

Department of Mathematics and Statistics

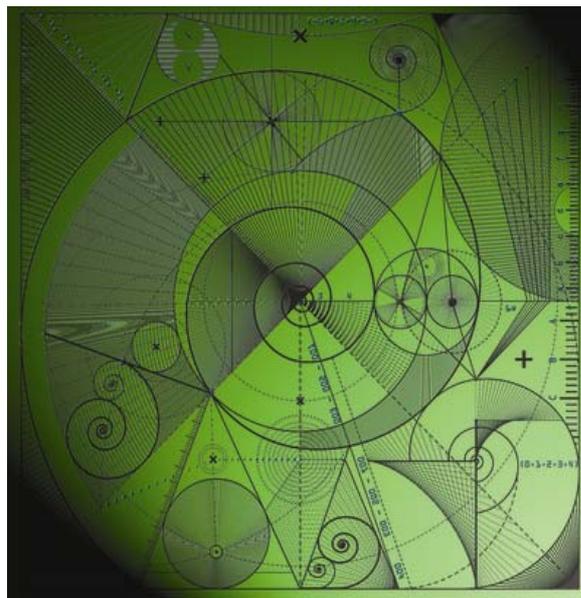
Preprint MPS-2012-14

6 September 2012

A New Error Measure for Forecasts of Household-level, High Resolution Electrical Energy Consumption

by

Stephen Haben, Jonathan A. Ward, Danica
Vukadinovic-Greetham, Peter Grindrod and
Colin Singleton



A New Error Measure for Forecasts of Household-level, High Resolution Electrical Energy Consumption

Stephen Haben¹, Jonathan A. Ward¹, Danica Vukadinovic Greetham¹,
Peter Grindrod¹, and Colin Singleton²

¹Centre for the Mathematics of Human Behaviour (CMoHB), Department of Mathematics and
Statistics, University of Reading

²CountingLab Ltd

September 6, 2012

Abstract

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 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

Forecast verification hinges on the ability of quantitative measures to assess the similarity between forecasts and observations, what Murphy refers to as forecast *quality* [15]. Hence measure-orientated approaches based on point-wise comparisons, such as mean absolute error (MAE) and root mean square error (RMSE), can often lead to spurious conclusions [3, 4, 9]. In particular, an observed feature that is forecasted accurately in terms of size and amplitude but displaced in time, incurs a “double penalty” [11]. Thus, as we illustrate in this paper, it can be difficult for skilled forecasts to out perform even a flat forecast that provides almost no informative value, particularly when the data is 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 [4] for a review. The class of distribution-orientated approaches [14, 3] offer many insights but require large quantities of data and increased computational effort [3]. The additional complications introduced by such techniques may often be unnecessary, particularly in situations where skilled benchmark forecasts are yet to be established. This motivates the development of improved forecast error measures, the topic of this paper.

One approach for managing the displacement errors, 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 data and forecast [9]. Although such verification methods have been widely developed, they have limited appeal in the setting that we are primarily interested — 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 [6]. These typically compare the average state of ‘events’ occurring within a neighbourhood of interest. For real-valued variables, such as the amount of rainfall or 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, thus one must consider a matrix of errors that captures both of these variations.

Our interest stems from the context of fine-grained domestic electrical energy usage forecasting. To date, data concerning household load profiles has been almost exclusively aggregated over large numbers of houses [17] and demand side management has benefited from a range of forecasting techniques [12, 1]. However, the UK smart meter roll out planned to start in 2014 will pave the way towards a radically different, data-rich energy sector. High-resolution domestic household electrical energy usage data is volatile, noisy and typically consists of many different types of behaviour, containing frequent but irregular peaks. The responsibility of reducing carbon emissions related to domestic energy use rests on the energy suppliers [8] and smoothing out peaks in demand is critical to success. Potential widespread adoption of low carbon technologies, such as electric vehicles, photo-voltaics and ground source heat pumps means that there is also a need for energy companies to understand future loading in order to reduce strain on the network. Forecasting household electrical energy usage could help to mollify peak demand, e.g. via the use of battery storage to efficiently manage supply from domestic micro-generation [20]. In addition, using forecasts to better inform households could lead to improved usage behaviour, particular in the presence of time of day tariffs. Ultimately, forecasting domestic electrical energy usage at the micro-scale will assist energy suppliers to achieve their carbon emissions targets.

Before sophisticated forecasting techniques for household electrical energy usage can be developed, we need to be able to quantitatively assess their veracity against data. However, we illustrate in this paper that the capricious nature of energy usage means that traditional point-wise measure-orientated 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. 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. This method can be applied to almost any type of error measure and has broad scope for application in other areas such as finance. We begin in Section 2 with a formal description of point-wise error measures, particularly the p -norm, we then introduce the “adjusted p -norm” and illustrate its advantages using a simple, synthetic example. In Section 3, we apply our new metric to measure the accuracy of a hierarchy of daily forecasts of half-hourly electrical usage taken from individual household smart meter data. In Section 4, we present a detailed discussion of the effect of the ‘adjustment limit’, i.e. the maximum allowed permutation displacement. Finally, we draw conclusions and discuss the advantages and disadvantages of our method in Section 5.

2 Measuring Errors

2.1 Standard Error Estimates: The p -Norm

Let $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ and $\mathbf{f} = (f_1, f_2, \dots, f_n)^T$ be the actual and forecasted data vectors respectively, such that each f_i is a prediction of the actual data x_i for $i = 1, \dots, n$. We focus on one-dimensional data (i.e. time-series), however the methods that we describe can be generalised to higher dimensions. Error measures can be described in terms of a vector function

$$E = F(\mathbf{f}, \mathbf{x}), \quad (2.1)$$

where $F : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$. In this paper we focus on the absolute p -norm,

$$E_p = \|\mathbf{f} - \mathbf{x}\|_p = \left(\sum_{i=1}^n |f_i - x_i|^p \right)^{1/p}, \quad (2.2)$$

for some $p \geq 1$ (See [7]). For example, this type of error includes the Mean Absolute Error (MAE) and the root mean square (RMS) error, which are simply constant multiples of the 1-norm and 2-norm errors respectively.

2.2 The Adjusted Error

In the applications that we are interested in, it is much more important to predict peaks at approximately the right times rather than not at all. As stated in Section 1, such forecasts incur a double penalty from point-wise error measures and may be judged incorrectly as poor forecasts. This motivates the idea that the error measure should allow for small, possibly discontinuous, displacements in time of the forecast values. We note that there exist many perfect matchings between the forecast values and actuals. Each match can be described by a permutation matrix P . To restrict the magnitude of the displacements of the forecast values, we impose an ‘adjustment limit’, denoted $w \geq 0$, on the permutations such that $P_{ij} = 0$ for $|i - j| > w$. We define the *adjusted error* to be the solution to the minimisation

$$E^w = \min_{P \in \mathcal{P}} F(P\mathbf{f}, \mathbf{x}), \quad (2.3)$$

for the given error measure F , where \mathcal{P} is the complete set of restricted permutations. The adjusted p -norm is then

$$E_p^w = \min_{P \in \mathcal{P}} \|P\mathbf{f} - \mathbf{x}\|_p. \quad (2.4)$$

The error minimisation is a variant of the assignment problem, a well-known combinatorial optimisation problem that can be solved in polynomial time [13] using the ‘Hungarian method’, details of which can be found in [18]. To incorporate the adjustment limit into the algorithm, if $|i - j| > w$ then we set $|f_i - x_j|^p = \Omega$, where Ω is a large constant that prevents such high error matches. The methods time complexity is $O(n(m + n \log n))$ [19] where m is the number of potential error matches, $|f_i - x_j|^p$ for $i, j = 1, \dots, n$.

The adjustment limit w is a time-scale parameter that is problem dependent and has an important effect on the efficacy of our verification method. If $w = 0$ then we recover the original p -norm (2.2). Increasing w reduces the adjusted error, but a small error resulting from large displacements is not necessarily indicative of a good forecast. Thus the mean displacement, which can be obtained from the permutation matrix P , is an additional measure of accuracy that can be used to compare different forecasts. We discuss these points in detail in Section 4.

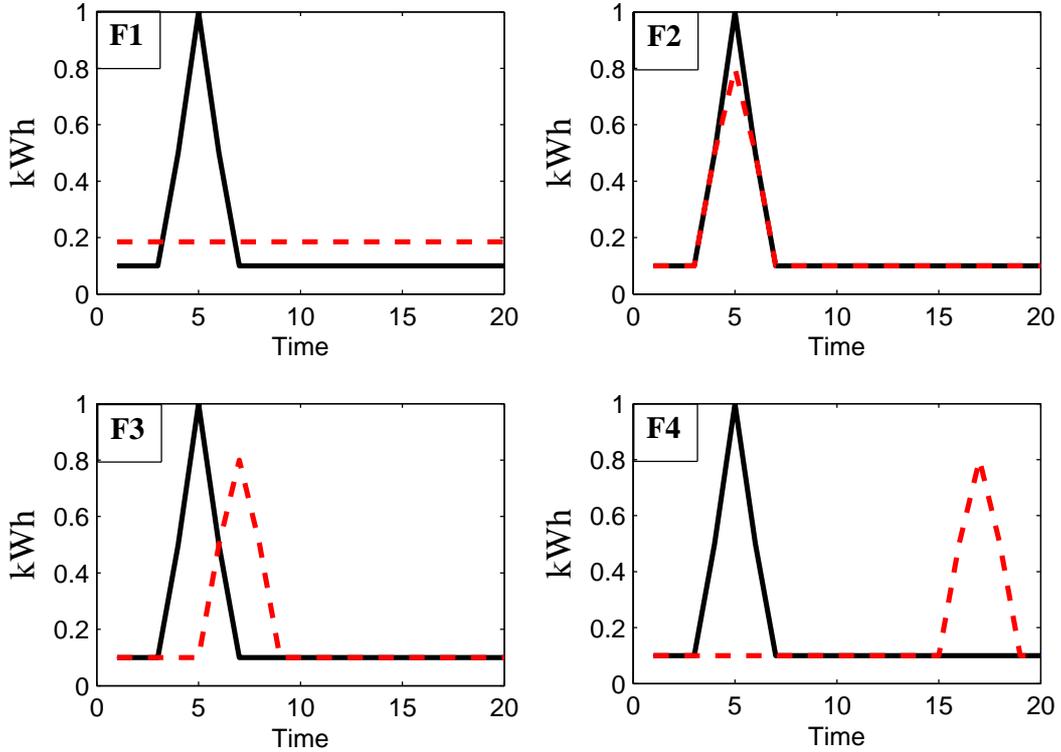


Figure 1: Four ‘Forecasts’ F1, F2, F3, F4 (Dashed line) together with the actual data (Solid Line) for a simplified example.

2.3 Simple Example

In this subsection we compare four qualitatively different forecasts of a simple energy load profile using the absolute and adjusted p -norm errors. The synthetic data, illustrated with solid black lines in each panel of Figure 1, consists of a single peak centred around $t = 5$ with a constant background usage over a 20 time point domain. The forecasts, illustrated with dashed lines, consist of a flat forecast (F1) (corresponding to the average usage) and a single peak centred around three different times (F2–F4) with the correct background usage. Subjectively, F2 is a very good forecast, F3 is reasonable and both F1 and F4 are poor.

The absolute and adjusted p -norm errors, with $p = 4$, for each of the forecasts illustrated in Figure 1 are presented in Table 1. We have used the 4-norm error, rather than the more common 2-norm, because we want to penalise large errors (i.e. missed peaks) much more than small errors. Different values of p yield qualitatively similar results. Table 1 illustrates the following:

- **Absolute 4-norm.** The good forecast F2 has the smallest error while the flat forecast F1 has smaller error than both the poor forecast F4 *and* the reasonable forecast F3. This illustrates the double penalty effect present in point-wise error measures.

Error	Forecast			
	F1	F2	F3	F4
Absolute Error	0.824	0.200	0.987	1.00
Adjusted Error ($w = 1$)	0.824	0.200	0.785	1.00
Adjusted Error ($w = 2$)	0.824	0.200	0.479	1.00
Adjusted Error ($w = 3$)	0.824	0.200	0.200	1.00

Table 1: Comparison of the error measurements given by the different norms for the 4 different forecasts F1–F2 described in the main text.

- **Adjusted 4-norm, $w = 1$ and $w = 2$.** The reasonable forecast F3 error is reduced to about 95 % and 58% of the F1 flat forecast error for adjustment limit $w = 1$ and $w = 2$ respectively. The F1, F2 and F4 forecast errors are the same for both the adjusted 4-norm and the absolute 4-norm—displacing the forecast values does not change the errors.
- **Adjusted 4-norm, $w = 3$.** The good F2 forecast and the reasonable forecast F3 errors are equal. However, we can still distinguish F2 as the better forecast with this method by considering the mean displacement. F2 has zero mean displacement of the forecast values (implying the minimum permutation is achieved by the forecast) whereas F3 has a mean displacement of 0.6 grid points over the 20 forecasted values.

In summary, the synthetic example illustrates how the adjusted norm can give a more accurate representation of the forecast accuracy than the standard 4-norms.

3 Application to Household Energy Load Forecasting

In this section we use the adjusted 4-norm to compare the performance of three forecasting methods applied to half-hourly domestic household electrical energy usage data. The data was collected by household smart meters as part of the Ofgem managed Energy Demand Research Project (EDRP) trial run by Scottish and Southern Energy (SSE)¹. A wide variety of energy usage behaviours are observed between households and individual household demand is both volatile and noisy. However there are daily, weekly and seasonal patterns that could potentially be exploited by forecasting methods. Such forecasts can have a positive impact on network operations and planning. Figures 2(a)–(c) illustrate a week’s worth of half hourly electrical energy usage profiles in kilowatt-hours (kWh) for three representative UK households. Household (a) consumes most of their energy during one or two peak periods at regular daily intervals. Thus we would hope to be able to forecast their usage fairly accurately. Household (b) has irregular peak demands that are smaller than the other households, but they maintain a fairly constant background usage. Household (c) is the most volatile, having large irregular peak demands and periods of low usage. We would expect this household’s energy usage to be difficult to forecast. The average daily energy usage for households (a), (b) and (c) is 5.51 kWh, 9.89 kWh and 18.12 kWh respectively.

¹See <http://www.ofgem.gov.uk/sustainability/edrp/Pages/EDRP.aspx> for further details

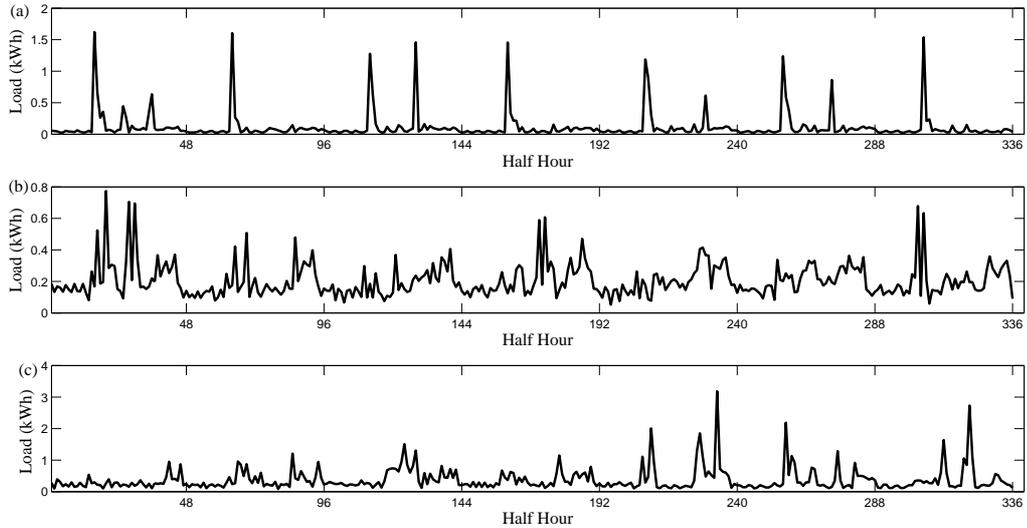


Figure 2: Example of half hourly smart meter electrical energy usage (in kWh) for household (a), household (b) and household (c) as described in the main text.

Our household energy usage data-set consists of 10 weeks of half-hourly kWh records (3360 in total) for each of the three households. Each forecast generates an unsupervised rolling daily prediction from midnight over the course of the 10th week with access to the full data-history of each household separately. Our aim is to assess validation techniques and consequently the forecast methods that we implement are chosen to form a clear hierarchy. The three methods by which each of the daily forecasts are generated are as follows:

1. **Flat** forecast: The average usage over the previous 7 days, used as the forecast for all time periods.
2. **Last Week (LW)** forecast: The usage on the same day of the previous week.
3. **Averaged Adjustment (AA)** forecast: A combination of a historic average and baseline usage. A detailed description can be found in Appendix A.

A snap shot of a single day's data from each household and the corresponding forecasts are illustrated in Figure 3.

Clearly the flat forecast provides little informative value, while the LW forecast is innately realistic but performs poorly for irregularities in week to week behaviour. The AA forecast is subjectively better than the other forecasts but volatility still reduces its performance.

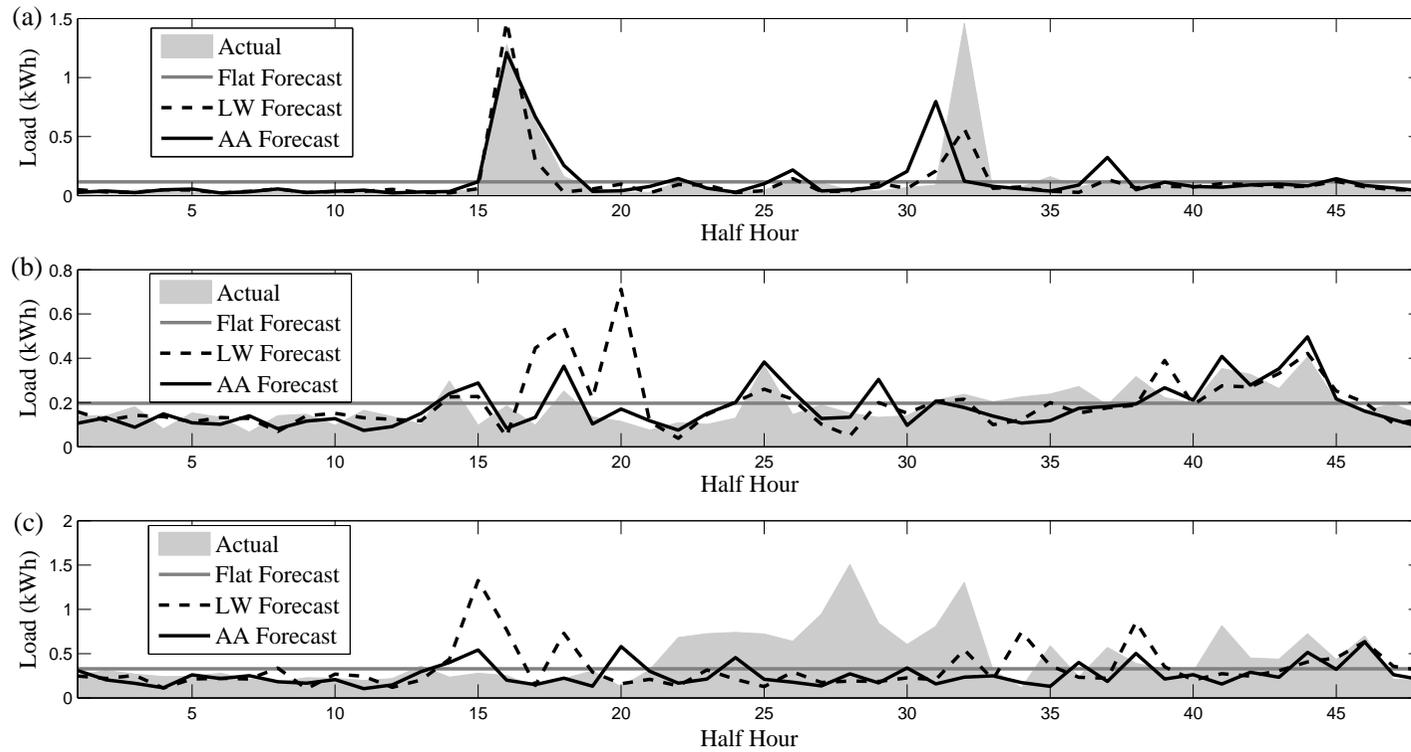


Figure 3: Forecasted usage on Wednesday of the final week of the data set of each household described in text. Plot shows actual usage (shaded area) together with forecasts for household (a), (b) and (c) using the AA (black line), LW (dashed line) and Flat (gray line) forecast methods.

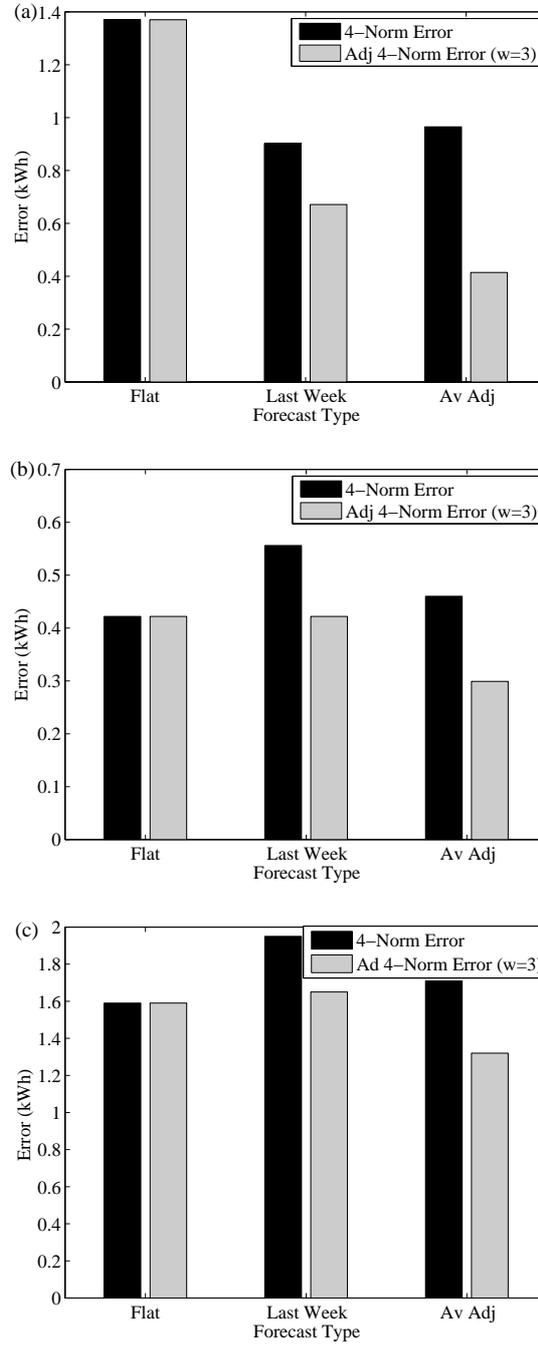


Figure 4: Panels (a)–(c) correspond to households (a)–(c) respectively. Each panel depicts the daily averages of the 4-norm (Black) and adjusted 4-norm (Gray) errors for the three forecasts.

As in the simple example described in Section 2.3, we compare the absolute and adjusted p-norms with $p = 4$ in order to penalise larger peaks to a greater extent than smaller peaks. We use $w = 3$ as the adjustment limit, hence forecasts can be displaced up to one and a half hours *either side* of their original forecast time. The effects of w are considered in more detail in Section 4. Because the forecasts produce rolling daily predictions, we calculate the i^{th} day’s errors for each measure, e_i , and then use the mean absolute error,

$$\langle E \rangle = \frac{1}{7} \sum_{i=1}^7 e_i, \quad (3.5)$$

to compare forecasts.

The daily mean errors of each forecast method are shown in Figures 4(a)–(c) for each household respectively. The black bars show the daily-mean 4-norm error and the gray bars show the daily-mean adjusted 4-norm error. Focusing first on the 4-norm errors, we note that the flat forecast out-performs the other forecasts for both households (b) and (c). Additionally, the AA forecast is beaten by the LW forecast for household (a). Clearly these results do not agree with the proposed forecasting hierarchy. In particular, we know that the flat forecast reproduces none of the daily household usage patterns. By ignoring peaks altogether, the flat forecast avoids the double penalty and can appear better than more sophisticated forecasts.

We now consider the 4-norm adjusted errors, illustrated with gray bars in panels (a)–(c) of Figure 4. We note that the adjusted norm does not change the flat forecast errors, but reduces all of the LW and AA errors. The AA forecast is now the most successful forecast for all households with a marked improvement for household (a) in particular. This can be attributed to the regular peak demands observed in the data being forecasted close to when they actually occur and the absence of the double penalty in the adjusted norm. Relative to the flat forecast errors, the improvement in the errors for the AA forecast decreases from households (a) through to (c), owing to the relative increase in volatility respectively. The magnitude of the errors for household (c) are by far the largest and the relative difference between methods is the smallest, indicating that forecast sophistication only introduces marginal relative improvements as volatility increases.

To illustrate that our results hold more generally, we consider the 4-norm and adjusted 4-norm errors of the three forecast methods applied to the usage data of 600 individual domestic households. As in the example above the data set for each household consists of half hourly electrical energy usage over a 10 week period, collected by smart meters during the EDRP trial. Using the Flat, LW and AA methods, a rolling daily forecast of each household’s energy usage was produced for the final week of each data set. Figure 5 shows the difference between the flat forecast errors and the 4-norm and adjusted errors (with $w = 3$) for both the LW and AA forecasts. The horizontal-axis represents the difference between the 4-norm errors of the Flat forecast and the LW or AA forecast and the vertical-axis represents the difference between the adjusted 4-norm errors of the Flat forecast and the LW or AA forecast. The diagonal line indicates where the 4-norm and deformed 4-norm errors are equal. Since the adjusted 4-norm is always smaller than the 4-norm, no forecasts can occupy the area below the line.

The three occupied quadrants of the graph establish a 3 cluster segmentation of the forecasts in terms of their accuracy:

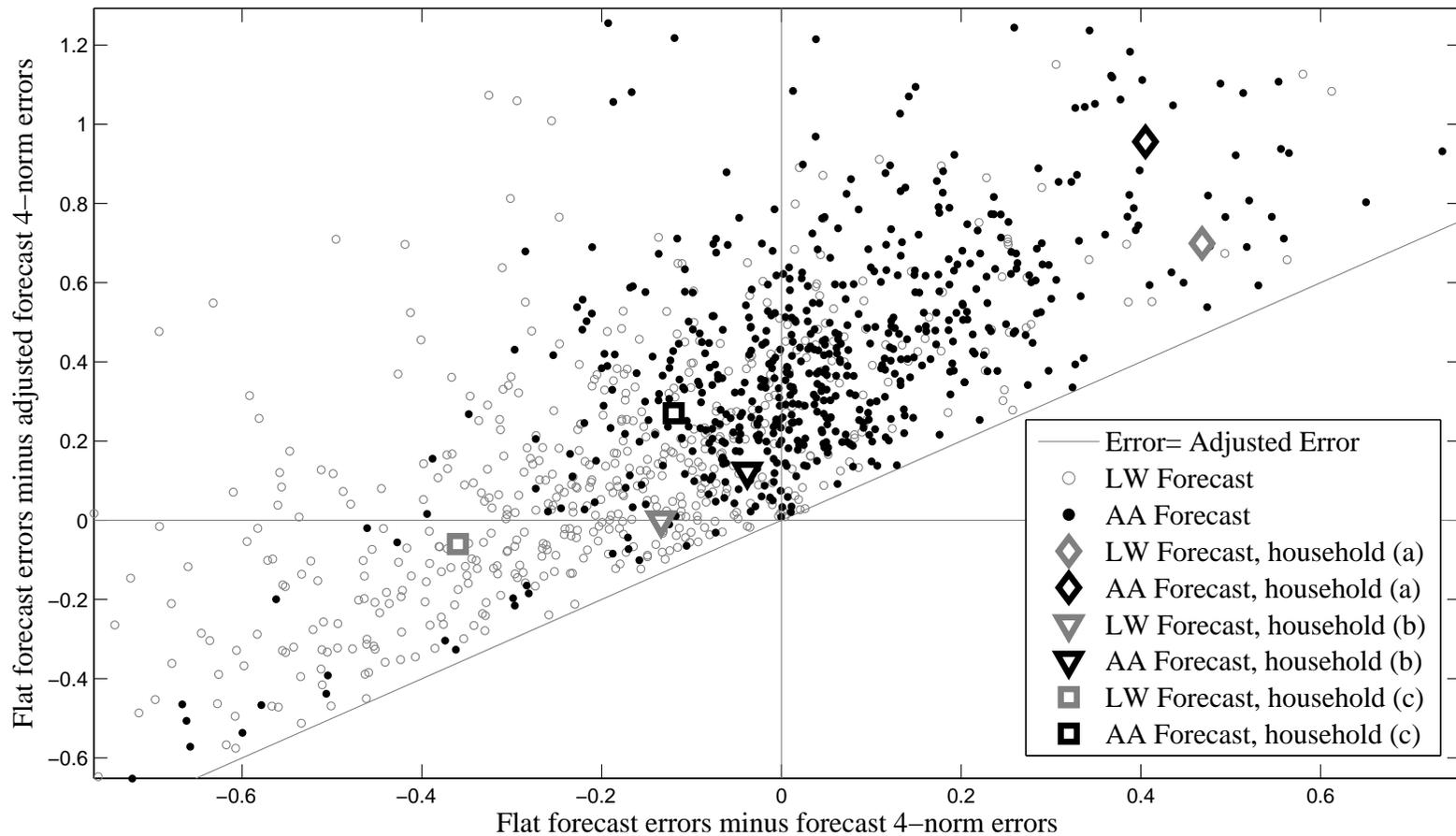


Figure 5: Difference in 4-norm forecast errors of the flat and LW (unfilled circles) or AA (filled) forecasts versus Difference in 4-norm adjusted forecast errors of the flat and LW (unfilled circles) or AA (filled) forecasts. Also included are the data for the Households (a) (Diamonds), (b) (Triangles) and (c) (Squares).

1. Points in the lower-left quadrant represent forecasts whose 4-norm and adjusted 4-norm errors are larger than or equal to the flat forecast errors. We refer to these forecasts as *Poor*.
2. Points in the upper-left quadrant represent forecasts whose flat forecast error is smaller than the 4-norm error forecast but larger than the adjusted 4-norm error. Since the small temporal re-alignment has reduced the error compared to the 4-norm error we refer to these forecasts as *Good after adjustment*.
3. Points in the top right quadrant represent forecasts whose 4-norm and adjusted 4-norm errors are smaller than the flat forecast errors. We refer to these as *Good* forecasts.

The plot shows that the AA forecasts (filled circles) are in general superior to the LW forecasts (unfilled circles). The majority of the AA forecast are either good (360) or good after adjustment (208). Only 32 of the AA forecast are poor whereas 225 of the LW forecasts are poor. For the LW method, only 105 are good forecasts and just less than half (270) are good after adjustments. Of the 600 households, the LW forecast only out-performs the AA forecasts for 30 households in the 4-norm but for 46 households in the adjusted 4-norm. In Figure 5 we also include the data for the LW and AA forecasts of households (a), (b) and (c). In terms of our accuracy classification both the LW and AA are good forecasts for household (a) whereas the AA forecast is good after adjustment for households (b) and (c) while the LW forecast is poor for households (b) and (c). The large proportion of forecasts that are good after adjustment are particularly important. If only the 4-norm is used as an accuracy measure then these forecast methods could potentially be mistakenly rejected, despite their improved performance with respect to the adjusted norm.

4 The adjustment limit

In this section we analyse the adjusted error in more detail. The choice of the adjustment window, w , is largely subjective and application specific. In section 3 we chose $w = 3$ based on the assumption that a reasonable forecast of household electrical energy usage should only misplace a peak by a maximum of an hour and a half. Other criteria, such as requiring the forecast to out-perform the flat forecast, can also be used to inform on a suitable adjustment limit. For a given application it may be necessary to consider the error as a function of w , as described in this section, in order to make a more informed decision on the size of the adjustment window.

Figure 6 displays the adjusted 4-norm error for each of the households introduced in Section 3 for the AA and LW forecast for different values of w , illustrated in panels (a) and (b) respectively. Each curve is a monotonically decreasing function of the adjustment limit. The black markers on the graph of each line shows where the forecast error equals the error of the flat forecast (The forecasts for household (a) have smaller errors than the flat forecast in these examples hence the absence of a marker). For all households, in order to outperform the flat forecast the AA forecast must use $w = 1$, whereas the LW forecast must use $w = 4$. As we increase w , large reductions in the adjusted error indicate that large peaks in the forecast are being matched to the actuals. We focus on the AA forecast for our analysis, similar results hold for the LW forecast. As we increase w from 0 to 2 there are large decreases in the adjusted error of the forecast for household (a) due to the closeness (within 3 half hours) of the peaks in the forecast and actual

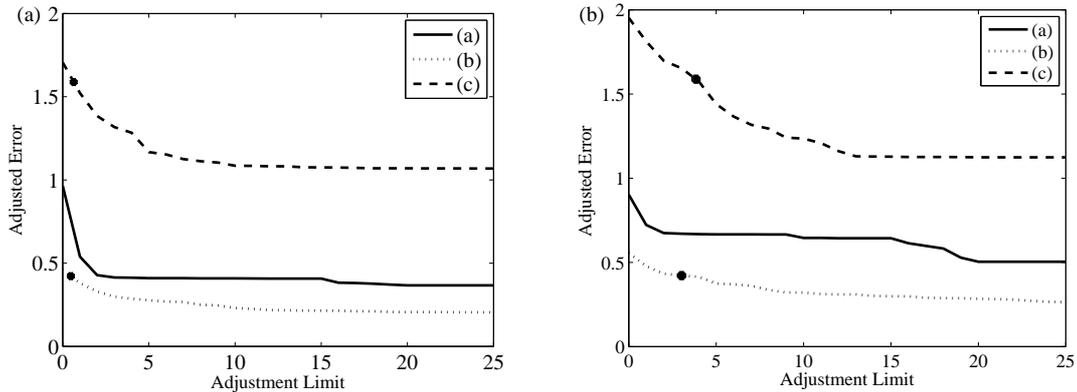


Figure 6: The adjusted errors for (a) the AA forecast and (b) the LW forecast for the usage of three different households (a) (solid line), (b) (dotted) and (c) (dashed) as a function of w . The black marker on each line shows where the forecast errors equal the errors of the flat forecast.

usage. Moderate decreases in the forecast errors are also observed for household (c), although even with $w = 20$ the errors are relatively large compared with the errors in the forecasts for households (a) and (b). Household (b) has a slow rate of reduction as w increases. As shown in Section 3, the general behaviour of household (b) can be forecasted accurately and so the slow reduction is likely to be due to the matching of the small daily irregularities.

The adjusted error decreases with increasing w but this is likely to simultaneously increase the mean displacement of the forecast positions. Smaller displacements are more desirable as they indicate a closer proximity of the features of the forecast with the actuals. To fully describe the accuracy of a forecast we must consider both the mean displacement and the adjusted error of the forecast. As shown for the synthetic example in Section 2.3, the mean displacement can be used to distinguish between the accuracy of two forecasts with the same adjusted error. Since we are primarily interested in the displacement of the peak loads, we consider a weighted mean displacement. Suppose that the forecast at point i , f_i is matched to the actual at j , and $d_i = |i - j|$ is the forecast displacement then we define the average displacement for each day as

$$\hat{D} = \frac{\sum_{i=1}^{48} f_i^4 d_i}{\sum f_i^4}. \quad (4.6)$$

The power of 4 ensures that our measure is representative of larger peaks.

Figure 7 shows the mean displacement of the AA and LW forecasts over the final week as a function of w for each household, together with a plot of the expected average displacement if the forecast was assigned randomly. (The random displacement is found by calculating the expected displacement within the adjustment limit assuming any displacement is equally likely. The mean over the 48 daily points is then calculated.) We present the results for the AA forecast, the LW forecast results are similar. The mean displacements of the forecasts for households (b) and (c) closely match the random displacement curve when $w < 10$. It is likely that the features of the forecasts are being matched to the irregular week to week behaviours of the households.

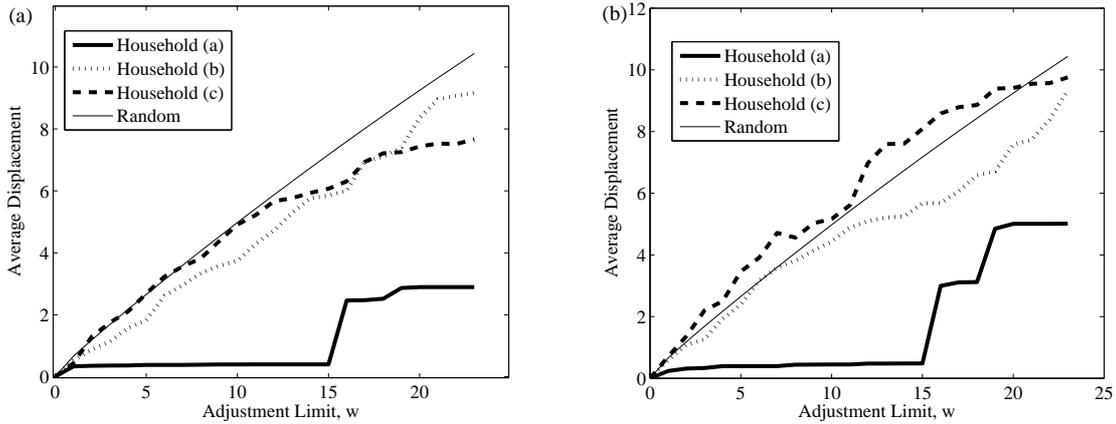


Figure 7: Average weighted displacement of forecasted points for different adjustment limits for (a) the AA forecast and (b) the LW forecast for the three different households (a) (solid line), (b) (dotted) and (c) (dashed). Also included is the expected displacement if forecast points were assigned randomly.

As we showed in Section 3, the regular behaviour of household (b) is accurately forecasted but the small irregular demands are poorly forecasted. Household (c) has no regular week to week behaviour and is largely unpredictable. In contrast, household (a) has regular weekly behaviour and thus the peaks are accurately forecasted and therefore the mean displacement remains small for all w values. As the adjustment limit is increased beyond $w = 15$ some of the afternoon and morning peaks are matched resulting in the small increase in the size of the average displacement.

Figures 6 and 7 together reveal extra information about the usage patterns and forecast accuracy for each of the different households. In particular, for household (a), sharp drops in the forecast error as w is increased from 0 to 2 indicate the forecast closely approximates the large features in the data. The small average displacements confirm that regular peaks are being matched. In contrast, for household (c) the large reduction in forecast error is likely to be the result of matching random, irregular behaviour as shown by the mean displacement being similar to a random assignment in Figure 7. Similarly we find that the small reduction in the adjusted error for household (b) as we increase w are mainly the consequence of matching the small irregular behaviour which are missed by the forecast.

This section has shown how considering both the adjusted and displacement errors as a function of w can reveal properties of the forecasts and their accuracy in predicting large features in the data.

5 Conclusions

In this report we suggest an alternative metric for measuring the success of forecasts of volatile and noisy data. A standard treatment of the accuracy is to consider the p-norm of the error, but due to the “double penalty” effect such measures have been shown to be inadequate, especially

when attempting to forecast peaks and troughs in the data. Any successful forecast method requires a degree of flexibility. Our proposed solution, the adjusted p -norm, allows for limited permutations of the forecasted data, which reduces the penalty imposed on shifted peaks. This was illustrated with a simple synthetic example and then demonstrated on forecasts of real, high resolution household electrical energy usage.

To test the forecast measure, three forecast methods were applied to three different households usage data with varying degrees of week to week regularity and hence forecastability. The forecasts ranged in skill with a clear hierarchy: an innately poor flat forecast, a poor, yet realistic ‘last week as this week’ forecast and an intelligent-average of previous weeks behaviour method. We found that the flat forecast could outperform many of the more realistic, informative forecasts, using point-wise metrics. This was not the case with our new metric. We also applied the measure to forecasts of 600 independent households which confirmed the ability of the new measure to successfully distinguish between the accuracy of each of the 3 forecasts. In addition, we also considered the effect of changing w on the adjusted error and the average displacement of the matched forecasts which offered further insights into the accuracy of the forecasts.

The new metric deforms the forecast in a discontinuous way which may not be appropriate for all applications. However when the data is volatile and irregular the smoothness of the deformation may be less significant. Additionally, for any particular application the method can be applied to any suitable error measure and is very simple to implement with only 1 controllable parameter, w . In summary, in this paper we have presented a new method for verifying forecast accuracy which has shown to be effective and efficient for dealing with shifted features of volatile and noisy data sets.

Acknowledgments We wish to thank Scottish and Southern Energy Power Distribution (SSEPD) for their support and for the funding of this project via the Ofgem ‘Innovation Funding Incentive’². We also wish to thank Scottish and Southern Energy (SSE) for providing the EDRP data for use in this project. JAW also acknowledges the EPSRC for support of MOLTEN (EP/I016058/1).

A The Averaged Adjustment Forecast

In this section we briefly describe the Averaged Adjustment (AA) forecast as implemented in this report. For clarity, we show how we forecast for one particular day, the other days of the week are forecasted in an analogous way. We assume that we have N daily usage profiles of half hourly resolution of the d^{th} day of the week ($d = 1, \dots, 7$) which we notate $\mathbf{G}^{(k)} = (g_1^{(k)}, g_2^{(k)}, \dots, g_{48}^{(k)})^T$ for $k = 1, 2, \dots, N$, where $\mathbf{G}^{(1)}$ is the previous week usage of the d^{th} day and $\mathbf{G}^{(2)}$ is the usage over the d^{th} from 2 weeks before etc. We create a baseline profile $\mathbf{F}^{(1)} = (f_1^{(1)}, f_2^{(1)}, \dots, f_{48}^{(1)})^T$ where each half hour is defined to be the median value over all N half hours. We iteratively update the baseline profile using matching with each successive previous weeks data. This is performed as follows. Suppose $\mathbf{F}^{(k)}$ is the current baseline for the k^{th} iteration ($1 \leq k \leq N - 1$). We define $\hat{\mathbf{G}}^{(k)} = \hat{P}\mathbf{G}^{(k)}$, where $\hat{P} \in \mathcal{P}$ is a permutation matrix such that

$$\|\hat{P}\mathbf{G}^{(k)} - \mathbf{F}^{(k)}\|_4 = \min_{P \in \mathcal{P}} \|P\mathbf{G}^{(k)} - \mathbf{F}^{(k)}\|_4. \quad (\text{A.7})$$

²See <http://www.ofgem.gov.uk/Networks/Techn/NetwrkSupp/Innovat/ifi/Pages/ifi.aspx> for Innovation Funding Incentive reports.

where \mathcal{P} represents the set of restricted permutations of the half hour loads (i.e. each half hour i moved to some half hour j where $|i - j| \leq w$ and w is the deformation limit as described in Section 2.2). In other words, $\hat{\mathbf{G}}^{(k)}$ is the usage from the previous week that minimises the deformed norm error between the baseline load usage and the usage of the current week $\mathbf{G}^{(k)}$. The new baseline is defined to be

$$\mathbf{F}^{(k+1)} = \frac{1}{k+1} (\hat{\mathbf{G}}^{(k)} + k\mathbf{F}^{(k)}). \quad (\text{A.8})$$

The process is repeated for each of the remaining weeks to give the final forecast $\mathbf{F}^{(N)}$. Hence, the forecast is defined to be an average of the initial baseline and permutations of the previous weeks

$$\mathbf{F}^{(N)} = \frac{1}{N+1} \left(\sum_{k=1}^N \hat{\mathbf{G}}^{(k)} + \mathbf{F}^{(1)} \right). \quad (\text{A.9})$$

References

- [1] H.K. Alfares and M. Nazeeruddin, *Electric load forecasting: literature survey and classification of methods*, Int. J. Syst. Sci., **33**(1), 2002, pp.23–34.
- [2] H.M. Al-Hamadi and S.A. Soliman, *Short-term load forecasting based on Kalman filtering with moving window weather and load model*, Electrical Power Systems Research, **68**, 2004, pp.47–59.
- [3] H.E. Brooks and C.A. Doswell III, *A Comparison of Measures-Oriented and Distributions-Oriented Approaches to Forecast Verification*, Wea. Forecasting, **11**, 1996, pp.288–303.
- [4] B. Castati, L.J. Wilson, D.B. Stephenson, P. Nurmi, A. Ghelli, M. Pocernich, U. Damrath, E.E. Ebert, B.G. Brown and S. Mason, *Review Forecast verification: current status and future directions*, Meteorological applications, **15**, 2008, pp. 3–18.
- [5] E.E. Ebery and J. L. McBride, *Verification of precipitation in weather systems: determination of systematic errors*, Journal of Hydrology, **239**, 2000, pp.179–202.
- [6] Elizabeth E. Ebert, *Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework*, Meteorological applications, **15**, 2008, pp. 51–64.
- [7] G. H. Golub and C. F. Van Loan, *Matrix Computations, third edition*, Johns Hopkins University Press, 1996.
- [8] Great Britain. *Climate Change and Sustainable Energy Act 2006*. London : The Stationery Office.
- [9] Ross N. Hoffman, Zheng Liu, Jean-Francois Louis and Christopher Grassotti *Distortion Representation of Forecast Errors*, Monthly Weather Review, **123**, 1995, pp.2758–2770.
- [10] Rob J. Hyndman and Anne B. Koehler, *Another look at measures of forecast accuracy*, International Journal of Forecast Accuracy, **22**, 2006, pp.679–688.
- [11] Christian Keil and George C. Craig, *A displacement and Amplitude Score Employing an Optical Flow Technique*, Weather and Forecasting, **24**, 2009, pp. 1297–1308.
- [12] I. Moghram and S. Rahman, *Analysis and evaluation of five short-term load forecasting techniques*, IEEE Transactions on Power Systems, **4**(4), 1989, pp.1484–1491.
- [13] James Munkres, *Algorithms for the Assignment and Transportation Problems*, Journal of the Society for Industrial and Applied Mathematics, **5**, No. 1, 1957, pp.32–38.
- [14] A.H. Murphy and R.L. Winkler, *A general framework for forecast verification*, Monthly Weather Rev., **115**, 1987, pp.1330–1338.
- [15] A.H. Murphy, *What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting*, Wea. Forecasting, **8**,1993, pp.281–293.
- [16] Thomas Nehrkorn, Ross N. Hoffman, Christopher Grassotti and Jean-Francois Louis, *Feature calibration and alignment to represent model forecast errors: Empirical regularization*, Quarterly Journal of the Royal Meteorological Society, **129**, 2003, pp.195–218.

- [17] J.V. Paatero and P.D. Lund, *A model for generating household electricity load profiles*, Int. J. Energy Res., **30**, 2006, pp.273–290.
- [18] Alexander Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*, Springer, 2002.
- [19] N. Tomizawa. *On some techniques useful for solution of transportation network problems*, Networks, **1**, 1971, pp. 33–34
- [20] N.S. Wade, P.C. Taylor, P.D. Lang and P.R. Jones, *Evaluating the benefits of an electrical energy storage system in a future smart grid*, Energy Policy, **38**, 2010, 7180–7188.