
  – Generating units that can take load immediately or in a short period

• The TSO (Transmission System Operator) must define this reserve for the next hours

• The recent increase in wind power penetration turned this exercise more difficult, due to the forecasting uncertainty
  – in a recent past, TSO defined the reserve based on empirical rules (such as, the largest unit + 2% of the forecasted load)
Application: Setting the Power System Operating Reserve

Tool developed in the EU Project Anemos.plus and demonstrated during 6 months for a end-user

Avenues for Future Research

- Method/rule for computing the “optimal” kernel bandwidth for the wind power problem

- Simultaneous density forecast (includes the temporal correlation of forecast errors)
  - the alternative is to have time trajectories (i.e. random vectors)

- Extreme events forecasting (e.g. extreme wind speed → disconnection of several wind farms)

- Evaluation of probabilistic forecasts (utility-based evaluation)
Wind Power Forecasting Algorithms and Application

Statistics Seminar – Toulouse School of Economics

Ricardo Bessa (rbessa@inescporto.pt)