International Journal of Forecasting 30 (2014) 246–256

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

A new error measure for forecasts of household-level, high resolution electrical energy consumption

Stephen Haben[a,c,*], Jonathan Ward[b], Danica Vukadinovic Greetham[a], Colin Singleton[a], Peter Grindrod[a,c]

[a] University of Reading, UK
[b] University of Leeds, UK
[c] University of Oxford, UK

ARTICLE INFO

Keywords:
Verification methods
Load forecasting
Volatile data
Smart meter
Error measure

ABSTRACT

As low carbon technologies become more pervasive, distribution network operators are looking to support the expected changes in the demands on the low voltage networks through the smarter control of storage devices. Accurate forecasts of demand at the individual household-level, or of small aggregations of households, can improve the peak demand reduction brought about through such devices by helping to plan the most appropriate charging and discharging cycles. However, before such methods can be developed, validation measures which can assess the accuracy and usefulness of forecasts of the volatile and noisy household-level demand are required. In this paper we introduce a new forecast verification error measure that reduces the so-called "double penalty" effect, incurred by forecasts whose features are displaced in space or time, compared to traditional point-wise metrics, such as the Mean Absolute Error, and p-norms in general. The measure that we propose is based on finding a restricted permutation of the original forecast that minimises the point-wise error, according to a given metric. We illustrate the advantages of our error measure using half-hourly domestic household electrical energy usage data recorded by smart meters, and discuss the effect of the permutation restriction.

1. Introduction

As many countries progress towards a low carbon economy, the increased penetration of low-carbon technologies (LCTs) may produce new risks to the security and robustness of the electricity networks (Combrink & Vaessen, 2006). The decarbonisation of transport and heating (for instance, through the uptake of electric vehicles and heat pumps) is likely to increase the network demand, whilst household microgeneration increases the prospect of a two-way flow of electricity on the network, as consumers become suppliers and feed back into the grid. In short, electricity demand is likely to increase and become more unstable, particularly at the low voltage (LV) level (Combrink & Vaessen, 2006).

In response to these new challenges, the UK government is aiming to help network operators and suppliers prepare for a low carbon economy through initiatives such as the £500m low carbon network fund (LCNF) (Ofgem, 2012), and the roll-out of smart meters to every home in the UK by 2020 (National Grid, 2012). Smart meters are advanced energy meters with a two-way communication capability which record high resolution (typically half-hourly) energy consumption. These detailed patterns of energy demand provide opportunities to improve our understanding of energy consumption habits, to design smarter interventions for energy reductions, and to produce accurate forecasts of energy demand at the LV level. Such accurate forecasts at the level of households, or small aggregations of households, can help distribution network

* Correspondence to: Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Road, Oxford, OX2 6GG, UK. Tel.: +44 01865 611 511.
E-mail address: haben@maths.ox.ac.uk (S. Haben).

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.08.002

operators improve their management and planning of the LV networks. Forecasts can also be combined with network storage devices to improve peak demand reduction. As part of the New Thames Valley Vision[1] LCNF project, storage devices are being considered to help alleviate the high demand on the LV network at peak times. Simple set point control is the simplest and most common way of controlling battery storage, but often fails to reduce the peak demand (Thomas, 2010). However, accurate household-level forecasts could optimise the use of the battery by helping to plan the appropriate charging and discharging of the storage device (Molderink, Bakker, Bosman, Hurink & Smit, 2010; Xu, Xie, & Singh, 2010). Until recently, the majority of load forecasting has been at the medium voltage (MV) to high voltage (HV) substation levels, where the demand is relatively smooth and more regular (for instance, see the review papers by Alfares & Nazeeruddin, 2002; Moghram & Rahman, 1989; Taylor & Espasa, 2008). However, at the LV network to household level, the demand is volatile and noisy, and typically consists of many different types of behaviour, such as frequent but irregular peaks (Brabec, Konár, Pelikán, & Malý, 2008). Hence, forecasting methods developed for the MV and HV levels may not be appropriate for the household level. In order to produce and test the accuracy of household-level forecast demands, appropriate forecast verification methods are required.

Forecast verification hinges on the ability of quantitative measures to assess the similarities between forecasts and observations, what Murphy (1993) refers to as forecast *quality*. Hence, measure-orientated approaches based on point-wise comparisons, such as the mean absolute error (MAE) and root mean square error (RMSE), can often lead to spurious conclusions, see Brooks and Doswell III (1996); Castati et al. (2008), and Hoffman, Liu, Louis, and Grassotti (1995). In particular, an observed feature that is forecasted accurately in terms of size and amplitude, but displaced in time, incurs a "double penalty" (Keil & Craig, 2009). Thus, as we illustrate in this paper, it can be difficult for skilled, plausible forecasts to out-perform even a flat forecast that is of almost no informative value, particularly when the data are volatile and noisy. This problem has long been understood in the meteorology community. Consequently, a large number of alternative verification strategies have been proposed; see Castati et al. (2008) for a review. The class of distribution-oriented approaches (Brooks & Doswell III, 1996; Murphy & Winkler, 1987) offers many insights but requires large quantities of data and is computationally intensive (Brooks & Doswell III, 1996).

One approach to the calculation of displacement errors, which was also pioneered in meteorology, has been to formulate errors using an optimal distortion of the original field, i.e., *smooth* changes in position and amplitude that minimise the misfit between the data and the forecasts (Hoffman et al., 1995). Although such verification methods have been developed widely, they have limited appeal in the setting in which we are interested primarily—volatile, noisy and irregular data. In this case, it may be more appropriate to use verification measures that deform the forecast *discontinuously*. To some extent, such techniques are employed in 'fuzzy' verification techniques for high-resolution weather forecasting (Ebert, 2008). These typically compare the average states of 'events' occurring within a neighbourhood of interest. For real-valued variables, such as the amount of rainfall or the wind intensity, events are defined relative to some threshold. In essence, these methods produce new fields for both the observed and forecasted data, which are then compared using a traditional point-wise metric. Such measures are both scale and threshold dependent, and thus, one must consider a matrix of errors that captures both of these variations.

Many algorithms and metrics have been developed for measuring the similarity of time series, such as Dynamic Time Warping (DTW), longest common subsequence, edit distance on real sequences, and edit distance with real penalty (Chen & Ng, 2004). Often, these algorithms are applied in information retrieval and data mining techniques in order to measure the cost of morphing one time series into another. Dynamic Time Warping is one of the most popular techniques for measuring time series similarity, and has been used successfully in automatic speech recognition algorithms (Muller, 2007). DTW measures the differences between sequences which may vary in time or speed by stretching the time series through the duplication of local points. The difference in the deformed time series is then calculated using a standard L_p metric. A more recent method, called the Move-Split-Merge (MSM) metric, is similar to DTW, except that duplicated and deleted values incur a fixed cost (Stefan, Athitsos, & Das, 2013). For time series matching methods, although suitable for comparing series with the same (but perhaps stretched) shape in time, they are biased toward preserving ordering, and therefore are not flexible enough, in the context of energy demand, to cope with the natural irregularities in household energy usage behaviour. In addition, DTW and MSM will tend to underestimate the costs of repeated peaks by simply merging/duplicating the local peaks, with little or no penalty incurred for the inaccurate repetition. The additional complications and restrictions introduced by such techniques make them unsuitable for measuring the errors of household-level forecasts. This motivates the development of a new forecast error measure, which is the topic of this paper.

Before sophisticated forecasting techniques for household electrical energy usage can be developed, we need to be able to assess their veracity against data quantitatively. However, in this paper we illustrate the fact that the capricious nature of energy usage means that traditional point-wise measure-oriented approaches perform poorly at this task. Our main contribution is to suggest a new approach that allows for some flexibility in the timing of the forecast when computing the error, while retaining some simplicity. Specifically, for each forecast we define the error to be the minimum error (with respect to an appropriate norm) over the set of all restricted spatial/temporal permutations of the forecast. We begin in Section 2 with a formal description of point-wise error measures, particularly the p-norm,

[1] http://www.thamesvalleyvision.co.uk/.