Application of Ensemble Kalman Filter in Forecasting the Electricity Grid Carbon Factor

Eng Tseng Lau and Qingping Yang
College of Engineering, Design, and Physical Sciences, Brunel University London, Uxbridge, UB8 3PH, UK
Email: {eng.lau, ping.yang}@brunel.ac.uk

Alistair Forbes and Valerie Livina
Mathematics and Modelling Group, National Physical Laboratory, Hampton Road, Teddington, Middlesex, TW11 0LW, UK
Email: {Alistair.Forbes, valerie.livina}@npl.co.uk

***Abstract*—Several publications have discussed the application of Ensemble Kalman Filter (EnKF) in history matching problems. The EnKF provides updated approximations based on the conditioned constraints to the historical data. In this paper we show how the EnKF is capable of forecasting/recovering the unpredictable trends of Electricity Grid Carbon Factor (EGCF). We adopt the EGCF scenario in the UK based on the available energy data provided by the Balancing Mechanism Reporting System (BMRS). We apply EnKF for forecasting the incomplete datasets in the UK EGCF in 2014. We present the ability of EnKF to recover the EGCF.**

***Index Terms*—history matching, recovery, ensemble Kalman filter**

I. Introduction

The international energy system experiences increasing unpredictable changes in the demand and load on the infrastructures. Such limitations may cause network power system failures, with addition of conventional, more polluting power stations, which eventually increase the level of Greenhouse Gases (GHG) emissions, particularly the EGCF. High level of EGCF may indicate poor utilizations of the whole electrical grids. Hence, a good forecasting technique is required in order to provide the EGCF forecast. In this paper we propose EnKF in forecasting EGCF and reconstruction of its trends. The obtained estimations may provide carbon inventory for assessment of energy network upgrades.

II. Electricity Grid Carbon Factor

According to [1], carbon factors are reported in kilograms of carbon dioxide (CO_2) equivalent per unit of energy (kWh). The EGCF refers to the average carbon factor across the energy grid according to fuel mix from different power plant generations (renewables and non-renewables) [2], [3]. The UK EGCF was introduced and studied in [2].

Manuscript received February 5, 2015; revised September 23, 2015.

III. Ensemble Kalman Filter

Based on the UK EGCF defined in [2], [3], the historical EGCF results have shown fluctuations corresponding to different seasons. The fluctuations are mainly due to the effect of fuel generation that is required to balance the energy demand among the consumers. This shows the need to forecast the uncertain trends of the EGCF (corresponding to energy generations) ahead with the intention of mitigating the risk of grid failures during the peak period, monitoring and maintaining the acceptable level of carbon emissions. The forecast may need to be performed when there are incomplete fuel data available. This requires the robust approaches in modelling and forecasting the EGCF. In order to address the problem with such limitations, EnKF [4] is adopted and demonstrated in forecasting or recovering the EGCF from the incomplete fuel data in 2014. EnKF has been also known to have the ability in real-time updating (propagation) and historical matching of an ensemble in a model that matches the given production or historical data [1], [2], [4], [5].

IV. Methodology

A. Estimations of Electricity Grid Carbon Factor

The EGCF resulting from the variations of fuel mix during energy generation is defined in [2], [3]:

$$EGCF(t) = \frac{\sum_{t=1}^{N_t} \sum_{m=1}^{N_m} \left(C_m \cdot E_m(t) \right)}{\sum_{t=1}^{N_t} E_m(t)} \tag{1}$$

where m is the fuel type index (renewables and non-renewables), N_m is the total number of fuels, C_m is the carbon factor for fuel type m, t is the time index, E_m is the generated energy corresponding to fuel type m.

The UK fuel-mix generation data is available publicly in [6]. However, the 2014 Quarter 4 (Q4) fuel-mix data is not yet available in [6]. This has impeded the electrical grid experts from evaluating the trend of the 2014 Q4

doi: 10.18178/ijoee.3.4.209-212

season that allows pre-cautionary actions to be taken beforehand. Therefore, EnKF is applied in forecasting (or recovering) the missing 2014 Q4 fuel data to determine the EGCF.

B. Covariance and Correlation Coefficient

Before we apply EnKF, it is necessary to determine the relationship of two global variables (the available historical and 2014 EGCF datasets from [2], [6]). Such determination is important as strong correlation between two variables will enable us to determine initializations (the prior knowledge) later before simulating the EnKF model.

We measure the covariance in order to determine the possible relationship between the two variables. The covariance *COV* is calculated using:

$$COV(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \quad (2)$$

where x and y refer to the variables (historical and 2014 EGCF datasets, together with their sample means), i indexes the sample, and n is the total number of samples.

We further compute the correlation coefficient using the following equation:

$$\rho_{x,y} = \frac{COV(x,y)}{\sigma_x \sigma_y} \quad (3)$$

where x and y are the variables (historical and 2014 EGCF datasets), σ_x and σ_y refer to the standard deviations of the variable x and y, respectively.

C. EnKF Algorithm

In EnKF, the variables of interests (fuel energy generation) are collected into state vector y:

$$y = \begin{bmatrix} m \\ d \end{bmatrix} \quad (4)$$

The m is the dynamic or static variables (fuel data) and d represents the observation variable (EGCF). The state vector y is determined based on the initial conditions (the prior knowledge of the model or historical records of earlier observations of the data). In this paper, the prior observation EGCF data (d) is obtained by using (1). The computation of (1) is based on the available fuel mix data (m) published in [6].

We perturb m and d with the model errors - the randomized Gaussian process noise Q with zero mean and covariance W for m and similarly, the Gaussian measurement noise R with zero mean and covariance V for d. Values Q and R are assumed to be drawn from Gaussian distributions with $Q \sim N(0,W)$ and $R \sim N(0,V)$. Model errors are very important as without model errors the whole system may be over specified and therefore no solutions resulting from EnKF propagations are obtained [7].

We create the ensemble of state vectors denoted as matrix Y in EnKF:

$$Y = [y_1, y_2, ..., y_{N_e}]^T \quad (5)$$

where N_e is the total number of ensembles. The ensemble y^p (*a priori ensemble*) from the observations will be propagated using EnKF to obtain the newly updated ensemble y^u (*a posteriori ensemble*) [2]:

$$y_j^u = y_j^p + C_Y H^T (H C_Y H^T + R)^{-1} (d_{obs,j} - H y_j^p) \quad (6)$$

where j indexes the ensemble members, C_y is the covariance matrix of state vector y, H is the measurement operator relating to the model variables, R is the covariance matrix of the measurement noise, and d_{obs} is the perturbed observations from (4).

We commit to determine the forecasting errors resulting of EnKF by applying the Root Mean Square Error (RMSE). RMSE in this paper is calculated as the differences between the average forecasted ($\mathbf{E}[d^u]$) and the EnKF mean observations ($\mathbf{E}[d_{obs}]$):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{E}\left[d_{obs}\right] - \mathrm{E}\left[d^u\right])^2} \quad (7)$$

where n is the total number of samples.

V. Results

A. Estimation of Electricity Grid Carbon Factor

We substitute the earlier results of estimated carbon factors from [1] into C_m in (1). The latter includes the samples of historical fuel data courtesy of Balancing Mechanism Reporting System (BMRS) [6] in tabulating the daily mean of EGCF. Fig. 1 illustrates the daily data of EGCF in 2014 (incomplete).

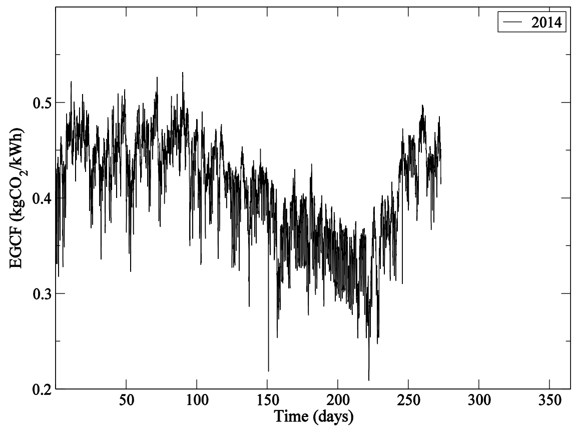

Figure 1. The incomplete 2014 EGCF dataset.

In Fig. 1, it is noted that the 2014 EGCF is only valid up to 273th day. Hence, we intend to forecast (or recover) the EGCF in the remaining Q4. Before we apply the EnKF forecast for 2014 Q4, we use (2) and (3) to find the correlation between the two datasets.

B. Estimation of Covariance and Correlation Coefficient

Table I shows the calculated covariance *COV* and correlation coefficient. The *COV* and the correlation coefficient of 0.5789 have confirmed a good correlation between the available 2014 datasets with the historical data from [6]. Hence, EnKF can be performed to forecast the EGCF in the remaining period of 2014. The historical EGCF data is later selected as the initial conditions (prior data) for initialling simulation for the EnKF model.

TABLE I. COVARIANCE AND CORRELATION COEFFICIENT

Covariance	0.0013
Correlation coefficient	0.5789

C. EnKF Propagation

Using the available historical EGCF data, we created ensembles of true measurements (the observations), taking the mean of the observations simulated at every time step. At the same time, we examine the model state (4) propagated at every time steps. An initial ensemble (*a priori* ensemble) at every realisation of the model state is created that forms the ensemble Y in (5). The *a priori* ensemble member is updated to form *a posteriori* ensemble member in (6), which represents the simulated observations. The tabulated *a posteriori* ensembles are finally computed as the $\mathbf{E}[y^u]$, to be comparable with $\mathbf{E}[d_{obs}]$. This allows us to compare the convergence in relation to the observations. 1000 ensemble members are created in this example. The plot with datasets of observations and EnKF mean propagation of EGCF is shown in Fig. 2. The figure shows that the EnKF mean estimations converge towards the observations of the EGCF data.

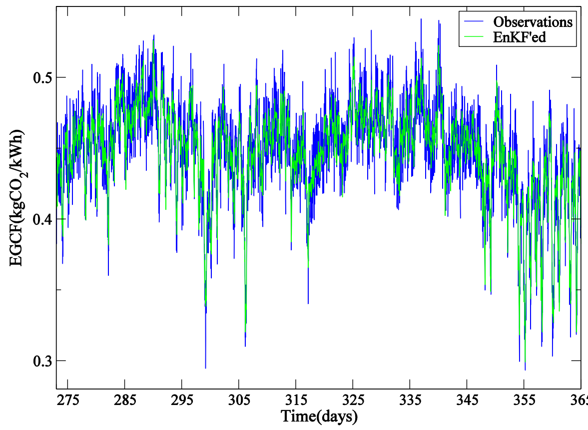

Figure 2. Plot of ensemble mean distributions of 2014 Q4 EGCF vs. the observations.

The RMSE values from Table II indicate that the larger the ensemble size, the smaller the RMSE value, and the better EnKF mean estimations converge towards the observations of EGCF data.

TABLE II. RMSE

Number of ensemble N_e	RMSE
10	0.0757
50	0.0359
100	0.0290
500	0.0224
1000	0.0209

We then map the EnKF forecasted results into the 2014 Q4 EGCF data and the resultant plot is shown in Fig. 3.

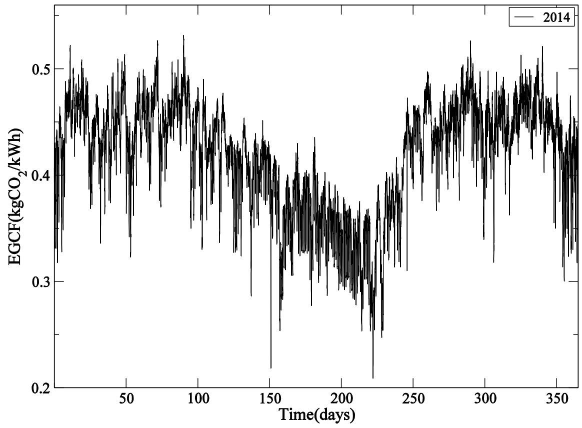

Figure 3. Plot of forecasted 2014 EGCF data.

VI. CONCLUSIONS

This paper presents the application of EnKF in modelling EGCF in the UK. The covariance and correlation coefficient are calculated based on the partly available 2014 EGCF data and the historical EGCF data. A good correlation between the historical and 2014 EGCF confirms that EnKF can be performed in forecasting the EGCF in the remaining period of 2014 Q4. The historical EGCF data has been selected as the priori data for initialization in the simulation of the EnKF model.

EnKF is capable of forecasting and recovering the EGCF from the incomplete fuel data. The ability of EnKF to match the forecast data with observations (historical data) demonstrates the strength of EnKF in estimations of the system state and this can be achieved under the condition that the ensemble size must be sufficiently large.

Although the forecasting or recovering of EGCF using EnKF has been demonstrated in a relatively simple model, the EnKF can be applied in highly nonlinear systems, for instance, in estimation or recovery of individual fuel mixes that balance the electricity market. The EnKF modelling will however become complex and require identifying state variables, initial conditions and prior knowledge in the EnKF model. We suggest that the recovery of EGCF should be taken into account in all aspects of the fuel generations, transmissions and distributions.

REFERENCES

[1] E. T. Lau, Q. Yang, A. B. Forbes, P. Wright, and V. N. Livina, "Modelling carbon emissions in electrical systems," *Energy Conversion and Management*, vol. 80, no. 59, pp. 573-581, 2014.

[2] E. T. Lau, Q. Yang, G. A. Taylor, A. B. Forbes, P. Wright, and V. N. Livina, "Carbon savings in smart interventions in electrical systems," submitted to IEEE Transactions on Power Systems.

[3] E. T. Lau, Q. Yang, G. A. Taylor, A. B. Forbes, P. Wright, and V. N. Livina, "Optimization of carbon emissions in smart grids," in *Proc. Universities' Power Engineering Conference*, 2014, pp. 1-4.

[4] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics," *Geophysical Research*, vol. 99, no. 5, pp. 10143-10162, 1994.

[5] V. E. J. Haugen, G. Nævdal, L. Natvik, G. Evensen, A. M. Berg, and K. M. Flornes, "History matching using the ensemble Kalman

filter on a north sea field case," *SPE Journal*, vol. 13, no. 4, pp. 382-391, 2008.
[6] Elexon. BM reporting. [Online]. Available: http://www.bmreports.com/bsp/bsp_home.htm
[7] J. P. Jensen, "Ensemble Kalman filtering for state and parameter estimation on a reservoir model," Master dissertation, Norwegian University of Science and Technology, 2007.

Eng Tseng Lau was born in Sarawak, Malaysia in 1989 and received his Advanced Diploma in Mechatronics from the Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia in 2011 and MSc in Engineering Management from the Brunel University London, UK in 2011. He is currectly pursuing PhD degree in Manufacturing and Enterprise Engineering, Brunel University London, UK. He is working as the PhD student at the National Physical Laboratory, Teddington, UK. His research interests are power balancing, optimization techniques, mathematical programming and smart grid connected renewable and non-renewable power systems.

Qingping Yang was born in Sichuan, China in 1964 and received his Diploma in Instrumentation and Measurement Technology from Chengdu Aeronautical Polytechnic in 1983 and his PhD in high precision dimensional metrology from Brunel University in 1992.
He first worked for an Aircraft Structural Integrity Research Institute in China before he joined Brunel University as a Visiting Scholar in 1988. After his PhD, he has been working as a Research Fellow and later a Lecturer at Brunel University London.
Dr. Yang has been a member of the IEEE and the IET (Institute of Engineering and Technology). His current research interests include advanced sensors, dimensional metrology, quality engineering, advanced measurement systems, data analysis and smart systems. He has published more than 90 papers in these areas.

Alistair B Forbes was born in Buckie, UK in 1955 and holds a BSc in mathematics from the University of Aberdeen, UK, (1977), an MSc in mathematics from the University of Newcastle, UK (1978) and an MA in mathematics from the University of Pennsylvania, Philadelphia, USA in 1984.
Since joining the National Physical Laboratory (NPL) in 1985, he has worked extensively in the application of mathematics and statistics in measurement. He is currently and NPL Fellow in the Mathematics and Modelling Group, NPL, Teddington, UK.
Prof. Forbes is a member of The Institute of Mathematics and Its Applications, a Chartered Mathematician and a Fellow of the Royal Statistical Society. Since 2000, he has been Visiting Professor of Mathematics and Huddersfield University, UK. In 2014, he received the Alex Hough-Grassby Memorial Award from the Institute of Measurement and Control.

Valerie Livina joined NPL as a Senior Research Scientist in 2012. She has more than 20 years of research and teaching experience in pure mathematics, statistical physics and time series analysis in Russia, Israel and UK. Since 2007, she has developed the so-called 'tipping point toolbox' for early warning, detection and forecast of transitions and bifurcations in time series of dynamical systems. The toolbox combines modified degenerate fingerprinting (monitoring autocorrelations for early warning signals) and potential analysis (detection and forecast of transitional dynamics using stochastic models). The methodology has been applied to a variety of dynamical systems, from paleoclimate to ecology and structure health monitoring.
Dr. Livina has published more than 30 papers in international physical journals and more than 10 papers in Russian mathematical journals. She has reviewed 40+ research papers for 20+ international journals in physics and environmental sciences since 2005. Since 2003, she attended 100+ conferences and workshops (more than 10 as an invited speaker) and 70+ training and developmental courses. She is a Chartered Mathematician of the Institute of Mathematics and its Application (CMath MIMA), Fellow of the Royal Statistical Society and a member of the European Geosciences Union.

