Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Influence-Driven Model for Time Series Prediction from Partial Observations

Saima Aman
Department of Computer Science
University of Southern California
Los Angeles, California
Email: saman@usc.edu

Charalampos Chelmis and **Viktor K. Prasanna**
Department of Electrical Engineering
University of Southern California
Los Angeles, California
Email: {chelmis, prasanna}@usc.edu

Abstract

Applications in sustainability domains such as in energy, transportation, and natural resource and environment monitoring, increasingly use sensors for collecting data and sending it back to centrally located processing nodes. While data can usually be collected by the sensors at a very high speed, in many cases, it can not be sent back to central nodes at a frequency that is required for fast and real-time modeling and decision-making. This may be due to physical limitations of the transmission networks, or due to consumers limiting frequent transmission of data from sensors located at their premises for security and privacy concerns. We propose a novel solution to the problem of making short term predictions in absence of real-time data from sensors. A key implication of our work is that by using real-time data from only a small subset of *influential* sensors, we are able to make predictions for *all* sensors. We evaluated our approach with a large real-world electricity consumption data collected from smart meters in Los Angeles and the results show that between prediction horizons of 2 to 8 hours, despite lack of real time data, our influence model outperforms the baseline model that uses real-time data. Also, when using partial real-time data from only $\approx 7\%$ *influential* smart meters, we witness prediction error increase by only $\approx 0.5\%$ over the baseline, thus demonstrating the usefulness of our method for practical scenarios.

Introduction

Low cost wireless sensors are increasingly being deployed in large numbers for performing monitoring and control in many sustainability domains such as in smart electric grids, transport networks, and natural resource and environment monitoring. These sensors are located at geographically dispersed locations and periodically send back acquired data to centrally located processing nodes (Ciancio and Ortega 2005) via wireless links and the Internet (Chong and Kumar 2003). They include sensors for monitoring natural resource and environment such as biodiversity and atmosphere (Lozano et al. 2009); smart meters for measuring energy consumption (Simmhan et al. 2013), (Marascu et al. 2013); loop detectors installed under pavements for recording traffic (Pan, Demiryurek, and Shahabi 2012); and meters on wind turbines that record wind speed and turbines' power output (Bullis 2014).

Due to several factors, data from all sensors is not available at central nodes in real-time or at a frequency that is required for fast and real-time modeling and decision-making. For example, wind turbines record data every few seconds, but transmit data every five minutes to far-off research centers for use in forecasting algorithms (Bullis 2014). *Physical limitations* of existing transmission networks, such as latency, bandwidth and high energy consumption (Ciancio and Ortega 2005) are key factors that limit the frequency of data transmission from sensors to central nodes (Bouhafs, Mackay, and Merabti 2012). Sometimes, consumers may also limit frequent transmission of information from sensors located at their premises for *security and privacy* concerns (McDaniel and McLaughlin 2009).

All these situations reflect the ***partial data problem***, *where only partial data from sensors is available in real-time*, and complete high resolution data become available only periodically, generally one or more times a day. Without addressing this problem, traditional solutions risk degradation in performance and inaccurate interpretation of generated insights. For instance, time series prediction methods are adversely affected by the prediction horizon length, and in the case of partial data, the effective prediction horizon for those sensors for which data is unavailable in real-time becomes larger, which in turn leads to detrimental increase in prediction error. Thus, the time series approach cannot be used for accurate predictions, for example, for up to 8 hours ahead. A possible approach - as we propose in this paper - is to develop creative solutions using data from a small subset of sensors selected on the basis of some heuristics or learning methods, while minimizing information loss resulting from leaving out data from remaining sensors. The intuition behind this approach is the fact that sensors located spatially close to each other or sensing activities driven by similar schedules - such as those on an academic campus or traffic on high density roads - are likely to be correlated. If this information can be leveraged, it will obviate the need for real-time transmission from *all* sensors to the central nodes, and thereby reduce the load on the transmission network. Also, it would make it simpler to add new sensors without straining the network.

We distinguish *partial data* from *missing data*, which is arbitrary unavailability of data at random time periods due to diverse factors, whereas partial data is systematic unavailability of data for known time periods for a known subset of smart meters, due to non-transmission of data in that period. While selective missing data is lost, partial data becomes available when batch transmission occurs, and can be used to re-train our models.

In this paper we address the partial data problem in the context of smart electricity grids, where high *volume* electricity consumption data is collected by smart meters at consumer premises and securely transmitted back to the electric utility over wireless or broadband networks. There, they are used to predict electricity consumption and to initiate curtailment programs ahead of time by the utility to avoid potential supply-demand mismatch. Partial data problem arises when data from smart meters is only partially available in real-time. To address this, we propose a two-stage solution: first, we learn the dependencies among time series of different smart meters, then, we use data from a small subset of smart meters which are found to have high *influence* on others to make predictions for all meters. Our main contributions are:

1) We leverage dependencies among time series sensor data for making short term predictions with *partial real-time data*. While time series dependencies have been used previously, the novelty of our work is in extending the notion of dependencies to discover *influential* sensors and using real-time data only from them to do predictions for *all* sensors.

2) Using real-world electricity consumption data, we demonstrate that despite lack of real time data, our prediction models perform comparably to the baseline model that uses real-time data, thus indicating their usefulness for practical sustainability domain scenarios.

Related Work

Many predictive modeling methods are designed for ideal scenarios where all required data is readily available. For example, time-series prediction methods such as Auto-Regressive Integrated Moving Average (ARIMA) (Box and Jenkins 1970) and Auto-Regressive Trees (ART) (Meek, Chickering, and Heckerman 2002) require observations from recent past to be readily available in real-time to make short-term future predictions. However, this assumption does not hold true for many sensor-based applications involving "big data" that is only *partially* available in real time. The solutions proposed to address this problem can be categorized into two types: 1) Reduce the volume of transmitted data by techniques such as data compression (Marascu et al. 2013), (Razzaque, Bleakley, and Dobson 2013), data aggregation (Karimi, Namboodiri, and Jadliwala 2013), model-driven data acquisition (Deshpande et al. 2004), and communication efficient algorithms (Sanders, Schlag, and Muller 2013); 2) Estimate missing real-time data by techniques such as interpolation based on regression (Kreindler and Lumsden. 2006), or through transient time models that use differential equations to model system behavior (Cuevas-Tello et al. 2010). Main challenge with these methods is that estimates depend on the accuracy of models and interpolation errors get propagated to subsequent analysis and decision-making steps. Another method for estimation is using spectral analysis of time series, though it is a more complex and involved process that is suitable only for periodic time series (Bahadori and Liu 2012). We use an orthogonal approach where instead of trying to estimate missing real-time data, we first discover *influential* sensors and then do predictive modeling using real time data from only these sensors.

Our approach involves learning dependencies among time series data from different sensors. Several techniques have been proposed to learn dependencies among time series data; the more popular among them are based on cross-correlations (Box and Jenkins 1970) and Granger Causality (Granger 1969). The latter has gained popularity in many domains such as climatology, economics, and biological sciences due to its simplicity and robustness (Bahadori and Liu 2012). It is however time consuming for evaluating pairwise dependencies when large number of variables are involved. Lasso-Granger (Arnold, Liu, and Abe 2007) is proposed to provide a more scalable and accurate solution. In our work, we leverage the Lasso-Granger method to discover dependencies among time series from different sensors, and then identify influential sensors based on these dependencies.

Our work brings the much needed focus to efficient data collection methods for sustainability domains. In smart grid, data streams from thousands of sensors are monitored for predictive analytics (Balac et al. 2013), and demand response (Kwac and Rajagopal 2013). With large scale adoption of smart meters, most cities would soon have millions of smart meters recording electricity consumption data every minute. For utilities, real-time data collection from meters all over a city would be prohibitive due to limited capacity of current transmission networks. Such scenarios necessitate development of alternative methods, such as ours, that could work with only partial data that is available in real-time.

Preliminaries

Consider a large set of sensors $\mathcal{S} = \{s_1, ..., s_n\}$ collecting real-time[1] data. Due to network bandwidth constraints, only some of these sensors can send data back to the central node in real-time, while the rest send the collected data in batches every few hours (Fig. 1). The problem we address is to use this *partial data* to make predictions for *all* sensors.

Problem Definition Given a set of sensors $\mathcal{S}$ with time series outputs $\{x_j^i\}, j = 1, ..., t, i = 1, ..., n$, make short-term predictions $\{x_j^i\}, j = t+1, ..., t+h, i = 1, ..., n$ for each sensor $s_i \in \mathcal{S}$, when readings $\{x_k^o\}, k = t-r+1, ..., t$ for $o \in \mathcal{O}$ are unavailable for a subset $\mathcal{O}$ of sensors, $\mathcal{O} \subset \mathcal{S}$.

For simplicity, we assume all time-series sensor outputs to be sampled at the same frequency and be of equal length.

We *hypothesize* that we can learn dependencies in past time series outputs from sensors and use them to identify the set of sensors that are more helpful in making predictions for

[1]In the context of this paper, data collected at 15-min intervals are considered real-time, even though our models would be applicable (with even greater impact) to data at smaller resolutions.

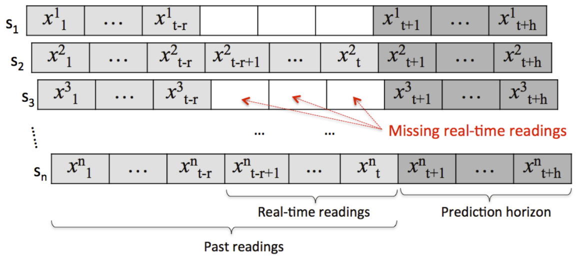


Figure 1: Some sensors can send readings to a central node in real-time, while the rest send every few hours resulting in partial real time data.

other sensors, so that we can collect real-time readings from only these sensors.

Definition 1 *A* ***dependency matrix*** $\mathcal{M}$ *is an* $n \times f$ *matrix, where each element* $\mathcal{M}[i, j]$ *represents the dependence of time series* $\mathcal{T}_i$ *on time series* $\mathcal{T}_j$.

Here, n is the number of sensors and f is the number of features used for each sensor.

Definition 2 *The* ***influence*** $\mathcal{I}^k$ *of a time series* $\mathcal{T}_k$ *is defined as the sum of all values in the column* k *in the dependency matrix* $\mathcal{M}$.

$$\mathcal{I}^k = \sum_{j=1}^{n} \mathcal{M}[j, k] \tag{1}$$

Definition 3 ***Compression Ratio****,* $\mathcal{CR}$ *is defined as the ratio between the total number of sensor readings that would be required for real-time prediction and the number of readings actually transmitted from selected influential sensors for prediction with partial data.*

$$\mathcal{CR} = \frac{\sum_{i=1}^{n} |\mathcal{P}_i|}{\sum_{i=1}^{n} |\mathcal{P}_i| - \sum_{o \in \mathcal{O}} |\mathcal{P}_o|} \tag{2}$$

where $\mathcal{P}_i$ is the sequence of past values from sensor s_i used for prediction and $|\mathcal{P}_i|$ is the length of this sequence; $\mathcal{O}$ is the subset of sensors with unavailable real-time readings and n is the total number of sensors. For simplicity, we consider same length l of past values for all sensors. Hence, $|\mathcal{P}_i| = l, \forall i$ and above equation can be simplified as $\mathcal{CR} = \frac{n}{n - |\mathcal{O}|}$.

Methodology

We propose a two-stage process, where we first learn dependencies from past data and determine influence for individual sensors, and then use this information for selecting influential sensors for regression tree based prediction.

Influence Discovery

We cast the problem of making predictions for a sensor $s_i \in \mathcal{O}$ in terms of recent real time data from other sensors as a regression problem. In ordinary least squares (OLS) regression, given data $(\mathbf{x}^i, y_i), i = 1, 2, ..., n$, the response y_i is estimated in terms of p predictor variables, $\mathbf{x}^i = (x_{i1}, ..., x_{ip})$ by minimizing the residual squared error.

We identify sensors that show stronger *influence* on other sensors using the Lasso Granger method. We thus use Lasso Granger as a novel way for feature selection. The lasso method is used in regression for shrinking some coefficients and setting others to zero by penalizing the absolute size of the coefficients (Tibshirani 1996). The OLS method generally gives low bias due to over-fitting but has large variance. The Lasso improves variance by shrinking coefficients and hence may reduce overall prediction errors (Tibshirani 1996).

Given n sensor outputs in form of time series $\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^n$, with readings at timestamps $\mathbf{t} = 1, ...T$, for each series $\mathbf{x}^i$, we obtain a sparse solution for coefficients $\mathbf{w}$ by minimizing the sum of squared error and a constant times the L1-norm of the coefficients:

$$\mathbf{w} = \arg\min \sum_{t=l+1}^{T} \left\| x_t^i - \sum_{j=1}^{n} \mathbf{w}_{i,j}^T \mathcal{P}_t^j \right\|_2^2 + \lambda \|\mathbf{w}\|_1 \tag{3}$$

where $\mathcal{P}_t^j$ is the sequence of past l readings, i.e., $\mathcal{P}_t^j = [x_{t-l}^j, ..., x_{t-1}^j]$, $\mathbf{w}_{i,j}$ is the *j*-th vector of coefficients $\mathbf{w}_i$ representing the dependency of series *i* on series *j*, and λ is a parameter which determines the sparseness of $\mathbf{w}_i$ and can be determined using cross-validation method.

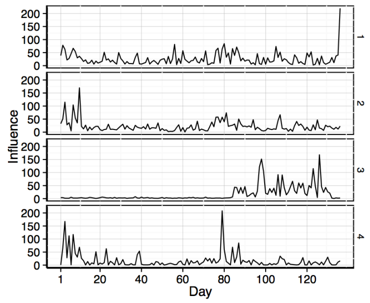
Influence Model (IM)

We first learn the dependency matrix $\mathcal{M}_j$ for each day as follows. Each sensor's data is split into a set of q daily series $\{\mathcal{D}_j^i\}_{i=1,...,n, j=1,...,q}$. Longer time series data is usually non-stationary, i.e., the dependence on preceding values changes with time. Splitting into smaller day-long windows ensures stationarity for time series data in each window. Dependency matrix is re-calculated daily to account for changes in influence over time. Weights for daily series for each day are calculated using eqn. 3. The weight vectors $\mathbf{w}_i$ form the rows of the dependency matrix $\mathcal{M}$. We set diagonals of the dependency matrix to zero, i.e., $\mathcal{M}[i, i] = 0$ in order to remove self-dependencies and simulate the case of making prediction without a sensor's own past real-time data. Given $\mathcal{M}$, *influence* $\mathcal{I}$ of all series can be calculated using eqn. 1.

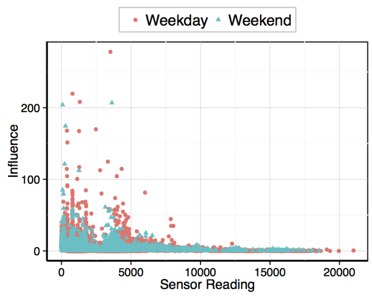
Predictions for a given day are based on training data from a previous *similar* day sim. We consider two cases of similarity: 1) *previous week* - same day in the preceding week, which captures similarity for sensor data related to periodic (weekly) human activities; 2) *previous day*, which captures similarity for sensor data related to human activities and natural environments on successive days.

We apply a windowing transformation to the daily series $\{\mathcal{D}^i\}$ in both training and test data to get a set of $\langle$predictor, response$\rangle$ tuples. Given time-series $\mathbf{x}$ with k values, the transformation of length l results in a set of tuples of the form $\langle (x_{t-l+1}, ..., x_t), x_{t+h} \rangle$ such that $l \leq t \leq k - h$.

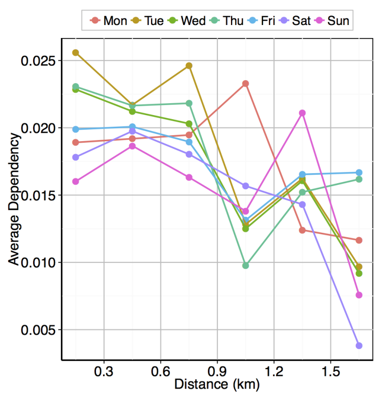
The prediction model for a sensor s_i is a regression tree (Breiman et al. 1984) that uses predictors from all sensors with non-zero coefficients in the dependency matrix learned from a similar day, i.e., predictors are taken from $\{\mathcal{D}^k\}, \forall k : \mathcal{M}_{sim}[i, k] \neq 0$. Since $\mathcal{M}[i, i] = 0$, sensor s_i's own past values are not used as predictors. Hence, *a key benefit of this model is that we are able to do predictions for a sensor in*



(a)

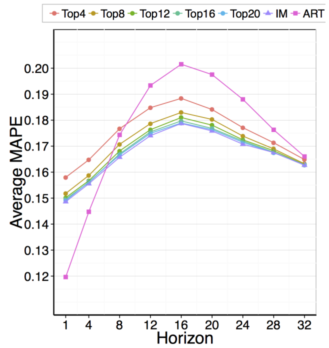


(b)

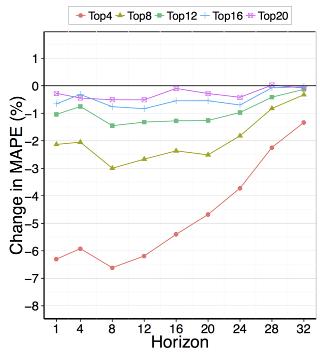


(c)

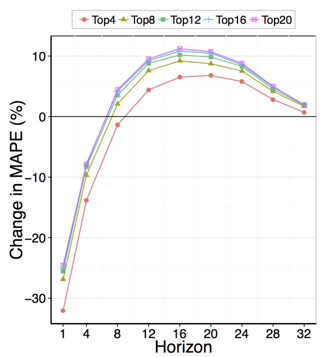
Figure 2: Influence/dependency w.r.t (a) time, (b) size, and (c) distance: higher values observed for weekdays than for weekends.



(a)

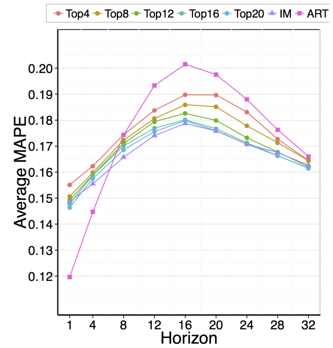


(b)

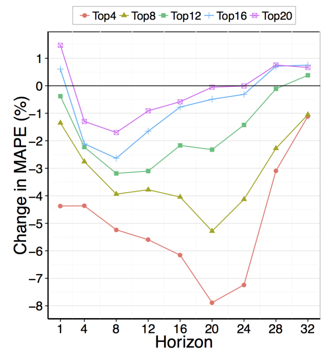


(c)

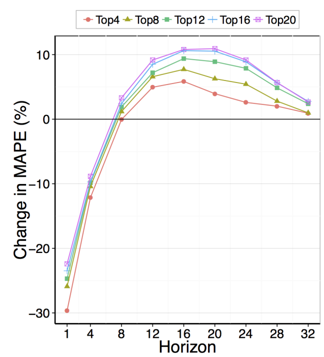
Figure 3: LIM's performance: (a) MAPE; and percentage change in MAPE with respect to (b) IM and (c) ART. (Negative change implies increase in error.)



(a)



(b)



(c)

Figure 4: GIM's performance: (a) MAPE; and percentage change in MAPE with respect to (b) IM and (c) ART. (Negative change implies increase in error.)

absence of its own past values by using past values of its influential sensors.

Local Influence Model (LIM)

In the previous section, we discussed how IM resolves the problem of partial data availability by using influential sensors (a small subset of sensors) to transmit data in real time. However, without restricting the number of influential sensors, the subset of influential sensors considered by IM may include the total number of sensors. Next, we discuss a policy to ensure that only a fraction of sensors is considered for real-time predictions. For each sensor s_i, we sort the corresponding row $\mathcal{M}[i,]$ in the dependency matrix and consider only readings from the top τ_l sensors in this model.

Global Influence Model (GIM)

In LIM, because local influencers are selected for *each* sensor, overall it may still require real-time data from a large number of sensors, thus defeating the goal of getting real-time data from only a few influencers. Thus, we are interested in finding *global* influencers. Using dependency matrices $\mathcal{M}_j$, we calculate daily influence $\mathcal{I}_j^i$ for each sensor s_i as described in equation 1. After sorting the sensors based on their influence values, we consider only readings from the top τ_g sensors in the influence model.

Experiments

Datasets

1) Electricity Consumption Data[2]: collected at 15-min intervals by over 170 smart meters installed in the USC campus microgrid (Simmhan et al. 2013) in Los Angeles.
2) Weather Data: temperature and humidity data taken from NOAA's (NOAA 2013) USC campus station, linearly interpolated to 15-min resolution.

Performance Comparison

We evaluate our models for up to 8 hour-ahead prediction[3]. Given the short horizon, the length of previous values used was set to 1-hour. Out of two choices of similar day for training, previous week and previous day, we found previous week to perform better.

Baseline Model We use the Auto-Regressive Tree (ART) Model as the baseline. Our proposed models are also based on regression tree concept so ART provides a natural baseline to compare performances. ART uses recent observations as features in a regression tree model and has been shown to offer high predictive accuracy on a large range of datasets (Meek, Chickering, and Heckerman 2002). ART's main advantage is its ability to model non-linear relationships in data, which leads to a closer fit to the data than a standard autoregressive model. We implement a specialized $ART(p, h)$ model that uses recent *p* observations of a variable for making *h* interval ahead prediction. While ART uses a variable's own recent observations, our models only use other variables' observations to make predictions.

Evaluation Metric We used MAPE (Mean Absolute Percentage Error) as the evaluation metric, as it is a relative measure and therefore scale-independent (Aman, Simmhan, and Prasanna 2014). MAPE $= \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - \hat{x}_i|}{x_i}$ where x_i is the observed value and $\hat{x}_i$ is the predicted value.

Influence Variation

Fig. 2(a) shows influence variation for the top 4 influencer sensors. Given this variation, we decided to *re-calculate influence for each day* in our experiments, rather than use a static value calculated over a large number of days. Fig. 2(b) shows the distribution of influence for each sensor with the size of sensor readings. It is interesting to note that buildings with smaller consumption values have higher influence.

[2]Available from the USC Facilities Management Services.

[3]Smart Grid applications such as Demand Response usually require up to 6 hours ahead predictions (Aman, Simmhan, and Prasanna 2014).

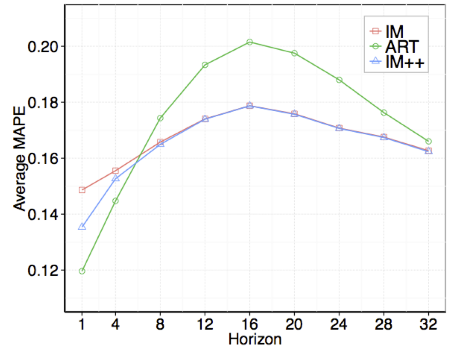


Figure 5: Prediction performance: partial data vs. complete data. For ART, recent values used as predictors at the time of prediction become increasingly ineffective for longer horizons, when IM's use of more recent real-time values of other sensors become more useful.

Also, influence for weekdays is higher possibly due to more activity and movement of people between buildings.

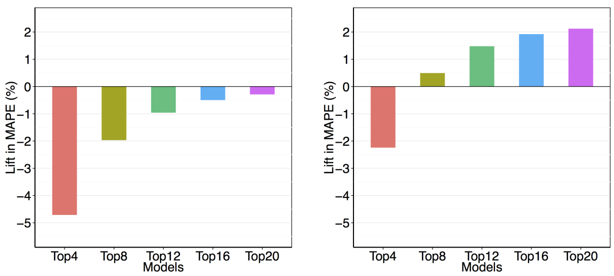
Fig. 2(c) shows average dependency decreasing with increase in the distance between the sensors. This validates our intuition about greater dependency among closely located sensors. This can be attributed to greater movement of people between neighboring buildings, and hence greater dependency in their electricity consumption. Also, there is more movement on weekdays, hence we observe that average dependency is higher for weekdays than for weekends.

Prediction Performance

Fig. 5 shows prediction errors of the *influence model*, averaged over all days and for all sensors. ART performs well up to 6 intervals (1.5 hour), as due to the very short prediction horizon, electricity consumption is not expected to drastically change from its previous 4 values. ART performs well as it has access to real-time data. Instead, IM achieves comparable accuracy despite the lack of real-time data. IM's accuracy also increases with the prediction horizon, where it consistently outperforms ART. While increase in IM's error is subdued, ART's error increases rapidly with increasing horizon implying that the previous 4 values used as predictors at the time of prediction become increasingly ineffective for predicting values beyond 1.5 hours ahead in time. Here, *more recent real-time values of other sensors become more useful predictors than a sensor's own relatively older values*. That IM achieves good accuracy despite the lack of real-time data is an important result and its main advantage.

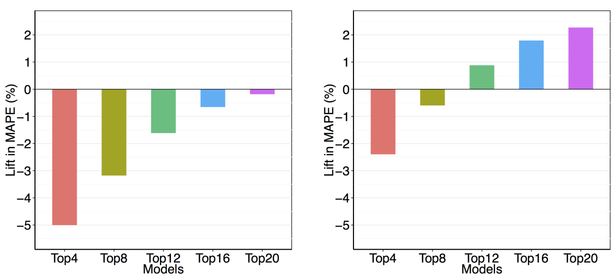
For comparison, we used an additional baseline (shown as IM++ in Fig. 5), where we include both other sensors' real-time data as well a sensor's own real-time data. IM++ outperforms IM initially, but beyond 2 hours its performance is similar to that of IM. This implies that even in absence of a sensor's own data, IM can achieve same results as when this data is available, again indicating its advantage.

For *local influence model* (Fig. 3(a)), we consider real-time values from top τ influential sensors for each sensor ($\tau = 4, 8, 12, 16, 20$). ART performs well initially due to very short prediction horizon, but its errors increase rapidly



(a) (b)

Figure 6: Lift in MAPE for LIMs w.r.t (a) IM and (b) ART. Positive lift is observed w.r.t. ART beyond Top 8. (Positive lift indicates reduction in MAPE.)



(a) (b)

Figure 7: Lift in MAPE for GIMs w.r.t (a) IM and (b) ART. Positive lift is observed w.r.t. ART beyond Top 12. In (b) Only $\approx 0.5\%$ increase in prediction error over ART is witnessed while using just top 8 ($\approx 7\%$) of smart meters.

with increasing horizon. *The LIMs show performance comparable to IM, while using real-time values from fewer sensors.* Using increasingly fewer predictors increases the prediction error for LIMs, but only slightly. LIM's performance deteriorates compared to IM in terms of percentage change (Fig. 3(b)) with increasing horizon. This can be the effect of very few sensors remaining influential over longer horizons. When averaged over all horizons, we observe 4.71% increase in error compared to IM for Top 4 model which comes down to 1.97% increase for Top 8 and less than 1% increase for Top 12, 16, and 20 models (Fig. 6(a)). We observe that beyond 1-2 hour horizon, all LIMs outperform ART (Fig. 3(c)) as for ART, the *effective horizon* now includes the prediction horizon and the unavailable real-time data. When averaged over all horizons, we observe that for Top 4, there is an increase in error by 2.24%, but for Top 8 (and Top 12, 16, 20), the error actually decreases (Fig. 6(b)) with respect to ART. Thus, we conclude that for this dataset, we need at least 8 influential sensors for each sensor to improve performance over the baseline.

The *global influence model* uses real-time values from only top τ influential sensors selected globally for *all* sensors ($\tau = 4, 8, 12, 16, 20$) GIM outperforms ART beyond 8 intervals (Fig. 4(a)). However as the number of predictors is reduced when moving from Top 20 to Top 4 model, we observe that increase in errors is more pronounced for GIM (Fig. 4(a)) than for LIM (Fig. 3(a)) as the number of unique influential sensors is significantly lower in the case of GIM as compared to LIM and IM. *While LIM used influential sensors selected separately for each sensor, GIM uses the same set of influential sensors for all sensors and still achieves comparable performance*. Top 20 and Top 16 GIMs outperform IM (Fig. 4(b)) for 1 interval ahead and later for 28 and 32 intervals. This could be due to the large number (20 and 16) of predictors selected in these models overlapping with those of IM. This is further supported by the average result over all horizons, where both Top 20 and Top 16 models show less than 1% increase in errors compared to IM (Fig. 7(a)). We also observe that all GIMs outperform ART beyond 12 intervals, i.e., 3 hour horizon (Fig. 4(c)). When at least 12 influential sensors are available, improvements are observed over the baseline across all horizons (Fig. 7(b)). We also found that as compression ratio (eqn. 2) was increased from 5 to 30, increase in MAPE was only by $\approx 1\%$. *GIM is able to provide a practical solution using real-time values from only a small fraction of sensors, thus achieving great compression ratio.*

Conclusion and Future Work

We address the *partial data problem* in sustainability domain applications that arises when data from all sensors is not available at central nodes in real time, either due to network latency or data volume, or when transmission is limited by the consumers for security and privacy reasons. Standard models for short term predictions are either unable to predict or perform poorly when trying to predict with partial data. We propose novel *influence* based models to make predictions using real-time data from only a few influential sensors, and still provide performance comparable to or better than the baseline. Thus, we provide a practical alternative to canonical methods - which assume real-time data availability for all sensors - for dealing with unavailable real-time readings in sensor streams. These models are generalizable to applications in several domains, and provide a simple and interpretable solution that is easy to understand and apply for domain experts.

Future extension of this work is towards a two stage process for influence discovery, guided by heuristics, and by a combination of local and global selection of influential sensors to further improve prediction performance. Another direction of research is for scenarios when time series data from different sensors is not sampled at equally spaced time intervals resulting in irregular time series.

Acknowledgments

This material is based upon work supported by the United States Department of Energy under Award Number DEOE0000192, and the Los Angeles Department of Water and Power (LA DWP). The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, the LA DWP, nor any of their employees.

References

Aman, S.; Simmhan, Y.; and Prasanna, V. 2014. Holistic measures for evaluating prediction models in smart grids. *IEEE Transactions in Knowledge and Data Engineering* PP(99).

Arnold, A.; Liu, Y.; and Abe, N. 2007. Temporal causal modeling with graphical granger methods. In *International conference on Knowledge discovery and data mining (KDD'07)*. ACM.

Bahadori, M. T., and Liu, Y. 2012. Granger causality analysis in irregular time series. In *SIAM International Conference on Data Mining (SDM 2012)*. SIAM.

Balac, N.; Sipes, T.; Wolter, N.; Nunes, K.; Sinkovits, R. S.; and Karimabadi, H. 2013. Large scale predictive analytics for real-time energy management. In *IEEE International Conference on Big Data*.

Bouhafs, F.; Mackay, M.; and Merabti, M. 2012. Links to the future: communication requirements and challenges in the smart grid. *IEEE Power and Energy Magazine* 10(1).

Box, G. E. P., and Jenkins, G. M. 1970. *Time series analysis, forecasting and control*. Holden-Day.

Breiman, L.; Friedman, J.; Stone, C. J.; and Olshen, R. A. 1984. *Classification and regression trees*. CRC press.

Bullis, K. 2014. Smart wind and solar power. *MIT Technology Review*.

Chong, C., and Kumar, S. P. 2003. Sensor networks: Evolution, opportunities, and challenges. *Proceedings of the IEEE* 91.

Ciancio, A., and Ortega, A. 2005. A distributed wavelet compression algorithm for wireless multihop sensor networks using lifting. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*.

Cuevas-Tello, J. C.; Tiňo, P.; Raychaudhury, S.; Yao, X.; and Harva, M. 2010. Uncovering delayed patterns in noisy and irregularly sampled time series: an astronomy application. *Pattern Recognition* 43(3).

Deshpande, A.; Guestrin, C.; Madden, S. R.; Hellerstein, J. M.; and Hong, W. 2004. Model-driven data acquisition in sensor networks. In *International conference on Very large data bases (VLDB 2004)*.

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3):424–438.

Karimi, B.; Namboodiri, V.; and Jadliwala, M. 2013. On the scalable collection of metering data in smart grids through message concatenation. In *IEEE International Conference on Smart Grid Communications (SmartGridComm)*.

Kreindler, D. M., and Lumsden., C. J. 2006. The eects of the irregular sample and missing data in time series analysis. *Nonlinear dynamics, psychology, and life sciences* 10(2).

Kwac, J., and Rajagopal, R. 2013. Demand response targeting using big data analytics. In *IEEE International Conference on Big Data*.

Lozano, A. C.; Li, H.; Niculescu-Mizil, A.; Liu, Y.; Perlich, C.; Hosking, J.; and Abe, N. 2009. Spatial-temporal causal modeling for climate change attribution. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*. ACM.

Marascu, A.; Pompey, P.; Bouillet, E.; Verscheure, O.; Wurst, M.; Grund, M.; and Cudre-Mauroux, P. 2013. Mistral: An architecture for low-latency analytics on massive time series. In *IEEE International Conference on Big Data*.

McDaniel, P., and McLaughlin, S. 2009. Security and privacy challenges in the smart grid. *IEEE Security and Privacy* 7(3).

Meek, C.; Chickering, D. M.; and Heckerman, D. 2002. Autoregressive tree models for time-series analysis. In *2nd International SIAM Conference on Data Mining (SDM)*. SIAM.

NOAA. 2013. Quality Controlled Local Climatological Data Improvements/Differences/Updates.

Pan, B.; Demiryurek, U.; and Shahabi, C. 2012. Utilizing real-world transportation data for accurate traffic prediction. In *IEEE International Conference on Data Mining (ICDM'12)*.

Razzaque, M. A.; Bleakley, C.; and Dobson, S. 2013. Compression in wireless sensor networks: A survey and comparative evaluation. *ACM Transactions on Sensor Networks* 10(1).

Sanders, P.; Schlag, S.; and Muller, I. 2013. Communication efficient algorithms for fundamental big data problems. In *IEEE International Conference on Big Data*.

Simmhan, Y.; Aman, S.; Kumbhare, A.; Liu, R.; Stevens, S.; Zhou, Q.; and Prasanna, V. 2013. Cloud-based software platform for data-driven smart grid management. *IEEE Computing in Science and Engineering*.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58(1).