A Cross-Validated Combined Forecasting Method And Its Application in Power Wind

Shangbin Wang, Xinghui Ren*, Honghai Li, Yong Zhai, Liwei Yuan and Zhimeng Yuan
Shandong LuNeng Software Technology Co., Ltd, Jinan, Shandong, 250002, China
ls_zhaiy@163.com

Abstract: A Cross-Validated combined forecasting method is proposed to retain more data information of wind power at each time level, so that the wind power forecasting results at different time scales meet the aggregation constraints. We used the actual data of a wind farm in Shandong Province to verify the effectiveness of the method. The results show that compared with the existing method, forecasting with Cross-Validated combined increases accuracy over conventional forecasting, particularly in different time levels with larger sampling interval.

1. Introduction

Wind power forecasting can be divided into ultra-short term, short term, medium and long term according to the length of time[1]. From the point of power grid dispatching, the above forecasting results are mostly used for real-time dispatching, day-ahead dispatching, maintenance plan, annual generation plan and others[2,3]. Different time scale wind power forecasting focuses on different aspects, and uses different information and statistical methods. For example, the medium and long term wind power forecasting focuses on the factors that long-term role in the time series, and the short-term and ultra short-term wind power forecasting focuses on the factors that short-term role in the time series[4-9]. Therefore, the forecasting results of wind power at different time scales often do not meet the aggregation constraints, resulting in inconsistent dispatching decisions.

In view of the above problems, the research on the forecasting results can not meet the aggregation constraints due to the differences of information and statistical methods can be traced back to 1972[10]. Before 2009, the general method was to calculate forecasting results at a single level and then aggregate them[11].Take the method of “Bottom Up” for example, the forecasting are generated at the lowest level and then aggregated to a higher level in the hierarchy. Reference 12 proposes a combined forecasting method, which is used to solve the problem of information loss in the aggregation process of single-level forecasting results. Reference 13 optimizes the methods in reference 12, proposing that each level of forecasting forms "original" or "basic" forecasting results, and weighted combination of all levels of forecasting results to ensure the consistency of the overall level of forecasting results. On the basis of the above research results, reference 14 proposes the concept of time hierarchy and the method of temporal hierarchies forecasting: first, the corresponding information is used to make the basic forecasting at different time levels, and then the basic forecasting is integrated and optimized based on the temporal hierarchies to get the correction value of the forecasting results of each time level, and make the forecasting results meet the aggregation constraints.

In this paper, we first study the existing time stratified combined forecasting method, then analyze the deficiencies of the existing method in the correction of time stratified forecasting, and finally propose

*Corresponding author
Email: xhren@163.com
to use cross validation to improve the existing method. We apply the improved algorithm to wind farm power forecasting and compare it with existing methods, in order to provide new ideas for improving the consistency of wind power forecasting results at different time scales.

2. Temporal Hierarchies Combination Forecasting Method

2.1. Time hierarchy

In this paper, the time series of wind power is analyzed with 15 minutes as the sampling interval, which is aggregated into time series with sampling interval of 1 hour and 1 day.

As shown in the figure 1, the time series from bottom to top can be used for ultra-short term, short term and medium-long term wind power forecasting. Figure 1 is only a schematic diagram of time hierarchy. The number of layers and the sampling interval of each layer of time series can be adjusted according to the actual situation. The change of the number of levels only affects the number of times of hierarchical aggregation, and the change of the sampling interval of time series in each layer only affects the number of nodes in each layer, and does not affect the application of time level combination method. As shown in Figure 1, the sampling interval of the second layer is aggregated into 30 minutes. The above changes do not affect the application of the time level combination method.

2.2. Temporal Hierarchies Combination Forecasting

First, define a multi-level time series \( \{ y_i \}, T = 1, 2, ..., T \), \( T \) is observation time length of time series. Set \( m \) as the sampling frequency of the top time level in the time hierarchy, it can be seen that \( T \) is a multiple of \( m \). Set \( k \) as the number of time series with the maximum sampling frequency in each time level, it can be seen that \( k \) is a divisor of \( m \). \( \{ k \} \) makes up a complete time hierarchy.

Taking Figure 1 as an example, \( m \) is the sampling frequency of wind power in one day, \( m = 96 \); If the sampling time is one year, then the observation time is \( T = m \times 15 \times 365 \), At the same time, we know the most complete time level \( k \in \{ 96, 48, 32, 24, 12, 8, 6, 4, 3, 2, 1 \} \). There are three levels in Figure 1, \( k \in \{ 96, 24, 1 \} \). Set \( i = 1, 2, ..., T / m \), \( i \) is different days in a year, Set \( p = 1, 2, ..., m / k \), \( i \) is the node change of each level in one day, the node values of each level in time series \( \{ y_i \} \) can be expressed as \( y^{[2]}_{m / k (i-1)+p} \). The time hierarchy shown in Figure 1 can be specifically expressed as follows:

Figure 2. Hierarchical forecasting structure of wind power with sampling interval of 15 minutes, 1 hour and 1 day
For each time level in period $i$, it can be expressed as

$$Y_i^{[k]} = (Y_{m/k(i-1)+1}^{[k]}, Y_{m/k(i-1)+2}^{[k]}, \ldots, Y_{m/k2}^{[k]})^T$$

(1)

Let $L$ represent the level matrix after the hierarchy descending, $k_j = m$, $k_j = I$. Thus, the time series can be further expressed as

$$Y_i = (Y_i^{[k_j]}, Y_i^{[k_{j-1}]}, \ldots, Y_i^{[k_1]})^T$$

(2)

From here we see that $Y_i = SY_i^{[h]}$, $S$ is a summation matrix, which can be stacked by submatrix $S_k$, the submatrix $S_k$ can be obtained by copying the $k$ times of the unit matrix of $m/k$ size row by row. So $S$ is a matrix of order $(\Sigma m/k \times m$. $S$ is a matrix of order $252 \times 96$ when $m = 96$.

Assuming the basic forecasting of wind power in step $h^*$ at the bottom of time hierarchy, Then $h = 1, \ldots, h^*/m$ is the number of forecasting steps for the entire hierarchy. The basic forecast of each time level can be expressed as follows:

$$\hat{Y}_i(\cdot) = (Y_i^{[k_1]}, Y_i^{[k_{j-1}]}, \ldots, Y_i^{[k_1]})$$

(3)

Furthermore, the basic forecasting of $h$ step in the whole time level can be expressed as follows:

$$\hat{Y}(\cdot) = S\beta_i(h) + \varepsilon_h$$

(4)

Among them, $\beta_i(h) = E(Y_i^{[h]} | Y_i, \ldots, Y_1)$ is the unknown mean value of the lowest wind power in the future. $\varepsilon_h$ is the error of adjustment of wind power forecasting value between time levels, the covariance is $\sum h$.

Using generalized least squares to estimate $\beta_i(h)$, the following formula as follows:

$$\beta_i^{GLS}(h) = (S^T \sum h^+ S)^{-1} S^T \sum h \hat{Y}_i(h)$$

(5)

Among them, $\sum h^+$ is the generalized inverse of $\sum h$. So we can get the combined forecasting model of time level as follows:

$$\tilde{Y}_i(h) = S\beta_i^{GLS}(h) = S(S^T \sum h^+ S)^{-1} S^T \sum h^+ \hat{Y}_i(h)$$

(6)

In fact, $\sum h$ is unknown. For the solution of $\sum h$, the current overall idea is to introduce relevant parameters to simplify the solution, the main methods include Bottom Up(BU), Bottom Average(BA), Global Average(AG), Weighted Least Squares (WLS), etc.

In reference 13, Wickramasuriya and others introduced the minimum trace estimator $W_i$ of single step basic forecasting error covariance. Get the following formula

$$\beta_i^{MinT}(h) = (S^TW_i^{-1}S)^{-1} S^TW_i^{-1}\hat{Y}_i(h)$$

(7)

Among them, $\hat{Y}_i = \frac{1}{T/m} \sum_{i=1}^{T/m} e_i e_i^T$, $e_i$ is the single step basic forecasting error.

Based on the time hierarchy, using the Weighted Least Squares method we can get that
\[ \hat{W}_i = \Lambda = \text{diag}[m, k_{i-1}, k_{i-1}, ..., k_{i-2}, k_{i-2}, k_1, k_1] \]  \hspace{1cm} (8)

\[ \beta^{\text{MinT}}_i(h) \text{ can be further expressed as} \]

\[ \beta^{\text{WLS}}_i(h) = \left( S^T A^{-1} S \right)^{-1} S^T A^{-1} \hat{Y}_i(h) \]  \hspace{1cm} (9)

\[ \hat{Y}_i(h) = SP^{\text{WLS}}(h) = S \left( S^T A^{-1} S \right)^{-1} S^T A^{-1} \hat{Y}_i(h) \]  \hspace{1cm} (10)

Direct estimation \( \Lambda \) is more complex, in reference \cite{14}, Athanasopoulos and others further simplifies \( \Lambda \) into three kinds of diagonal matrix: Hierarchical variance scale matrix \( \Lambda_H \), Variance scale matrix \( \Lambda_V \), Structure scale matrix \( \Lambda_S \). Since \( \Lambda \) is a diagonal matrix and the non-diagonal elements are all zero, the information data of different time hierarchies will be lost when the above methods are used for forecasting.

2.3. Improved Time Combination Forecasting Method

In order to solve the problem of information loss in the process of introducing relevant parameters to simplify the solution of time level combined forecasting model \( Y_i(h) \), In this paper, Cross-Validation is used to obtain the optimal solution of \( Y_i(h) \), this method can retain more information at different time levels, and make the forecasting results have better aggregation constraints at different time levels.

Set \( P = (S^T \sum h^+ S)^{-1} S^T \sum h^+ \), equation (6) can be expressed as

\[ \hat{Y}_i(h) = SP \hat{Y}_i(h) \]  \hspace{1cm} (10)

Then the key to solve \( \tilde{Y}_i(h) \) is to solve matrix P.

The matrix P is solved as follows:

(1) The multi-level time series \( \{ y_i \} \) is divided into three parts which are not overlapped, training set \( \{ y_i \}^{\text{train}} \), verification set \( \{ y_i \}^{\text{val}} \), test set \( \{ y_i \}^{\text{test}} \).

(2) Using the data in training set \( \{ y_i \}^{\text{train}} \) to estimate model parameters and express these estimates as \( \hat{\theta}^{\text{train}} \). The Cumulative Distribution Function (CDF) is introduced to predict each time level in validation set \( \{ y_i \}^{\text{val}} \), and get the unconsolidated cumulative distribution function \( \hat{F}(y_i^{t+h} | f_i^l; \hat{\theta}^{\text{train}}) \).

The cumulative distribution function \( \hat{F}(y_i^{t+h} | f_i^l; \hat{\theta}^{\text{train}}) \) of combined forecasting is obtained by left multiplying projection matrix \( SP \) of \( \hat{F} \).

(3) Take \( \hat{F}_i^{t+h} \) as the boundary value of the cumulative distribution function corresponding to the \( j \)-th node in time hierarchy level \( l \). Using Continuous Ranked Probability Score (CRPS) as the scoring rule for generalization ability of model. F is the predicted value based on the cumulative distribution function, and Z is the hierarchical scaling weight parameter.

In the above method, the cross validation objective function value is

\[ CV(P) = L^{-1} \sum_{l=1}^{L} CV_l(P) \]  \hspace{1cm} (11)
Because the matrix P is large (P is the 96 × 121 order matrix in Figure 1). This paper presents a sparse structure of matrix P. Taking \( f_i = \{4, 2, 1\} \) as an example, matrix P can be optimized with the following sparse structure.

\[
P_{CV} = \begin{bmatrix}
v_{1,1} & v_{2,1} & 0 & 0 & 0 & 0 \\
v_{1,1} & v_{2,1} & 0 & 0 & v_{3,2} & 0 \\
v_{1,1} & 0 & v_{2,1} & 0 & 0 & v_{3,3} \\
v_{1,1} & 0 & v_{2,1} & 0 & 0 & v_{3,4}
\end{bmatrix}
\]

(13)

Where \( v_{r,j} \) represents the weight of the r-th element in time level \( l \).

In the cross validation process, the following three cases are considered to constrain the medium weight in \( P_{CV} \):

1. All elements in \( P_{CV} \) are positive, and the sum of elements in each row is 1;
2. The sum of elements in each row is 1;
3. All elements in are unconstrained.

Using Continuous Ranked Probability Score (CRPS) as the scoring rule, and the R as follows:

\[
R(F, z) = \int_u (F(u) - 1)^2 du \quad z \leq u
\]

\[
R(F, z) = \int_u F^2(u)du \quad z > u
\]

3. Application Analysis of Method

As a case study of the method, we use 15 minutes time series of wind power from a wind farm. The installed capacity of the wind farm in Shandong Province is 49.5MW. The time series are the wind power output data from October 2018 to September 2019.

Firstly, the 12-month wind power output observation data is divided into training sample data, verification sample and test sample according to 6:3:3 time ratio. The wind power output data of six months from October 2018 to March 2019 will be used as model training to train the wind power probability forecasting models at all time levels; The wind power output data of three months from April 2019 to June 2019 are used as model validation, and the weight of probability forecasting results in each time level is obtained by cross validation; The wind power output data of three months from July 2019 to September 2019 are used to test and evaluate the generalization ability of the model.

Then, the time series is divided into four levels according to 15 minutes, 1 hour, 4 hours and 24 hours, \( f = \{96, 24, 6, 1\} \), and the wind power of each time level is predicted separately to get the basic forecasting value. The time level and forecasting method are shown in table 1.

| Time level | Forecasting model |
|------------|-------------------|
| 24 hours   | ARFIMA-FIGARCH    |
| 4 hours    | VARMA-GARCH       |
| 1 hour     | ARMA-GARCH        |
| 15 minutes | ARMA-FIGARCH      |

The cross validation method in 1.2.2 is used to optimize \( P_{CV} \) weight under different constraints based on the validation sample data. Using the sparse structure in equation (13) and taking equation (11) as the objective function, the forecasting results of each time level model in table 1 are modified. Table
shows the weight mean under different constraints of each time level.

| Constraint condition | Time level sampling interval |
|----------------------|-----------------------------|
|                      | 24 hours | 4 hours | 1 hour | 15 minutes | Sum |
| $\sum v_i = 1 \& \forall v_i \geq 0$ | 0.00 | 0.02 | 0.87 | 0.11 | 1.00 |
| $\sum v_i = 1$ | -0.26 | 0.49 | 0.64 | 0.13 | 1.00 |
| Unconstrained | -0.02 | 0.24 | 0.37 | 0.04 | 0.63 |

Test set $\{y_{i test}\}$, Based on BU, BA, GA, WLS and other methods, the CRPS value of the wind power forecasting value of 127 (96 + 24 + 6 + 1) nodes at each forecasting origin under each method is evaluated. Take the mean value in $\{y_{i test}\}$, and then take the mean value again on all nodes in each time level to get the data. The last column of table 3 is the average value of all the first rows in the same row.

| Method types | Time level sampling interval |
|--------------|-----------------------------|
|              | 24 hours | 4 hours | 1 hour | 15 minutes | average |
| No correction | 1.69 | 1.76 | 1.70 | 1.74 | 1.72 |
| Bottom Up | 1.34 | 1.67 | 1.71 | 1.74 | 1.61 |
| Bottom Average | 1.34 | 1.66 | 1.71 | 1.74 | 1.61 |
| Global Average | 1.32 | 1.65 | 1.70 | 1.73 | 1.60 |
| LA | 1.38 | 1.72 | 1.77 | 1.80 | 1.67 |
| Weighted Least Squares | 1.32 | 1.65 | 1.70 | 1.73 | 1.60 |
| Cross-Validated | 1.27 | 1.64 | 1.69 | 1.72 | 1.58 |
| $\sum v_i = 1 \& \forall v_i \geq 0$ | 1.28 | 1.64 | 1.69 | 1.72 | 1.58 |
| $\sum v_i = 1$ | 1.28 | 1.65 | 1.70 | 1.73 | 1.59 |
| Unconstrained | 1.28 | 1.65 | 1.70 | 1.73 | 1.59 |

Under different methods, the average CRPS of each level in the time hierarchy is shown in table 3. Analyzing the average CRPS of each level in table 3, we can get the following conclusions

1. Based on the time hierarchy, no matter which combination forecasting method is chosen to adjust and optimize the basic forecasting, the aggregate constraints of the forecasting results are better than the independent forecasting;
2. In the time level combination forecasting method, the optimization effect of the time level with larger sampling interval is better than that of the time level with smaller sampling interval. That is to say, the time level combination forecasting method improves the accuracy of the time level with larger sampling interval, which is more obvious than that with smaller sampling interval;
3. The cross validation method proposed in this paper optimizes the forecasting results of other conventional combination forecasting methods, and the aggregation constraints of the forecasting results are better than those of other conventional combination forecasting methods.

4. Summary
This paper introduces the concept and related methods of time hierarchical composite forecasting. Aiming at the difficulty of covariance estimation in current composite forecasting methods, a cross validation time hierarchical composite forecasting method is proposed. By optimizing the structure of scale scaling matrix $P$, more information of different time hierarchical structures is retained.

The forecasting results of the actual power data of the wind farm show that the cross validation time layered combination forecasting method can effectively improve the forecasting accuracy of each time level, especially the time level with large sampling interval, and the forecasting results are better than
the aggregation constraint effect of other conventional combination forecasting methods.

References

[1] XUE Yusheng, YU Chen, YU Chen, ZHAO Junhua, et al. A Review on Short-term an Ultra-short-term Wind Power Forecasting[J]. *Automation of Electric Power Systems*, 2015, 39(06):141-151.

[2] WU Wenzu, QIAO Ying, LU Zongxiang, et al. Forecasting Method and Prospect of Wind Power Probability [J]. *Automation of Electric Power Systems*, 2017, 41(18):167-175.

[3] CHANG Kang, DING Maosheng, XUE Feng, et al. Ultra-short-term wind power forecasting and its application in early-warning system of power systems security and stability[J]. *Power System Protection and Control*, 2012, 40(12):19-24+30.

[4] YAN Jie. Uncertainty of Wind Power Forecasting and Power System Economic Dispatch [D]. *North China Electric Power University (Beijing)*, 2016.

[5] CHEN Siyu. Research on Ultra Short-term Power Forecasting of Wind Power Generation [D]. *Liaoning University of Technology*, 2016.

[6] LI Guo. Wind Power Forecasting Method for Power System Decision. *Tianjin University*, 2017.

[7] XU Yansong. Research on Medium and Long Term Wind Power Forecasting Method [D]. *Shenyang Institute Of Engineering*, 2017.

[8] SHI Jie. Research on Optimization of Ultra-short-term Forecasting Algorithm of Wind Power [D]. *North China Electric Power University*, 2012.

[9] PENG Xiaosheng, XIONG Lei, WEN Jinyu, et al. A Summary of the State of the Art for Short-term and Ultra-short-term Wind Power Forecasting of Regions[J]. *Proceedings of the CSEE*, 2016, 36(23):6315-6326+6596.

[10] AMEMIYA, T and WU, R. Y. The Effect of Aggregation on Forecasting in the Autoregressive model[J]. *Journal of the Am-erican Statistical Association*, 67 628-632, 1972.

[11] ATHANASOPOLOS, G, AHMED, R.A, HYNDMAN, R.J. Hierarchical forecasts for Australian domestic tourism[J]. *International Journal of Forecasting*, 146–166, 2009.

[12] HYNDMAN R J, AHMED R A, ATHANASOPOULOS G. and SHANG H. L. Optimal combination forecasts for hierarchical time series[J]. *Computational Statistics & Data Analysis*, 55(9), 2579-2589, 2011.

[13] WICKRAMASURIYA S L, ATHANASOPOULOS G, HYMDAN R J. Forecasting Hierarchical and Grouped Time Series Through Trace Minimization[R]. *Technical Report, Monash University*, 2015.

[14] ATHANASOPOLOS. G, HYNDMAN.R.J,NIKOLAOSKOURENTZES. Forecasting with Temporal Hierarchies [J].*European Journal of Operational Research, February 2017.*