# Improving forecast accuracy for grid demand and renewables supply with pattern-match features

Usman S. Sanusi
Dept. of Computer Science
Heriot-Watt University
UK
us.sanusi@gmail.com

David Corne
Dept. of Computer Science
Heriot-Watt University
UK
dwcorne@gmail.com

*Abstract*— Short term forecasts of electricity demand and renewables supply are vital for smart grid applications, and in particular for the successful integration of renewables. Improved forecasting of these quantities supports several aspects of smart grid control and deployment, including planning, pricing, staffing, and more. Globally, there is a target for renewables engagement of ~36% (percentage of renewables in the global energy mix) by 2030. This is expected to increase global GDP by 1.1%, and to substantially reduce GHG emissions. It is well known that one of the main barriers to the integration of renewables is the relatively unpredictable nature of renewables availability. A further difficulty (in many grid applications, irrespective of renewables), is the similarly unpredictable nature of energy demand. In this paper we explore an idea in the area of time series prediction that was developed specifically for grid-relevant applications, and we test this method on wind-speed data and on energy demand data. The basis of the approach is the exploitation of patterns in the time series that are maximally similar to the current context. However, unlike previous work that has explored the use of such pattern-matching directly for prediction, we instead use the outcome of pattern matching to generate one or more additional *features* for the subsequent machine learning. Experiments suggest that (i) the approach reliably leads to improved forecasts; and (ii) the approach has wide applicability, independent of the machine learning approach in use.

***Keywords— forecasting; prediction; renewables; time series; pattern matching; regression; feature engineering; feature selection.***

## I. Introduction

Artificial intelligence techniques such as machine learning and other optimization methods, are being increasingly used in the energy sector. In this context, the requirement for accurate predictions for solar and wind energy availability, and local electricity demand, cannot be overemphasized. Accurate supply forecasts are crucial for unlocking the potential of renewable energy. Meanwhile, it is important to note that for many applications we need to forecast the *surplus* or *deficit* in renewables (i.e. the difference between supply and demand), and therefore accurate demand forecasts are equally essential for de-risking a variety of applications in the smart grid sector.

Recent figures from the International Renewable Energy Agency have indicated that substantial engagement of renewables, to become about 36% of the global energy mix by 2030, would have positive impact on the global economy [16]. The report predicted that this would lead to a global GDP increase of 1.1%, and would potentially provide sufficient emission reductions for the active campaign against global warming. The prospects of environment-friendly generation, on-site source, and reduced losses in transmission and distribution, are some of the benefits of renewable generation. However, there are also some well-known problems, primarily including lack of coincidence between availability and demand, and differences between the generated and dispatched energy. These and many other challenges make projections on the uptake and penetration, of green energy utilization difficult to achieve. Efficient harnessing and integration of different energy sub-systems, remains the focal point in contemporary research. Thus, strategies of scheduling demand and the techniques for improving the accuracy of renewables forecasts can be expected to have a significant impact.

Forecasting of renewable energy supply naturally falls into a series of two main steps: (i) forecasting the appropriate weather variables, and then (ii) based on the weather forecasts, predicting the supply on the basis of the installed equipment. In this paper (when considering forecast of supply) we focus on the former (arguably more fundamental) aspect. Forecasting of weather variables has been an active research area for many years, and essentially divides into two areas. The first is large-scale numerical weather prediction (NWP) approaches, which essentially simulate weather systems and are used to understand and predict weather patterns across geography, emphasizing broad accuracy large spatial areas and day–to-week timescales. In contrast, the second area is concerned with improving the accuracy of a highly localized forecast (e.g. for a specific building or wind farm). The area of relevance in this paper is the latter, and it is invariably treated as a time series forecasting problem. The list of approaches

978-1-5090-4240-1/16/$31.00 ©2016 IEEE

that have been researched and used for this task essentially reflect the majority of approaches that have been developed for time series analysis; more recently, advanced statistical and machine learning methods have added to the mix.

In context, the task of predicting local weather variables (perhaps one or more hours ahead) tends to take place in a data rich environment, which recent and historical data are readily available, and current observations are also recorded for use in future predictions. It is therefore no real surprise that 'data-driven' statistical, time-series and/or machine learning methods are favored for accurate localized predictions of weather variables. As with all such applications, however, accuracy is arguably more dependent on the *features* than on the learning algorithm used. That is, the engineering of the features that are used as the input to the learning algorithms is increasingly recognized as key to performance, and an active area of research. In [6], for example, we showed how wind-speed forecasting could reliably be improved by also using temperature and pressure data as features rather than wind-speed alone, while in [18] we showed additional performance benefits from incorporating 'derived features' (e.g. differences in raw features, weighted averages, and so on) in co-operation with a suitable feature selection strategy.

In this paper we explore an additional idea in feature engineering, which is to derive new features based on pattern-matching. The basic idea is to introduce new features that capture a 'historical signal'; pattern matching is done to find one or more historical data patterns that best match the current context, and the wining historical patterns then provide one or more candidate prediction values. Unlike other research that has then used such values as the basis of the prediction itself, however, we use them instead as additional data features.

The remainder of this paper is set out as follows. In section II we briefly discuss related research efforts, and in section III we discuss the setup and rationale for the experiments that follow. Section IV then shows and discusses the results of our series of experiments. Section V offers a summary and some concluding discussion.

## II. FURTHER NOTES ON RELATED WORK

Time series forecasting in the context of renewables prediction is a highly active area, and many excellent reviews are available, e.g. [2, 10, 14]. Similarly, electricity load forecasting is crucial for effective planning and operations, and a review of recent approaches can be found in [19], which similarly emphasis time-series forecasting based approaches. The approaches that tend to be used in both contexts echo the standard statistical methods of exponential smoothing and autoregressive multivariate (ARIMA) models [12, 15, 24]. Other approaches used are modern machine learning algorithms, especially Artificial Neural Networks (ANN) [3, 7, 9], and support vector machines (SVM) [13, 15, 23]. While smoothing and ARIMA techniques are more of linear models, they faced difficulties modelling complex nonlinear trends. ANN and SVM are increasingly investigated in this context since they are believed to be more robust against noise, errors and nonlinear dynamics; however they are often not as effective as ARIMA models, in dealing with seasonal or trend variations. The full variety of approaches is reviewed in [9], while [19] focusses on their application in predicting electrical energy demand and supply.

'Pattern matching' approaches to forecasting are approaches in which we make a prediction based on a close match between historical data and the current context. For example, if we want to predict wind-speed two hours from now, we could proceed as follows: (i) find a pattern of historical wind-speeds in a five-hour (for example) window at the location of interest that matches closely the most recent five-hour window; (ii) use the value 'two hours to the right' of the historical window as the current prediction. It is clear from perusal of the related literature that 'Pattern-matching' based approaches are not particularly prominent in time series forecasting methods and their application; the reason for this is likely to be the fact that pattern-matching tends to be less effective than the standard approaches in general. However to some extent this depends on the patterns of variability in the application at hand. In [21, 22], for example, pattern-matching methods (among others) are explored for road traffic forecasting. A related approach is also considered for electricity demand forecasting in [4], and for financial forecasting in [25]. In these and other works that have explored pattern-matching, however, the pattern-match 'value' is used as the basis for the prediction itself. However, we have found no examples of work where the pattern-match value is instead used as a *derived feature*, and rolled in with other features to enhance the input data to the learning algorithm.

## III. ALGORITHMS AND EXPERIMENTS

### A. *pmReg: pattern-matching based regression*

Our pattern matching based regression algorithms (pmReg) is based on the simple subsequence matching approach reviewed in [10]. A longer time series is scanned sequentially, to find subsequences that match a given query sequence. A query sequence simply corresponds to the most recent window (of suitable length) finishing with the latest observed value. The matching is based on straightforward Euclidean similarity measures (see later for variations on this). In pmReg slide windows of size 6, 12, 18 and 24 values along the entire training set, then return the location of the windows with minimal error, breaking ties based on recency. Thus, in order to support prediction of the value '*N* steps ahead', four features are returned for any given query sequence. These four features correspond to the '*N* steps ahead' values to the right of the best matching window for each window size.

In the following we refer to these features as the 'pattern match values' or 'PM values'; terms such as 'PM18' denote the PM value corresponding to the 18-step window. These PM values are then used along with the corresponding historical window values, as input features to a set of machine leaning algorithms available in Weka [11] for experimentation.

In some variations of the pmReg feature extraction process we adopted a weighting strategy. In one such strategy, recent values are given more weight. Two other cases considered weighting of partial Euclidean distances: (i) distance formed between only the raw values, and (ii) distance formed between differences of any two consecutive normal values (derived values). In the latter case, for example, this means that a one-step increase in wind-speed of $x$ m/s from a historical window would match perfectly with a one-step increase in wind-speed of $x$ m/s from the query sequence, even if the absolute wind-speed values involved were different.

### B. *The Datasets*

The datasets used in the experiments are time series datasets of weather variables and domestic electricity demand. The weather dataset comprises five months of data, comprising hourly weather variable datapoints from May to September 2004, monitored from a location in Lekki City, Nigeria. just off the coast of the Atlantic Ocean. The observed variables include wind speed, wind direction, and temperature. We also perform experiments with two energy demand datasets. The first is from the "Smart*" project [5], and is a collection of data on household electricity usage recorded every minute, and also including weather statistics such as temperature, humidity, wind chill and heat index. These were monitored for a number of months, from a home in the United States. Lastly, we use an aggregated demand dataset, the United Kingdom National Grid's demand profile, representing periodic electricity demand for England and wales throughout the year 2013 [20].

### C. *Experiments*

A range of experiments were conducted to explore whether the use of PM values as features can lead to benefits in forecasting of renewables supply and/or electricity demand. In all of these experiments, a small selection of appropriate machine learning methods were used to build predictive models, namely: linear regression (LR), support vector regression (SVR – typically using an RBF kernel), and MLP (multi-layer perceptron – equivalent to an ANN); from preliminary and related work, these methods are representative of the techniques usually applied in this context, and also tend to have complementary strengths as the forecast horizon moves from 1 to 24 hours. The Weka toolkit implementations [11] were applied in each case. Meanwhile, our own pattern matching based regression algorithms (pmReg) was used to extract features, as discussed, which were then used to augment the input data for LR, SVR and MLP.

In each experiment, six predictive models were trained, for each of one hour, three hours, six hours, 12 hours, 18 hours, and 24 hours ahead. Benchmark results were first obtained by learning predictive models from a 24 input (24 hour) window of data values. In other words, the benchmark results established the accuracy obtainable for each 'hours ahead' horizon when given only the 24 data points up to the current observation. Following the benchmark experiments, further experiments were done which involved augmenting the 24 raw data points with PM values derived in a variety of ways. All experimental results are presented, as the root mean squared error (RMSE), averaged over the test set results, and further averaged over ten cross-validation folds.

## IV. Results

### A. *Wind-speed data: benchmarking*

In order to have baseline results, the complete 5 months wind speed hourly data was investigated (i.e. in this experiment, all results are based on ten-fold cross-validation where the training and test sets were always 9/10 of the 5 months of data, and 1/10 respectively). Using 24 most recent values as input, prediction accuracies for a number of predictive models (1, 3, 6, 12, 18 and 24 hours ahead) were obtained for different settings: (1) using just the normal historical values, (2) using normal values with an added feature (PM) discovered by pmReg, (3) using the normal values + multiple PM features based on different past windows (24, 18, 12 & 6). Weka's Linear Regression (LR), Multi-Layer Perceptron (MLP) and Support vector Regression (SVR) were tested.

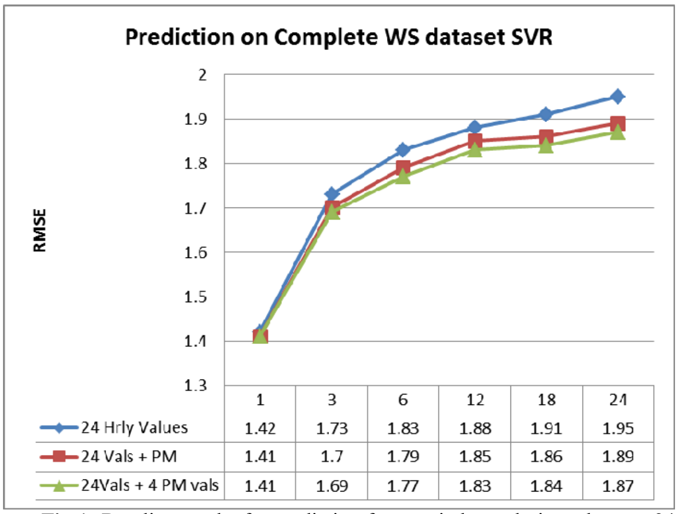


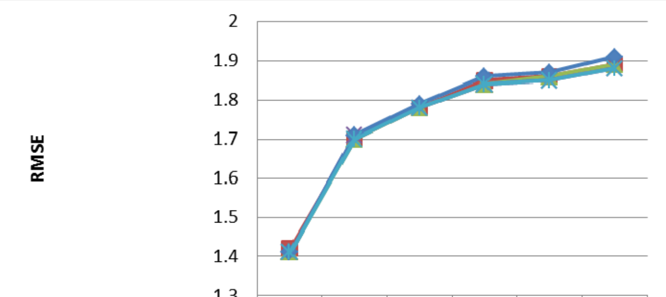

**Fig 1:** Baseline results for predicting future wind-speed given the past 24 hrs. Comparing best performance achieved on the normal values with the best on a dataset with added features from pattern matching (PM values).

Figure 1 shows the performance of the best-performing approach in each of the three scenarios (normal values only, normal values + single PM value and with added multiple PM values). The best performing approach in each case was SVR, however LR achieved similar accuracy levels.

Analysis of the figures above leads to the following observations:

- Introducing the single PM value (based on 24 window values) has a positive and consistent effect on the predictive accuracies.
- Addition of other PM values arising from various window lengths (6, 12 and 18), further improve accuracies.

Figure 2 below shows baseline results when all the discovered PM values were used, against instances of not having any one of them. LR and SVR (RBF-kernel) interestingly recorded nearly identical results, with difference of 0.01(LR) in some cases.



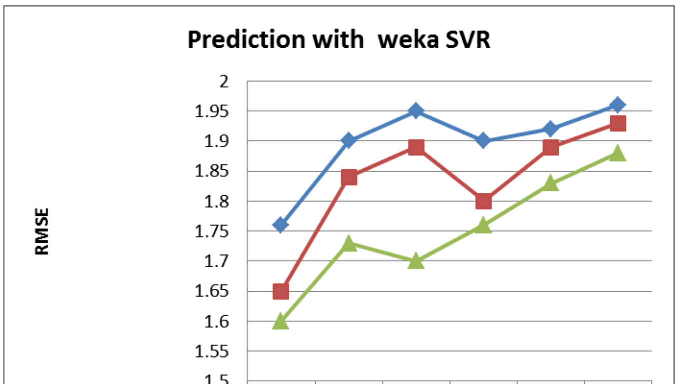

| | 1 | 3 | 6 | 12 | 18 | 24 |
|---|---|---|---|---|---|---|
| 24 Vals + 3 PM1 Vals | 1.42 | 1.7 | 1.78 | 1.85 | 1.86 | 1.89 |
| 24 vals + 3 PM2 Vals | 1.41 | 1.7 | 1.78 | 1.84 | 1.86 | 1.89 |
| 24 Vals + 3 PM3 vals | 1.41 | 1.71 | 1.78 | 1.84 | 1.85 | 1.88 |
| 24 vals + 3 PM4 vals | 1.41 | 1.71 | 1.79 | 1.86 | 1.87 | 1.91 |
| 24Vals + 4 PM vals | 1.41 | 1.7 | 1.78 | 1.84 | 1.85 | 1.88 |

**Fig 2:** Baseline prediction results comparing performance on the past 24 values + all the 4 PM features with similar settings but omitting any one PM feature at a time.

The mutual benefit of all the discovered PM features is clear from figure 2 as observed here:

- Omitting PM feature based on the window of 6 values (PM1), prediction accuracies of 1, 12, 18 and 24 hours begin to deteriorate.
- Similarly, absence of 12-window PM feature (PM2) negatively affects 18 and 24 hours predictions.
- Also, omitting 18-window PM feature (PM3), causes drop in accuracies of the 3 and 18 hours predictions.
- Lastly, not having 24-window based PM feature, negatively affect accuracies for all the 3-24 hours look-a-heads.

Targeting additional results for comparison, more investigations were designed. Hence conducting further baseline experiments, but with reduced datasets of just 2-4 weeks values. These were aimed to capture as closely as possible, the dynamic reality, variability and complex trend of time series such as wind speed, stocks indices and etc.



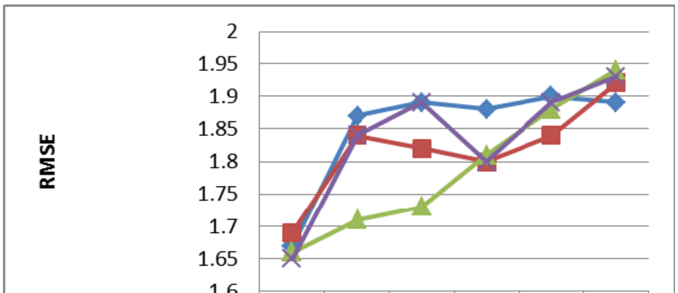

| | 1 | 3 | 6 | 12 | 18 | 24 |
|---|---|---|---|---|---|---|
| 24 h-Vals | 1.76 | 1.9 | 1.95 | 1.9 | 1.92 | 1.96 |
| 24 Vals+PM | 1.65 | 1.84 | 1.89 | 1.8 | 1.89 | 1.93 |
| 24 Vals + 4 PM vals | 1.6 | 1.73 | 1.7 | 1.76 | 1.83 | 1.88 |

**Fig 3:** Baseline results for predicting future wind-speed given the past 24 hours on the 2 weeks data. Comparing best performance achieved on the normal values, those with added single and multiple PM values.

The figure above presented remarkable performance improvement with added PM-based features. Additionally:

- It showed robustness of building predictive models at short intervals, which would be repeated periodically for regular updates.
- The best curve has 24 past hourly values + 4 PM features (based on 24, 18, 12 and 6 windows).

Concluding baseline expriments, smaller historical values were taken as input, contrasting the previous settings of always using past 24 data values. Figure 4 showed results on using different history lengths (6, 12, 18 and 24).




| | 1 | 3 | 6 | 12 | 18 | 24 |
|---|---|---|---|---|---|---|
| 6-Vals + PM-6 | 1.67 | 1.87 | 1.89 | 1.88 | 1.9 | 1.89 |
| 12-Vals + PM-12 | 1.69 | 1.84 | 1.82 | 1.8 | 1.84 | 1.92 |
| 18-Vals + PM-18 | 1.66 | 1.71 | 1.73 | 1.81 | 1.88 | 1.94 |
| 24-Vals + PM-24 | 1.65 | 1.84 | 1.89 | 1.8 | 1.89 | 1.93 |

**Fig 4:** Baseline results for predicting from the variable input values. Comparing performance on the 24 hrs values with others (6, 12, and 18), on datasets with added corresponding PM values.

Observation from figure 4 further established the relative importance of the PM-based features. While there could be lots of possible matches for small windows, trend might be difficult to repeat involving large window values. But due to the PM-based features, performances were relatively similar despite differences in the length of historical values. And 12-18 past values recorded better performances.

*B. Weighted Pattern Matching I (wpmReg)*

Evidence for consistent positive effects of the PM-based features, showed in the previous results, are quite insightful as improvements scaled very well with different data quantities. Contrasting the simple implementation of pattern matching, some variations were devised by means of weighting strategy. Figure 5 below presents accuracies of a variant pmReg , where squared differences of the most recent values are given more weights.

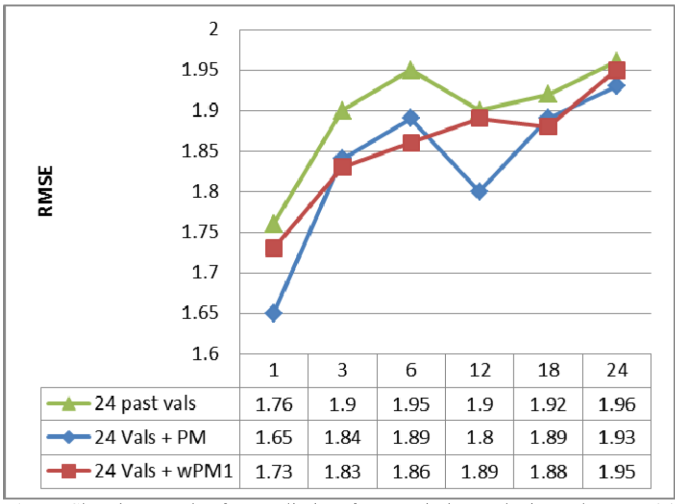



**Fig 5:** Showing results for predicting future wind speed given the past 24 historic values. Comparing best performance achieved on the normal values with the best on datasets with added variations of PM features.

Linear Regression and SVR continue to record best performances, and also accuracies further improve with the presence of any of the PM-variant features.
Figure 5 above, clearly indicated the advantage of adding the weighted PM feature (wPM1). Though, it is not better than the simple-minded one (PM), which allocated equal weights to all the squared differences.

*C. Weighted Pattern Matching II*

Other implementations considered weighting of partial Euclidean distances: (i) distance formed between only the normal values (EUD_Normal), and (ii) distance formed between differences of any two consecutive values of the windows (EUD_Derived). This resulted into four cases: (1) Allocating 80 and 20 per cent weights to EUD_Normal and EUD_Derived respectively.
(2) EUD_Normal (60%), EUD_Derived (40%). (3) EUD_Normal (40%), EUD_Derived (60%). (4) EUD_Normal (20%), EUD_Derived (80%).

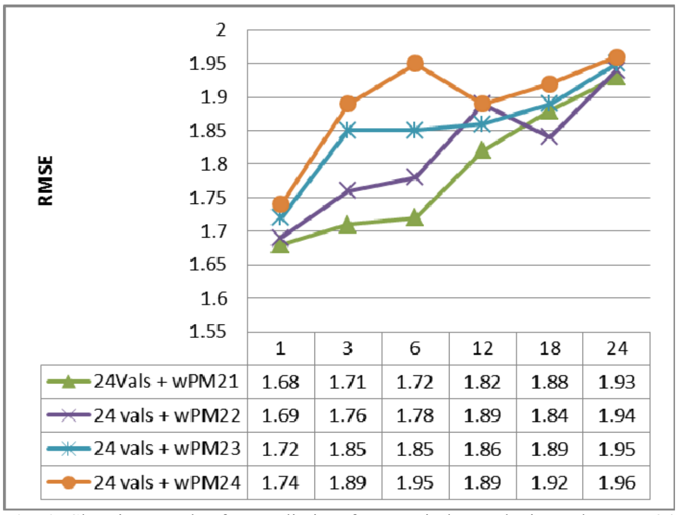



**Fig 6:** Showing results for predicting future wind speed given the past 24 historic values. Comparing best performance achieved with added variations of the weighted PM features.

Key observation from fig 6 pointed case 1 as best setup. Ie accuracies are better with minimum weight given to the derived values (difference of any 2 consecutive window values). And the setting helps to ascertain the contribution of derived values, which were found to be quite beneficial, and validating the finding in [18].

The figure below, bring together all the different instances; predicting from only the past values, as well as when any PM feature is added.

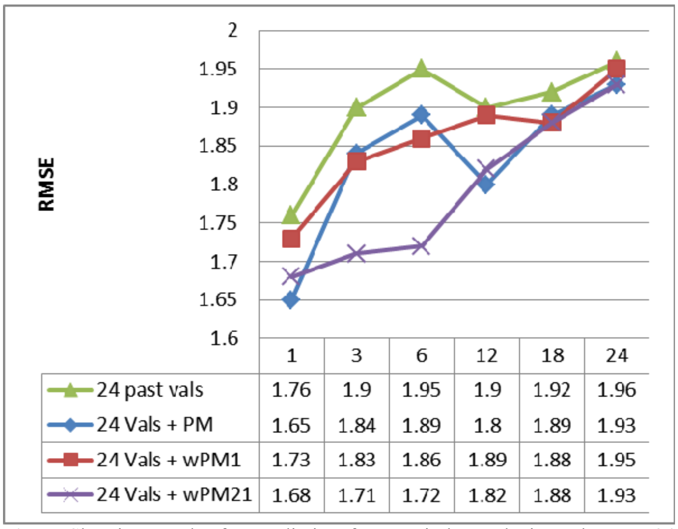



**Fig 7:** Showing results for predicting future wind speed given the past 24 historic values. Comparing best performance achieved on the normal values with the best on datasets with added PM features of all variations.

The figure above corroborated the remarkable influence of PM features regardless of implementation differences.
But the important thing to note is the effect of derived values, which can be observed by comparing the curve with no derived values (24vals + PM) with the one having 20% derived values (24vals + wPM21).

Obviously the setting allocating per cent weighting to separate distances (EUDs), accorded the opportunity to figure the impact of individual contributors.

*D. Validation I: Smart Electricity and weather data*

We validated our findings by testing the selected machine learning algorithms on different sets of data input from the smart project. The household's electricity usage (watts) recorded every second, was reduced to hourly averages, leading to preprocessing of 86,400 values to just 24 each day. The experimentation proceeded as follows:

- Maximum of 2-4 weeks of data were used periodically for building the predictive models.
- Linear regression, MLP and SVR in weka [11], were used to get accuracies for different predictive models based on just the past 24 hours values.
- The usual 3 ML algorithms were applied to similar set up using 24 past values, plus an added PM feature discovered by pmReg based on 24 window values.
- Other results were obtained on the 24 past values plus all the 4 PM features (24, 18, 12 & 6 windows).

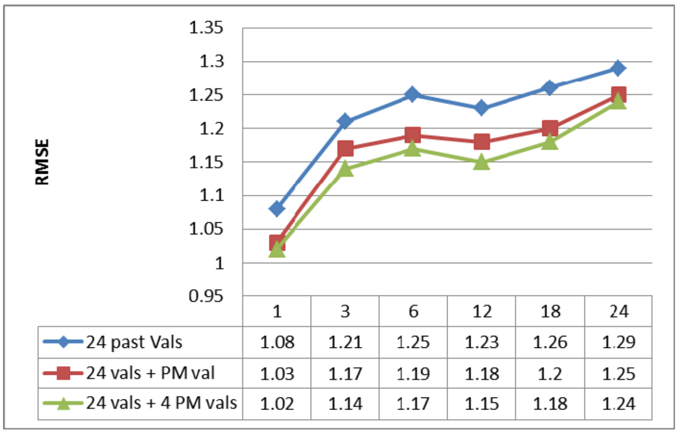



**Fig 8:** Showing results for predicting future electricity usage given the past 24 historic values. Comparing best performance achieved on the normal values with the best on datasets with added PM feature and with all 4 PM features added. **Legend: 1:750 watts**

Figure 8 summarises the results for predicting electrical usage, suggesting SVR as the most effective ML method, and the curves depicted similar patterns as the ones in fig 3 above.

The environment data provided temperature values taken every 5 minutes. Given its influence on energy demand, we tested our method on temperature forecasting.

Preprocessing temperature values in degrees (F), we obtained hourly averages, hence reducing the 288 values to 24 on each day, then follow experimental procedures as in the case of last validation test.

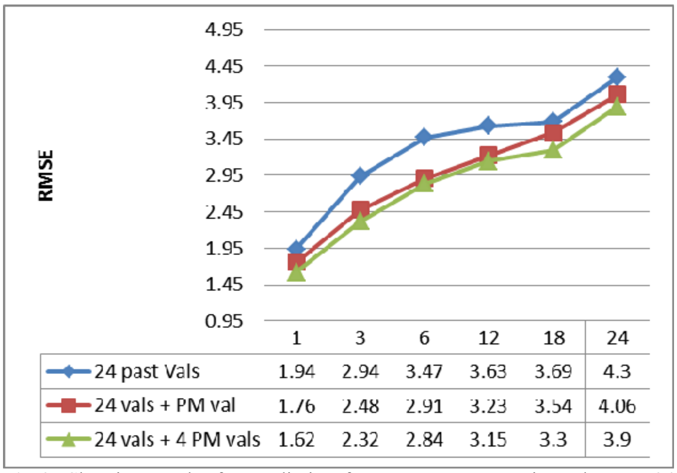



**Fig 9:** Showing results for predicting future temperatures given the past 24 historic values. Comparing best performance achieved on the normal values with the best on datasets with added PM feature and with all 4 PM features added.

The following are clear from the two validation tests, despite differences in the nature and dynamics of the datasets:

- SVR proved the most successful algorithm
- Accuracy always improve with an added PM value
- Presence of multiple PM values further help, though slightly in some cases.

*E. Validation II: England &Wales Electricity demand profile*

The half hourly electricity demand profile was reduced to hourly averages, which was then experimented in the same way as other validations.

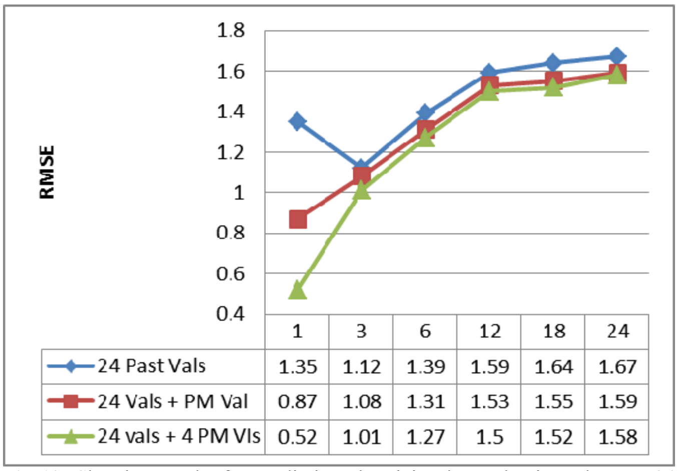
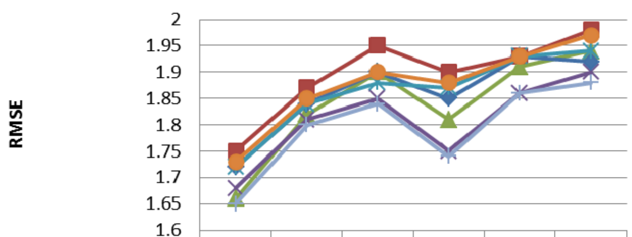



**Fig 10:** Showing results for predicting electricity demands given the past 24 historic values. Comparing best performance achieved on the normal values with the best on datasets with added PM feature and with all 4 PM features added. **Legend: 1:2000 Mega Watts**

Multilayer Layer Perceptron achieves the best result, in this case. While the relative benefit of PM values has fully manifested in fig 10, as the curve for past values only, looked anomalous compared to the smooth PM-added curves.

### F. *k-NN based Pattern matching*

In the previous subsections, pattern-matching features, were obtained on the basis of 1-NN similarity. In this section we explore the use of *k*-NN in this context, where *k* PM value features are used, one for each of the *k* nearest matches in the historical data. Results are summarized in Figure 11.




| | 1 | 3 | 6 | 12 | 18 | 24 |
|---|---|---|---|---|---|---|
| 24 Vals | 1.75 | 1.87 | 1.95 | 1.9 | 1.93 | 1.98 |
| 24vals +1NN | 1.66 | 1.82 | 1.9 | 1.81 | 1.91 | 1.94 |
| 24 Vals + 3NN | 1.68 | 1.81 | 1.85 | 1.75 | 1.86 | 1.9 |
| 24 Vals + 5NN | 1.72 | 1.84 | 1.9 | 1.85 | 1.93 | 1.92 |
| 24 vals + 7NN | 1.72 | 1.84 | 1.88 | 1.87 | 1.93 | 1.94 |
| 24 Vals + 9NN | 1.73 | 1.85 | 1.9 | 1.88 | 1.93 | 1.97 |
| 24 Vals + All-NN | 1.65 | 1.8 | 1.84 | 1.74 | 1.86 | 1.88 |

**Fig 11:** Showing results for predicting future wind speed given the past 24 historic values. Comparing best performance achieved on the normal values with the best on datasets with added single k-NN based PM features and with multiple PM features.

Figure 11 is achieved by the SVR, and echoes the findings recorded in figure 1:

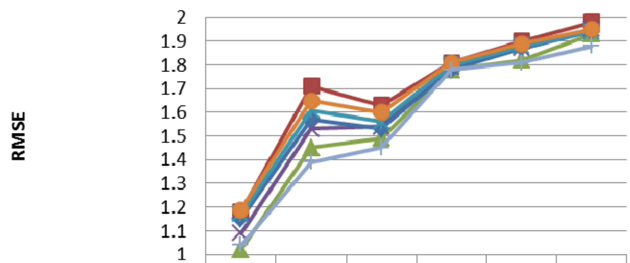
- Introduction of any of the PM values, lead to a consistent positive effect on predictive accuracies.
- Multiple PM values have an overall best improvement.

It is worth noting that multiple PM values here were based on multiple matches with the same window size, while in figure 1 the multiple PM values were based on variable window lengths (6, 12, 18 and 24).

### G. *Other Distance Measures*

All results shown so far were achieved using (perhas weighted) Euclidean distance. In this subsection we investigate alternative distance metrics.

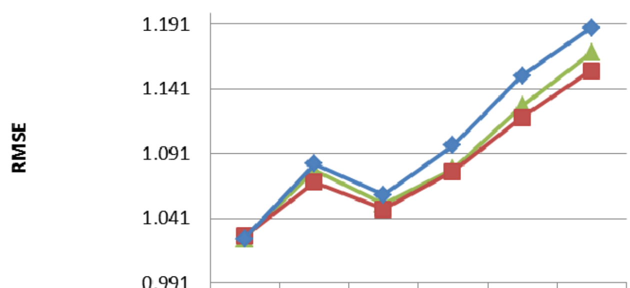
Figure 12 shows results on the electricity demand data where PM features are found based on the Dynamic Time Warping (DTW) distance measure (implemented in the java machine learning library [1]) which attempts to correct for cases of two patterns being fundamentally similar, although one may be 'stretched out' more than the other over time.




| | 1 | 3 | 6 | 12 | 18 | 24 |
|---|---|---|---|---|---|---|
| 24 Vals | 1.18 | 1.71 | 1.63 | 1.81 | 1.9 | 1.98 |
| 24vals +1NN | 1.02 | 1.45 | 1.49 | 1.78 | 1.82 | 1.93 |
| 24 Vals + 3NN | 1.09 | 1.53 | 1.54 | 1.79 | 1.87 | 1.95 |
| 24 Vals + 5NN | 1.15 | 1.57 | 1.53 | 1.78 | 1.87 | 1.94 |
| 24 vals + 7NN | 1.17 | 1.61 | 1.56 | 1.8 | 1.88 | 1.94 |
| 24 Vals + 9NN | 1.19 | 1.65 | 1.6 | 1.81 | 1.89 | 1.95 |
| 24 Vals + All-NN | 1.04 | 1.39 | 1.45 | 1.78 | 1.81 | 1.88 |

**Fig 12:** Showing results for predicting electricity demands given the past 24 historic values. Comparing best performance achieved on the normal values with the best on datasets with added single k-NN based PM features and with multiple PM features. **Legend: 1:1700 Mega Watts**

Leveraging the nonlinearity capabilities of the Radial Basis Function (RBF) kernel and DTW metrics, we investigated these distance measures alongside Euclidean distance (again, available from the javaml library). And the results (again, for the electricity demand case) are summarized in the Figure 13.




| | 1 | 3 | 6 | 12 | 18 | 24 |
|---|---|---|---|---|---|---|
| 24 vals + PM on EUC | 1.025 | 1.078 | 1.052 | 1.079 | 1.128 | 1.169 |
| 24 Vals + PM on DTW | 1.027 | 1.069 | 1.048 | 1.077 | 1.119 | 1.155 |
| 24 Vals + PM on RBF | 1.025 | 1.083 | 1.059 | 1.097 | 1.151 | 1.188 |

**Fig 13:** Showing results for predicting future electricity usage given the past 24 historic values + PM features. Comparing performance achieved using different distance measures. **Legend: 1:700 watts**

Figure 13 indicates the relative performance of three distance metrics on the electricity usage dataset. Although DTW may provide marginal improvement, it seems clear that alternative measures do not have a significant advantage over straightforward Euclidean distance in this case.

## V. Concluding Discussion

Predictions on renewables and that of the energy demand, heavily rely on weather conditions. While the wind speed forecasts serve as principal inputs to the predictive models in wind-generation. Temperature on the other hand, played a major role in energy demand forecasts, mainly due to cooling and/or heating purposes in residential areas. For effective renewable and energy demand predictions, methods suitable for time series forecasts are predominantly employed. These include advanced statistical and machine learning algorithms.

But good performance by any of the methods involved, depend on the input features, which many a times are selected few, often derived or transformed from the pool of the original features set. A number of efforts have successfully proved usefulness of exploring available features. While [6] recorded remarkable benefits when input features are chosen creatively. The advantage of derived features (e.g. simple differences) and features from other variables were reported in [18]. It is evident that a lot more work is required to discover additional means of getting precious input features.

This paper pursued features set exploration from an unusual perspective, as we aimed to leverage site-specific forecasts such as at the wind farm installations, solar panels points or the household's energy demand. Tasks of this nature involve data intensive operations as well as the advanced computing. Our procedure exploited pattern discovery techniques, to suggest a set of features, to enable more robust machine learning on any time series.

While several works used pattern matching techniques for the actual forecasting, such that the pattern match value was always taken as the forecast, or in combination with another forecast. Our novel approach use pattern match values (PM) as features, i.e. applying pattern matching strategy mainly for discovering promising input features.

## Acknowledgment

The first author is sponsored by the Nigerian Petroleum Technology Development Fund (PTDF), and we are also grateful to the Energy Commission of Nigeria (ECN) for providing the majority of the data used.

## References

[1] Abeel, T., Peer, Y. V. D., & Saeys, Y. (2009). Java-ml: A machine leraning library. Jounal of Machine Learning Research, 10(Apr), 931-934

[2] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. Renewable and Sustainable Energy Reviews, 12:1725–1744, 2008.

[3] A. More and M. C. Deo. Forecasting wind with neural networks. Marine Structures, 16:35–49, 2003.

[4] Al-Qahtani, F. H., & Crone, S. F. (2013, August). Multivariate k-nearest neighbour regression for time series data—A novel algorithm for forecasting UK electricity demand. In Neural Networks (IJCNN), The 2013 International Joint Conference on (pp. 1-8). IEEE.

[5] Barker, S., Mishra, A., Irwin, D., Cecchet, E., Shenoy, P., & Albrecht, J. (2012). Smart*: An open data set and tools for enabling research in sustainable homes. SustKDD, August, 111, 112

[6] Corne, D., Dissanayake, M., Peacock, A., Galloway, S., & Owens, E. (2014, December). Accurate localized short term weather prediction for renewables planning. In Computational Intelligence Applications in Smart Grid (CIASG), 2014 IEEE Symposium on (pp. 1-8). IEEE.

[7] E. Cadenas and W. Rivera. Short term wind speed forecasting in La Venta, Oaxaca, M´exico, using artificial neural networks. Renewable Energy, 34:274–278, 2009

[8] E. Cadenas and W. Rivera, "Short term wind speedforecasting in La Venta, Oaxaca, Mexico, using artificial neural networks", Renewable Energy, vol. 34, no. 1, 2009, pp. 274–278.

[9] Fu, Tak-chung. "A review on time series data mining." Engineering Applications of Artificial Intelligence 24, no. 1 (2011): 164-181.

[10] G. Giebel, R. Brownsword, G. Kariniotakis,M. Denhard, and C. Draxl. The state-of-the-art in short-term prediction of wind power: A literature overview, 2nd edition. Technical report, ANEMOS.plus and SafeWind projects, Jan. 2011

[11] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.

[12] J. H. Kim, "Forecasting autoregressive time series with biascorrected parameter estimators", International Journal of Forecasting, vol.19, 2003, pp. 493–502.

[13] K. J. Kim, "Financial time series forecasting using support vector machines", Neuro-computing, vol. 55, 2003, pp. 307- 319.

[14] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan. A review on the forecasting of wind speed and generated power. Renewable and Sustainable Energy Reviews, 13:915–920, 2009.

[15] Mounce, S. R., Mounce, R. B., & Boxall, J. B. (2011). Novelty detection for time series data analysis in water distribution systems using support vector machines. Journal of hydroinformatics, 13(4), 672-686.

[16] Renewable Energy Benefits: Measuring the Economics, International Renewable Energy Agency (2016), Abu Dhabi. Available at http://www.irena.org/DocumentDownloads/Publications/IRENA_Measuring-the-Economics_2016.pdf

[17] S. Gelper, R. Fried, and C. Croux, "Robust forecasting with exponential and Holt-Winters smoothing", Journal of Forecasting, vol. 29, 2010, pp. 285-300.

[18] Sanusi, U., & Corne, D. (2015, November). Feature selection for accurate short-term forecasting of local wind-speed. In 2015 IEEE 8th International Workshop on Computational Intelligence and Applications (IWCIA) (pp. 121-126). IEEE.

[19] Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting—A review. Renewable and sustainable energy reviews, 16(2), 1223-1240.

[20] The United Kingdom National Grid Historical Demand Data (2016). Available at: http://www2.nationalgrid.com/UK/Industry-information/Electricity-transmission-operational-data/Data-Explorer/

[21] Van Lint, J. W. C., & Van Hinsbergen, C. P. I. J. (2012). Short-term traffic and travel time prediction models. Artificial Intelligence Applications to Critical Transportation Issues, 22, 22-41.

[22] Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. Transportation Research Part C: Emerging Technologies, 43, 3-19.

[23] Y.Radhika and M.Shashi, "Atmospheric Temperature Prediction using Support Vector Machines," International Journal of Computer Theory and Engineering, vol. 1, no. 1, 2009, pp. 55-58.

[24] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

[25] Zhang, Z., Jiang, J., Liu, X., Lau, R., Wang, H., & Zhang, R. (2010, January). A real time hybrid pattern matching scheme for stock time series. In Proceedings of the Twenty-First Australasian Conference on Database Technologies-Volume 104 (pp. 161-170). Australian Computer Society, Inc..