

Hierarchical Electricity Time Series Forecasting for Integrating Consumption Patterns Analysis and Aggregation Consistency

Yue Pang¹, Bo Yao¹, Xiangdong Zhou¹, Yong Zhang², Yiming Xu¹ and Zijing Tan¹

¹ School of Computer Science, Fudan University, Shanghai, China

² State Grid Shanghai Municipal Electric Power Company, Shanghai, China

¹{ypang15, byao16, xdzhou, 16210240068, zjtan}@fudan.edu.cn, ²ecgc@163.com

Abstract

Electricity demand forecasting is a very important problem for energy supply and environmental protection. It can be formalized as a hierarchical time series forecasting problem with the aggregation constraints according to the geographical hierarchy, since the sum of the prediction results of the disaggregated time series should be equal to the prediction results of the aggregated ones. However in most previous work, the aggregation consistency is ensured at the loss of forecast accuracy. In this paper, we propose a novel clustering-based hierarchical electricity time series forecasting approach. Instead of dealing with the geographical hierarchy directly, we explore electricity consumption patterns by clustering analysis and build a new consumption pattern based time series hierarchy. We then present a novel hierarchical forecasting method with consumption hierarchical aggregation constraints to improve the electricity demand predictions of the bottom level, followed by a “bottom-up” method to obtain forecasts of the geographical higher levels. Especially, we observe that in our consumption pattern based hierarchy the reconciliation error of the bottom level time series is “correlated” to its membership degree of the corresponding cluster (consumption pattern), and hence apply this correlations as the regularization term in our forecasting objective function. Extensive experiments on real-life datasets verify that our approach achieves the best prediction accuracy, compared with the state-of-the-art methods.

1 Introduction

Electricity demand forecasting is a crucial problem for power grid [Li *et al.*, 2012; Taylor, 2012]. Precise prediction can give effective guidance to enable reasonable decision making and management on power generation and supply, and benefits the economy and environmental protection issues. As the smart grid develops faster, massive number of individual electricity consumption data (such as smart electricity meter data) can be obtained conveniently [Zheng *et al.*, 2013; Siano, 2014]. The available electricity consumption data

with large scale and wide variety leads to new challenges to the communities of research and industry [Ramchurn *et al.*, 2012].

In electricity demand forecasting, the system structure is usually built on the geographical hierarchy [Yang *et al.*, 2015]. For instance, electricity demand in a city can be disaggregated into ones of districts through the administrative divisions, which are further disaggregated into ones of streets and blocks, etc.. The “aggregation consistency” is a critical point in hierarchical forecasting, that is the disaggregated time series should add up equally to the high level aggregated ones. Since the aggregation consistency is very unlikely ensured in the forecasting of the entire time series independently, the “bottom-up” approach is adopted. It forecasts all of the bottom level disaggregated series and then adds the forecasts results to form various aggregated series forecasts to achieve the aggregation consistency. However, when the disaggregated data tends to have low signal-to-noise ratio, the overall prediction accuracy especially for the disaggregated series forecasting is poor [Taieb *et al.*, 2017]. In recent work, various optimal reconciliation approaches become the mainstream. Hyndman *et al.* (2011) propose a solution of using least squares reconciliation to deal with the hierarchical forecasting problem. They apply a regression model to optimally reconcile the forecasts of the disaggregated and aggregated series correspondingly. However, the aggregation consistency is often achieved at the expense of the forecasting accuracy, since the side effects of the reconciliation and adjustments may enlarge the prediction errors.

In this paper, we propose a novel approach for electricity demand hierarchical forecasting. Instead of dealing with the geographical hierarchy directly, we make time series clustering analysis for electricity consumption pattern detection. Individual household electricity consumption (power usage) usually follows a steady pattern of growth over time in some extent. In this work we build a new consumption pattern based time series hierarchy based on the clustering results. When the time series with similar patterns are grouped together, the signal-to-noise ratio of the aggregated time series will be risen significantly, and the prediction accuracy of the aggregated ones will be improved. Figure 1 gives more illustration on this phenomenon. Note that the forecast of the electricity usage is composed of two parts: the ground truth of the electricity usage (denoted by T_A and T_D for aggregated

and disaggregated forecasts respectively) and the prediction error (denoted by E_A and E_D for aggregated and disaggregated forecasts respectively). With aggregation constraints, we have the equation: $T_A + E_A = \Sigma(T_D + E_D)$, and as the ground truth of the electricity consumptions at different levels are always equal, it can be deduced that $E_A = \Sigma E_D$. Therefore, if the E_A is decreased then in overall the prediction error of the disaggregated forecasts E_D will be decreased either with proper reconciliation methods. Based on the above motivation, we propose a new optimal reconciliation approach to improve the forecasting ability of the disaggregated time series. Finally the “bottom-up” approach is applied for adding up the disaggregated forecasts to form various high level aggregated forecasts, hence the aggregation consistency on the geographical hierarchy is realized.

In this paper, our main contributions are threefold:

1. We present a novel hierarchical electricity demand forecasting approach. Unlike the traditional “single path” solution, we exploit the idea of electricity consumption pattern analysis, the optimal reconciliation regression and the “bottom-up” approach altogether to deal with the hierarchical electricity time series forecasting problem.

2. In our approach we present a novel electricity consumption pattern based hierarchical forecasting method. It is based on the observation that the reconciliation error of the disaggregated time series is “correlated” to its membership degree of the corresponding cluster (the consumption pattern). Hence we propose a new optimization object function with a regularization term of penalizing the correlation of the two distributions.

3. We conduct experiments on two real-life datasets for performance evaluations. In our experiments, the state-of-the-art methods and the strong baselines are compared with our method on hierarchical electricity demand forecasting extensively. The experimental results show that our method achieves the best forecasting accuracy while keeping geographic aggregation consistency. Specially, Mean Absolute Percentage Error (MAPE) of our approach is 2.50% lower than that of the state-of-the-art methods.

2 Related Work

Base forecasting (BASE) [Hyndman *et al.*, 2011] estimates time series at all levels in a hierarchical structure independently. Some common models, such as ETS (Smoothing State Space model), are applied in the classical forecasting [Box and Jenkins, 1976; Hyndman *et al.*, 2008; De Livera *et al.*, 2011]. Due to the characteristic of independent forecasting, the predictions at different levels do not adversely affect each other. But the aggregation consistency is not taken into account in BASE. To satisfy the aggregation consistency, Bottom-up [Athanasopoulos *et al.*, 2009] adds up the bottom-level forecasts to obtain the upper ones.

Recently, the optimal combined forecasting becomes the mainstream in hierarchical time series forecasting with aggregation consistency. It estimates initial forecasts at bottom level by using BASE, then these forecasts are reconciled according to the geographic aggregation constraints. The method of Ordinary Least Square (OLS) [Hyndman *et al.*, 2011] re-

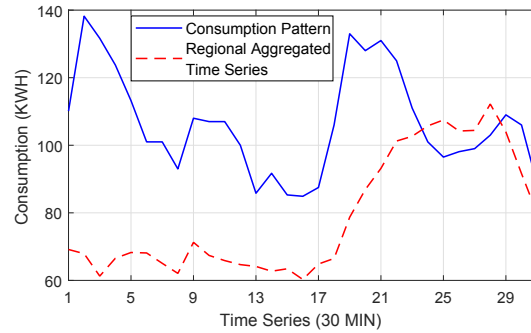


Figure 1: The electricity consumption pattern vs the region aggregation. By time series clustering, 278 time series are aggregated as the consumption pattern (the blue curve). The same number of 278 time series in one region are aggregated as the region aggregation time series (the red curve). Note that the consumption pattern has more regularity and its prediction error is less than that of the regional aggregated series. In addition, compared with single individual series, consumption pattern also has more regularity [Wijaya *et al.*, 2015].

gards the covariance matrix as an identity matrix in the adjustment model learning. Although OLS ensures aggregation consistency, the forecasting accuracy may be decreased due to the simple assumption of covariance matrix. HLS method [Athanasopoulos *et al.*, 2015] obtains weights from the hierarchical structure. It assumes that forecast errors of time series at the same level are equivalent while that of series at different levels are not. Different from HLS, other optimal combined forecasting methods adopt weight least square method [Wickramasuriya *et al.*, 2015]. WLS [Hyndman *et al.*, 2016] estimates the diagonal part of the covariance matrix by using the sample variance of the pre-forecasts. Because WLS estimates the covariance matrix based on the historical observations (pre-forecasts), the adjustments at the bottom level series may be affected negatively when the time series is not steady. MinT-Reg [Taieb *et al.*, 2017] obtains the sparse adjustments by solving a regression problem with the elastic net penalty.

For the general hierarchical forecasting problem, the Bayesian framework is used to obtain the proportion changes in hierarchical domestic tourism forecasting [Park and Nassar, 2014]; An adaptable regression method combined with multiple scalar forecasts to obtain accurate weather predictions is proposed [Williams *et al.*, 2016]. A data-driven decision support system [Huber *et al.*, 2017] is presented to predict the hierarchical demand of fast moving consumer goods.

3 Our Proposed Method

3.1 Consumption Pattern Analysis and Hierarchy Construction

We explore consumption patterns of electricity smart meter data. Consumption patterns are extracted from amounts of individual household time series by using the X-means [Pelleg *et al.*, 2000] clustering algorithm. As a toy example, two consumption patterns are extracted from clustering eight electricity time series, which is shown in Figure 2.

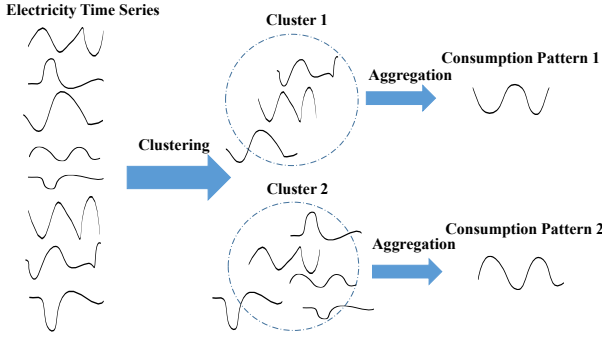


Figure 2: Consumption pattern analysis. The consumption pattern based hierarchy is build with the number of the high level series $N_a = 3$ and that of the bottom level series $N_b = 8$.

The ground truth of the electricity consumptions at different levels are always equal, as illustrated in Section 1. If the forecast of aggregated time series is accurate, the average prediction accuracy of individual time series will be likely improved under the aggregation constraints and reconciliation. We observe that the pattern of electricity time series is more regular than the aggregated time series based on geographical hierarchy (Figure 1). The prediction error of pattern is hence less than that of the regional aggregated time series. Motivated by the idea, we therefore construct a consumption pattern based hierarchy (Figure 3), consisting of patterns (the high level series) and individual time series (the bottom level series).

3.2 Hierarchical Forecasting Based on Aggregation Consistency

In the consumption pattern based hierarchy, the bottom level series adds up consistently over levels. We reconcile the bottom level forecasts under the constraints of $\bar{a}_{p,T+h} = S_{c,a}\bar{b}_{T+h}$, where \bar{a}_{T+h} and \bar{b}_{T+h} denote the mean aggregated h-period forecasts and disaggregated ones with T given historical observations respectively. S_c denotes an $N \times N_b$ summing matrix derived from the consumption pattern based hierarchical structure. It consists of an $N_a \times N_b$ submatrix $S_{c,a}$ and an $N_b \times N_b$ identity matrix. $N = N_a + N_b$.

$$S_c = \begin{bmatrix} S_{c,a} \\ I_{N_b} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ & & & & I_{N_b} & & & \end{bmatrix}. \quad (1)$$

Our goal is to obtain the reconciled bottom level forecasts by minimizing the expectation of square forecast errors.

$$\begin{aligned} & \arg \min .E \left[\left\| Y_{p,T+h} - S_c \tilde{b}_{T+h} \right\|_2^2 \right] \\ & = \arg \min_Q .E \left[\left\| Y_{p,T+h} - S_c Q \hat{Y}_{p,T+h} \right\|_2^2 \right], \end{aligned} \quad (2)$$

where

$$Y_{p,T+h} = \begin{bmatrix} y_{top,T+h} \\ y_{pattern1,T+h} \\ y_{pattern2,T+h} \\ y_{b1,T+h} \\ \vdots \\ y_{b8,T+h} \end{bmatrix}, \tilde{b}_{T+h} = \begin{bmatrix} \tilde{y}_{b1,T+h} \\ \vdots \\ \tilde{y}_{b8,T+h} \end{bmatrix}. \quad (3)$$

$$\hat{Y}_{p,T+h} = \begin{bmatrix} \hat{a}_{p,T+h} \\ \hat{b}_{T+h} \end{bmatrix} = \begin{bmatrix} \hat{y}_{top,T+h} \\ \hat{y}_{pattern1,T+h} \\ \hat{y}_{pattern2,T+h} \\ \hat{y}_{b1,T+h} \\ \vdots \\ \hat{y}_{b8,T+h} \end{bmatrix}. \quad (4)$$

$Y_{p,T+h}$ denotes an N -length vector of real forecasts at time $T+h$, \tilde{b}_{T+h} denotes an N_b -length vector of reconciled bottom level forecasts, Q denotes an $N_b \times N$ transformation matrix, \hat{b}_{T+h} and $\hat{Y}_{p,T+h}$ denote initial forecasts at bottom level and all levels based on ETS model, respectively.

The optimal solution in Equation 2 can be obtained by the best linear unbiased reconciled forecasts [Wickramasuriya *et al.*, 2015].

$$\tilde{b}_{T+h} = Q^* \hat{Y}_{p,T+h} = (S_c' W_h^{-1} S_c)^{-1} S_c' W_h^{-1} \hat{Y}_{p,T+h}, \quad (5)$$

where W_h denotes the covariance matrix of the h-period ahead forecast errors and its estimator of one-ahead base forecast errors is written as:

$$W_h = k_h \frac{1}{T} \sum_{t=1}^T (Y_{t+h} - \hat{Y}_{t+h})(Y_{t+h} - \hat{Y}_{t+h})', \quad (6)$$

where k_h denotes a positive constant. To simple the computation, W_h is assumed to be an identity matrix.

According to generalized least squares (GLS) [Kariya and Kurata, 2004], the bottom level forecasts given by Equation 5 can also be estimated from the regression model in Equation 7.

$$\hat{Y}_{p,T+h} = S_c \beta_{T+h} + \xi_h, \quad (7)$$

where β_{T+h} is the conditional mean of the reconciled bottom level forecasts \tilde{b}_{T+h} under the observations at all levels. ξ_h is the estimated error of initial forecasts, whose mean and variance are zero and covariance matrix \hat{W}_h , respectively. The GLS estimation of the bottom level forecasts then are obtained by minimizing the squared Mahalanobis length of the residual vector $\hat{Y}_{p,T+h} - S_c \tilde{b}_{T+h}$,

$$\arg \min_{\tilde{b}_{T+h}} (\hat{Y}_{p,T+h} - S_c \tilde{b}_{T+h})' \hat{W}_h^{-1} (\hat{Y}_{p,T+h} - S_c \tilde{b}_{T+h}). \quad (8)$$

In this way, our problem is formulated as an optimal GLS regression problem of forecast error, which can be effectively solved by using gradient descent method [Snyman, 2005]. This method is denoted by CHF in this work.

Let Θ_{T+h} be an N_b -length vector of adjustments of the bottom level time series, we have $\tilde{b}_{T+h} = \hat{b}_{T+h} + \Theta_{T+h}$.

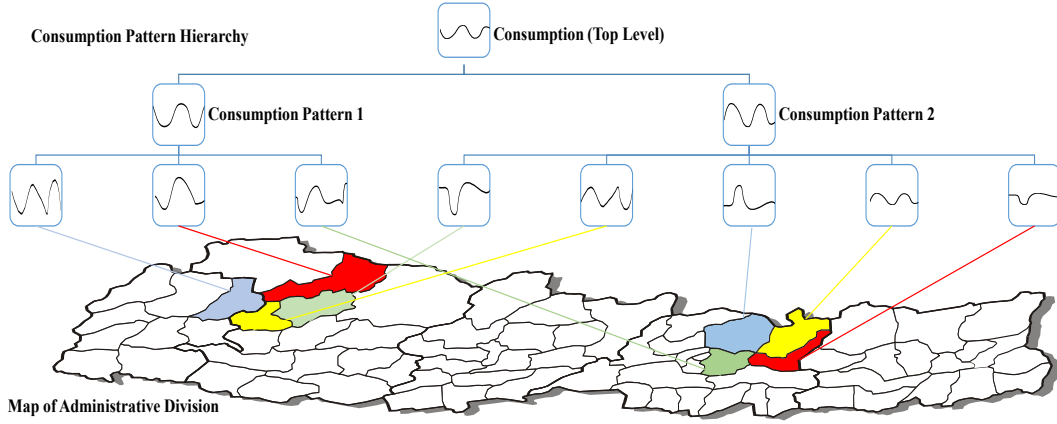


Figure 3: A clustering-based hierarchical electricity time series forecasting approach.

Then our new optimization object function is as follows:

$$\begin{aligned} & \arg \min_{\Theta_{T+h}} (\hat{Y}_{p,T+h} - S_c \hat{b}_{T+h} - S_c \Theta_{T+h})' \hat{W}_h^{-1} (\hat{Y}_{p,T+h} - \\ & S_c \hat{b}_{T+h} - S_c \Theta_{T+h}) + \lambda \|\Theta_{norm,T+h} - V_{norm}\|_2^2, \\ & s.t. \hat{b}_{T+h} + \Theta_{T+h} \geq 0 \end{aligned} \quad (9)$$

where $\Theta_{norm,T+h} = (\frac{|\theta_1|}{\sqrt{\frac{1}{N_b} \sum_{i=1}^{N_b} (\theta_i)^2}}, \dots, \frac{|\theta_{N_b}|}{\sqrt{\frac{1}{N_b} \sum_{i=1}^{N_b} (\theta_i)^2}})$, θ_i denotes the i -th unknown value of adjustment of the bottom level time series, $1 \leq i \leq N_b$, $V_{norm} = (\frac{d_1}{\sqrt{\frac{1}{N_b} \sum_{i=1}^{N_b} (d_i)^2}}, \dots, \frac{d_{N_b}}{\sqrt{\frac{1}{N_b} \sum_{i=1}^{N_b} (d_i)^2}})$, and d_i denotes the i -th distance between the bottom level time series and its corresponding clustering center (the consumption pattern), which can be estimated by using clustering analysis.

3.3 The Regularization

The regularization term in Equation 9 is defined by $L2$ norm [Hastie *et al.*, 2015] of the distance between $\Theta_{norm,T+h}$ and V_{norm} . $\lambda > 0$ is a penalty parameter. The constraint of $\hat{b}_{T+h} + \Theta_{T+h} \geq 0$ ensures the reconciled forecasts are greater than or equal to zero. This method is referred to as CHF-Reg in the rest of the paper. The regularization term is based on exploring the correlation between consumption pattern and individual electricity time series. Figure 4 shows that the distribution of the absolute value of ideal adjustment of real electricity individual time series is similar to that of the distances from the time series to the clustering center (the corresponding pattern).

A time series is more similar to the pattern means it is more stable and regular. That is, when the time series is closer to the clustering center, its forecast is more accurate and therefore requires smaller adjustment. In order to find the optimal adjustment, we hence introduce the regularization term to penalize the correlation between the distribution of the adjustment of series and that of its membership degree of the corresponding pattern. When the time series is closer to clustering center, the corresponding adjustment obtained by solving Equation 9 will be smaller.

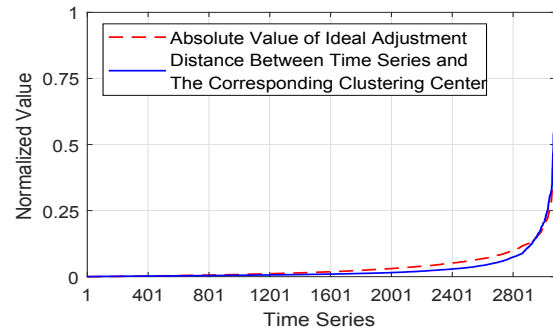


Figure 4: The distribution of the absolute value of ideal adjustment and the distance between the time series and the corresponding clustering center.

3.4 High Level Aggregated Forecasting

In order to keep geographical aggregation consistency, the “bottom-up” method is applied as shown in Figure 3. Based on the geographical hierarchy, we obtain the regional forecasts at all levels by summing the optimally reconciled bottom level forecasts. $Y_{region,T+h} = S_{region}(\hat{b}_{T+h} + \Theta_{T+h}^*)$. Where S_{region} denotes a summing matrix derived from the geographical hierarchical structure.

4 Experiments

4.1 Experimental Datasets

We use two real-life datasets on individual household electric power consumption. The first one is power grid data from State Grid Shanghai Municipal Electric Power Company. It contains 3078 time series, taken from January 1th, 2014 to February 19th, 2015, at a sampling rate of 24 hours. The second one is public electricity data from Energy Demand Research Project: Early Smart Meter Trials (EDRP) [Raw and Ross, 2011]. It contains 2501 time series at a sampling rate of 30 minutes.

4.2 Experimental Setup

In the experiments, we compute one-step-ahead rolling forecasts [Hyndman *et al.*, 2016] with historical consumption observations for the two datasets. Specifically, we repeat 15 forecasting tasks on the smart grid data from January 1th, 2014 to February 4th, 2015, so as to generate electricity demand in half a month. While for EDRP data, we repeat 48 forecasting tasks, from May 9th 00:00, 2009 to May 20th 09:30, 2009, in order to predict electricity demand in the whole day.

For each forecasting task, we use Mean Absolute Percentage Error (MAPE) [Wijaya *et al.*, 2015] and Root Mean Square Error (RMSE) [Yang *et al.*, 2017] to measure forecasting errors. They indicate the relative forecast errors and the average forecasting errors of household electricity consumption at one time, respectively.

In the experiments, we compare our method with the strong baselines and the state-of-the-art methods, including BASE, Bottom-up [Athanasopoulos *et al.*, 2009], HLS [Athanasopoulos *et al.*, 2015], WLS [Hyndman *et al.*, 2016] and MinT-Reg [Taieb *et al.*, 2017]. In our CHF and CHF-Reg, 32 types of patterns are extracted by clustering amounts of time series whose dimension is reduced to 50 by using PCA [Hotelling, 1933]. The value of λ are 3×10^5 and 10^4 for smart grid and EDRP dataset respectively.

4.3 Experimental Results and Analysis

The prediction accuracy at each level and the average statistical results for the two real-life datasets are shown in Table 1 - 4, respectively. The last columns in these tables indicate whether the forecasting methods satisfy the geographical aggregation consistency.

From the tables, we can see in general that the upper level predictions have less errors than the bottom level forecasts. Because the upper level forecasts have higher signal-to-noise ratios due to the aggregation. We can also observe that BASE has the least forecast errors in some upper levels, but it cannot preserve the constraints of geographical aggregation. In contrast, CHF-Reg not only satisfies the constraints, but also achieves a higher forecasting accuracy than BASE in most cases.

In Table 1 and Table 2, the MAPE of BASE is 0.14% lower than that of Bottom-up, and is 1.17% lower than that of HLS. It suggests that Bottom-up and HLS achieve the aggregate consistency at the expense of overall prediction accuracy. In contrast, CHF-Reg forecasts achieve lower MAPE and RMSE of 6.19% and 4.714 respectively than HLS. This is because that CHF-Reg optimally reconciles forecasts of the time series at bottom level according to the distances between these series and their corresponding clustering centers, while HLS uniformly reconciles the forecasts at the same level. The RMSE of CHF-Reg forecasts are 27.215 lower than that of WLS. The reason is that CHF-Reg estimates the adjustments based on patterns extracted from historical observations. Although both MinT-Reg and CHF-Reg have lower MAPE values than other methods, CHF-Reg still shows the best prediction performance. Specially, one-day electricity consumption forecasts of one household based on CHF-Reg is averagely 0.502 kilowatt hour more precise than MinT-Reg. This

is because CHF-Reg can estimate the appropriate adjustment values applied to the forecasts at all levels, which effectively mitigates the negative effect of estimation errors in the adjustments. In addition, CHF-Reg has better forecasting performance than CHF. It demonstrates the effectiveness of the introduced regularization in the object function.

Figure 5 and Figure 6 show the comparison of prediction accuracy for the six methods averaged over all time series at the bottom and upper levels, respectively. We can see

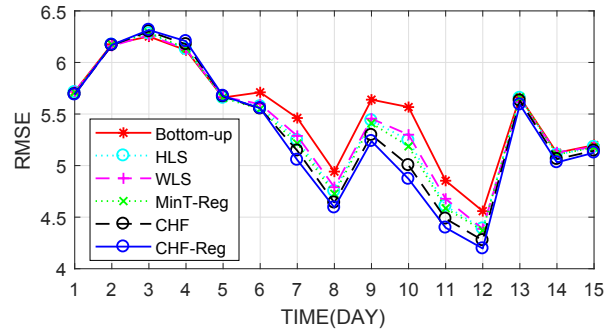


Figure 5: RMSE averaged over time series at the bottom level (smart grid dataset).

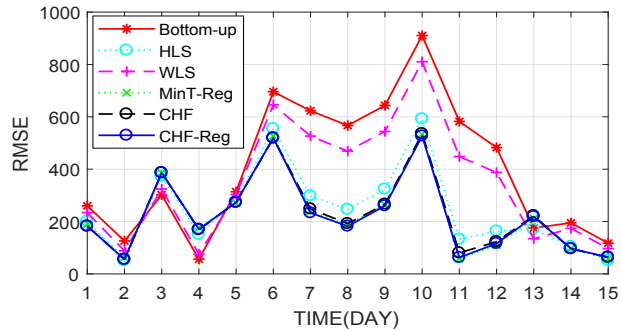


Figure 6: RMSE averaged over time series at the upper levels (smart grid dataset).

that CHF-Reg (solid line with circle markers) always has less forecasting errors at different horizontal coordinates than other methods.

In the forecasting results on the EDRP dataset in Table 3 and Table 4, it can also be found that CHF-Reg obtains the most precise forecasts than other methods that satisfy aggregate consistency. We can see that CHF-Reg forecasts show more obvious advantages on the first dataset than the second one. This is because the electricity time series data in the second dataset are more stable and hence their adjustments will be smaller. Overall, in term of both MAPE and RMSE, CHF-Reg has better predictive power and more robustness on two real datasets than strong baselines and the state-of-the-art methods.

Method	Top	Level 5	Level 4	Level 3	Level 2	Level 1/ Bottom	Average	Consistency
BASE	5.81	8.98	9.71	16.04	16.32	64.05	62.30	×
Bottom-up	11.56	13.54	14.42	20.01	19.97	64.05	62.44	✓
HLS (2015)	6.59	9.49	10.18	16.23	16.20	65.27	63.47	✓
WLS (2016)	9.60	11.20	11.94	17.75	17.80	62.05	60.43	✓
MinT-Reg (2017)	5.91	8.81	9.70	15.74	15.69	61.46	59.78	✓
CHF	5.87	8.67	9.50	15.56	15.54	60.34	58.70	✓
CHF-Reg	5.85	8.88	9.51	15.69	15.53	58.87	57.28	✓

Table 1: The comparison of MAPE obtained by different methods (smart grid dataset).

Method	Top	Level 5	Level 4	Level 3	Level 2	Level 1/ Bottom	Average	Consistency
BASE	2056.837	541.551	245.579	155.137	90.451	5.530	50.839	×
Bottom-up	3717.071	963.696	418.488	265.062	148.939	5.530	89.711	✓
HLS (2015)	2260.852	592.863	263.710	167.549	97.585	5.436	55.477	✓
WLS (2016)	3161.706	838.277	376.224	241.981	135.909	5.500	77.978	✓
MinT-Reg (2017)	2076.402	546.970	246.325	156.440	91.150	5.425	51.265	✓
CHF	2067.599	551.384	247.533	158.190	93.195	5.382	51.388	✓
CHF-Reg	2062.086	541.505	241.895	154.130	90.182	5.351	50.763	✓

Table 2: The comparison of RMSE obtained by different methods (smart grid dataset).

Method	Top	Level 5	Level 4	Level 3	Level 2	Level 1/ Bottom	Average	Consistency
BASE	3.60	22.29	21.14	21.65	26.04	68.69	67.21	×
Bottom-up	11.90	25.56	24.71	24.51	28.16	68.69	67.30	✓
HLS (2015)	4.45	22.12	20.89	20.92	24.55	66.34	64.91	✓
WLS (2016)	5.97	22.82	21.59	21.71	25.20	66.71	65.28	✓
MinT-Reg (2017)	3.72	44.83	35.02	31.98	32.09	66.46	65.36	✓
CHF	3.70	23.11	20.92	20.83	24.24	66.34	64.90	✓
CHF-Reg	3.69	22.93	20.61	20.48	24.08	65.23	63.82	✓

Table 3: The comparison of MAPE obtained by different methods (EDRP dataset).

Method	Top	Level 5	Level 4	Level 3	Level 2	Level 1/ Bottom	Average	Consistency
BASE	29.290	8.790	6.160	4.269	2.506	0.244	1.015	×
Bottom-up	70.673	16.395	10.893	6.832	3.574	0.244	1.966	✓
HLS (2015)	33.704	8.814	6.057	4.072	2.347	0.242	1.051	✓
WLS (2016)	40.661	10.076	6.833	4.459	2.499	0.243	1.206	✓
MinT-Reg (2017)	29.887	8.495	5.985	4.099	2.402	0.243	0.999	✓
CHF	29.293	8.066	5.590	3.752	2.192	0.242	0.952	✓
CHF-Reg	29.348	7.942	5.526	3.709	2.157	0.243	0.945	✓

Table 4: The comparison of RMSE obtained by different methods (EDRP dataset).

5 Conclusion

Aggregation consistency is one of the critical points in hierarchical time series forecasting. In most of the previous work, it is considered that the prediction accuracy is usually affected by the geographical aggregation constraints. In this work we demonstrate that a flexible usage of the aggregation constraints could bring the improvement of the electricity demand forecasting. To deal with the problem, we build a new data hierarchy by electricity consumption patterns analysis. Then based on the new hierarchy and our proposed optimization regression method with regularization, the prediction accuracy of the aggregated and disaggregated times series are significantly improved on the real-life datasets. In fu-

ture work, we will explore new hierarchy building approaches and the optimization prediction with aggregation consistency in the electricity demand forecasting problem and other application fields.

Acknowledgements

This work was supported by the National High Technology Research and Development Program (863 Program) of China (2015AA050203), NSFC grant No. 61370157 and No. 61572135, and the Shanghai Innovation Action Project (No.16DZ1100200, No.17DZ1203600).

References

- [Athanasopoulos *et al.*, 2009] George Athanasopoulos, Roman A Ahmed, and Rob J Hyndman. Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, 25(1):146–166, 2009.
- [Athanasopoulos *et al.*, 2015] George Athanasopoulos, Rob J Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos. Forecasting with temporal hierarchies. *Working paper 16/15, Department of Econometrics and Business Statistics, Monash University*, 2015.
- [Box and Jenkins, 1976] George EP Box and Gwilym M Jenkins. *Time series analysis: forecasting and control, revised ed.* Holden-Day, 1976.
- [De Livera *et al.*, 2011] Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527, 2011.
- [Hastie *et al.*, 2015] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical learning with sparsity: the lasso and generalizations.* CRC press, 2015.
- [Hotelling, 1933] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.
- [Huber *et al.*, 2017] Jakob Huber, Alexander Gossmann, and Heiner Stuckenschmidt. Cluster-based hierarchical demand forecasting for perishable goods. *Expert systems with applications*, 76:140–151, 2017.
- [Hyndman *et al.*, 2008] Rob Hyndman, Anne B Koehler, J Keith Ord, and Ralph D Snyder. *Forecasting with exponential smoothing: the state space approach.* Springer Science & Business Media, 2008.
- [Hyndman *et al.*, 2011] Rob J Hyndman, Roman A Ahmed, George Athanasopoulos, and Han Lin Shang. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, 55(9):2579–2589, 2011.
- [Hyndman *et al.*, 2016] Rob J Hyndman, Alan J Lee, and Earo Wang. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis*, 97:16–32, 2016.
- [Kariya and Kurata, 2004] Takeaki Kariya and Hiroshi Kurata. *Generalized least squares.* John Wiley & Sons, 2004.
- [Li *et al.*, 2012] Der-Chiang Li, Che-Jung Chang, Chien-Chih Chen, and Wen-Chih Chen. Forecasting short-term electricity consumption using the adaptive grey-based approach—an asian case. *Omega*, 40(6):767–773, 2012.
- [Park and Nassar, 2014] Mijung Park and Marcel Nassar. Variational bayesian inference for forecasting hierarchical time series. In *ICML Workshop*, 2014.
- [Pelleg *et al.*, 2000] Dan Pelleg, Andrew W Moore, et al. X-means: Extending k-means with efficient estimation of the number of clusters. In *ICML*, volume 1, pages 727–734, 2000.
- [Ramchurn *et al.*, 2012] Sarvapali D Ramchurn, Perukrishnen Vytelingum, Alex Rogers, and Nicholas R Jennings. Putting the ‘smarts’ into the smart grid: a grand challenge for artificial intelligence. *Communications of the ACM*, 55(4):86–97, 2012.
- [Raw and Ross, 2011] GJ Raw and DI Ross. Energy demand research project: Final analysis. 2011.
- [Siano, 2014] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.
- [Snyman, 2005] Jan Snyman. *Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms*, volume 97. Springer Science & Business Media, 2005.
- [Taieb *et al.*, 2017] Souhaib Ben Taieb, Jiafan Yu, Mateus Neves Barreto, and Ram Rajagopal. Regularization in hierarchical time series forecasting with application to electricity smart meter data. In *AAAI*, pages 4474–4480, 2017.
- [Taylor, 2012] James W Taylor. Short-term load forecasting with exponentially weighted methods. *IEEE Transactions on Power Systems*, 27(1):458–464, 2012.
- [Wickramasuriya *et al.*, 2015] Shanika L Wickramasuriya, George Athanasopoulos, Rob J Hyndman, et al. Forecasting hierarchical and grouped time series through trace minimization. *Department of Econometrics and Business Statistics, Monash University*, 2015.
- [Wijaya *et al.*, 2015] Tri Kurniawan Wijaya, Matteo Vasirani, Samuel Humeau, and Karl Aberer. Cluster-based aggregate forecasting for residential electricity demand using smart meter data. In *Big Data (Big Data), 2015 IEEE International Conference on*, pages 879–887. IEEE, 2015.
- [Williams *et al.*, 2016] John K Williams, Peter P Neillley, Joseph P Koval, and Jeff McDonald. Adaptable regression method for ensemble consensus forecasting. In *AAAI*, pages 3915–3921, 2016.
- [Yang *et al.*, 2015] Dazhi Yang, Gary SW Goh, Siwei Jiang, Allan N Zhang, and Orkan Akcan. Forecast upc-level fm-cg demand, part ii: Hierarchical reconciliation. In *Big Data (Big Data), 2015 IEEE International Conference on*, pages 2113–2121. IEEE, 2015.
- [Yang *et al.*, 2017] Dazhi Yang, Hao Quan, Vahid R Disfani, and Licheng Liu. Reconciling solar forecasts: Geographical hierarchy. *Solar Energy*, 146:276–286, 2017.
- [Zheng *et al.*, 2013] Jixuan Zheng, David Wenzhong Gao, and Li Lin. Smart meters in smart grid: An overview. In *Green Technologies Conference, 2013 IEEE*, pages 57–64. IEEE, 2013.