EDI-Net: A smarter approach to city-scale energy management

Graeme Stuart, Carl Holland
& Paul Fleming
De Montfort University
Leicester
UK

Keywords

data, data monitoring, energy analysis, communication, community

Abstract

Energy metering data is becoming 'smart' – sub-hourly and downloaded at regular intervals in a day – but the software which is used to display such data remains 'dumb'. Such software relies on the experience of the analyst to identify instances where consumption is outside of a typical pattern. The analyst must trawl through data looking for problems. Though alarms can be set in 'dumb' software – these alarms are also dumb – typically an energy manager sets maximum and minimum thresholds and is notified when they are exceeded.

This paper presents 'smart' software and advocates for a more innovative approach to energy management. The software maintains a detailed model of consumption patterns across a whole portfolio of buildings and communicates energy performance in a user friendly way. Rather than simply alarming on peaks and troughs it can track unexpected levels of consumption and manage a list of exceptions. The approach opens up energy management to a wider community of stakeholders through the use of simple visualisation, coherent information architecture and a flexible communication platform.

It is now possible to analyse large volumes of consumption data automatically with software, notifying the energy manager only when something unexpected occurs. With a continually evolving model of consumption patterns across the entire portfolio smart software can present stakeholders with interactive, high-level views on the performance of the portfolio as a whole whilst allowing a user to 'drill' down to see the detail on demand. Reports can be designed for energy experts or for other professionals such as finance managers and decision makers enabling better communication across an organisation and more effective energy management.

The Energy Data Innovation Network (EDI-Net) project is developing an approach to energy management which engages a community using modern, scalable communication tools with an ambitious approach to modelling, analysis and visualisation of large datasets.

Introduction

In a smart city, information is available to those who need it, when they need it and in a form they can use. Smart energy efficiency in buildings is no exception. The Energy Data Innovation Network (EDI-Net project) aims to increase the capacity of public authorities to make use of high-resolution energy consumption data by developing scalable tools for the collection, modelling, analysis and visualisation of near real-time energy consumption data and providing a platform for participating authorities to build communities of stakeholders around this important and accessible source of information.

The potential for energy efficiency in the non-domestic sector is well known and the public sector accounts for a significant proportion of this opportunity as well as having an important role in demonstrating best practice. Data collection and analysis has become increasingly important in supporting effective energy efficiency in such organisations. This paper is concerned specifically with energy management in European public authorities such as municipalities, in particular, those with a large portfolio of buildings. It describes an approach to data analysis designed to increase the capacity of an organisation to deliver energy efficiency in buildings.

Advances in metering and data management over the last few decades have led to a massive increase in the availability, quality and resolution of energy and water consumption data. It is still extremely common for buildings to be manually monitored at a resolution of 4–12 readings per year (quarterly or monthly) for billing purposes. However, it is also now fairly common to see datasets available with 17,520 or 35,040 readings per utility, per year (half-hourly or quarter-hourly). This increase in available data is a trend which is unlikely to reverse and it is reasonable to expect that in decades to come such data will become the norm in both non-domestic and domestic buildings.

The increase is largely driven by the need for high quality data. Organisations are installing their own automatic meter reading (AMR) systems in order to access the benefits of these data. Leicester City Council, in the East Midlands region of the UK were early adopters of this technology back in 2002 and now have 15 years of experience with this kind of data. Access to such data transforms what can be achieved with data analysis. Previously invisible wastage is now in many cases extremely obvious from a quick scan of the data (Ferriera et al. 2007).

However, a problem many organisations now experience is being overwhelmed by the sheer volume of data generated every day. Monitoring systems require that data are reviewed regularly to identify problems early. Reviewing an entire building portfolio in a large municipality is a significant task. For example, a municipality with 450 buildings and 1,350 (gas, water and electricity) meters will produce 64,800 new data points per day at half hourly resolution. This amounts to a total of 23,652,000 data points per year.

The reality of managing large numbers of buildings is that, even if an analysis of one building takes only a few minutes to conduct, it can take weeks or months to systematically review all the available datasets. If it is possible to conduct 30 analyses in a working day then it would take roughly nine weeks of uninterrupted work to complete a systematic review of 1,350 datasets. Shaving time off the individual analyses will reduce the total time by a corresponding proportion. Each additional minute on the individual analysis time adds 1,350 minutes to the total time! Thus, the speed of a single analysis will affect how often an individual dataset is looked at.

The time it takes to conduct an analysis of a single dataset depends on the depth of the analysis being conducted. It may take one or two minutes to access the data from a meter and to perform a cursory glance at the consumption profile. A more in-depth analysis looking at what is 'normal' for the given building and comparing recent consumption with historical trends may take twenty minutes or longer. The time taken is strongly influenced by both the software in use and the depth of the analysis. Thus, there is often a choice between depth of analysis and frequency of analysis and the available software can have a significant influence on this balance.

Clearly access to automatically generated, high resolution data has huge advantages. The ability to react quickly to faults in buildings is part of this. If each dataset can only be reviewed once in nine weeks then much of the benefit of high resolution data is lost. To realise the benefits of automated data collection requires a systematic approach to automated data analysis. There are obvious benefits of a continuous monitoring scheme able to identify datasets which require further analysis and to avoid the need to review the very large number of datasets for which no new information would be obtained.

This paper presents a brief review of standard industry practice, followed by a description of a new approach to energy data analysis under development in the Horizon 2020 Energy Data Innovation Network (EDI-Net) project.

Data analysis and visualisation

Data analysis in energy management provides insight into the opportunities for energy efficiency savings. It can help to isolate individual buildings from a portfolio where further investigation is required. This may be via a process of looking at consumption amongst similar buildings to identify outliers or it may be via a process of identifying buildings where the pattern of consumption has changed. These changes can have many causes. They may be associated with a failure of some kind (e.g. controls), a change of operation (e.g. hours of use) or an energy efficiency intervention. In all cases the identification of a change is the beginning of a deeper investigation into the data and the circumstances in the building. These investigations are the precursors to intervention and can lead to improved efficiency and reduced wastage.

Data collection and analysis for energy management purposes is well established. Standard approaches to data analysis are described in a large body of published guidance for energy management. The majority of these publications fall into two distinct categories. In the UK monitoring and targeting (M&T) is prevalent whilst measurement and verification (M&V) was developed in the USA. The two traditions are distinct but have many similarities (Stuart 2011). In general M&T is more exploratory; using consumption models and event-detection techniques to identify signs of wastage. M&V has a more limited scope and is more quantitative; using consumption projections to quantify the effects of known projects.

Monitoring and targeting is the process of developing an expected level of consumption for a building and continually checking to see if consumption diverges from the expectation. Expectation may be established via benchmarks (e.g. annual kWh per m^2) determined from a wide sample of similar buildings or it may be established by looking at historical short time series data for the given building. Since we are monitoring consumption at high resolution, the latter is the focus of this paper.

Historically the source data for such an analysis consists of monthly consumption figures, collected either from bills or manual meter readings and monthly degree days published for the region. A simple regression model of consumption against degree-days provides a means for basic consumption patterns to be quantified and tracked over time (Carbon Trust, 2012). The regression calculation provides the expectation for a given month (with given degree-days) and this can be compared with the actual measured consumption for the month. A systematic comparison can be made using the CUSUM method to study divergence from the model. Model residuals (the difference between actual and predicted consumption) are accumulated over time to highlight any persistent divergence from expectation. Of course, when dealing with monthly data any insights gained from analysis will only be confirmed after several months of data have been collected and incorporated into the analysis. Furthermore, monthly resolution data will give very little information about the nature of the divergence (e.g. the time of day/week when the divergence occurs). Though usually applied to monthly data, the CUSUM approach can be applied to higher resolution data with good results (Stuart et al. 2007).

A portfolio of 450 buildings requires the careful management of 1,350 datasets. If this is done manually (e.g. in spreadsheets) then there is a great risk of human error causing problems with the raw data even before any analysis occurs. The raw data are in any case vulnerable to data quality issues caused by the timing of readings, the accurate manual recording of timestamps and reading values and the use of estimated readings. Data quality problems can be the dominant feature of a dataset. Thus any analysis requires careful consideration. Even with good quality data, the variation in monthly consumption can lead to unstable consumption models and produce erroneous conclusions if not considered carefully. Such analyses require systematic reviews by experienced analysts (Stuart 2011).

Analysis software

Software have been available for decades to collect and analyse monthly data, typically collected manually, either directly from meters or indirectly from bills. Many proprietary software packages are available and most are continuously evolving to meet the needs of users. Desktop packages are gradually transitioning into web-based systems and data management is moving into the cloud.

A full review of the available software is beyond the scope of this paper. However, they fall into two main categories. Those intended for use with monthly billing data and those designed for managing and analysing the new, higher resolution data. In the last 10 years or so, software tools have been adapted to handle the new higher resolution data. These datasets have made it essential to automate the process of data collection and management at the source (typically directly communicating with AMR systems) which has led to many new proprietary systems being developed specifically to provide the energy manager with convenient access to these raw data and some simple analyses to make these datasets more usable.

Automation with software reduces some of the risk of human error and certainly reduces the workload associated with managing and processing large volumes of data. Basic outputs include simple reports showing consumption data plotted over time. Simply providing convenient access to visualisations of the raw data is the primary output and mode of operation of many packages. Some systems allow comparisons between similar buildings (e.g. as annual consumption per unit of floor area). Grouping buildings by type and comparing annual consumption with benchmarks provides a simple means to identify buildings which are more energy intensive than their peers.

More advanced reports include scatter plots of consumption against degree days and the fitting of regression models. These provide more insight into the detailed consumption patterns in each building and highlight the variation around those patterns. In particular, CUSUM charts allow changes in consumption patterns in a building to be isolated. Typically, this kind of analysis is considered an advanced feature and requires significant configuration including synchronisation with published degree day data. These analyses are typically produced at a monthly resolution.

Data quality is a serious issue with monthly billing data mainly due to the human error involved in the data collection and processing. High resolution data bring more problems with data quality, especially missing data. As readings are often being collected for many hundreds of meters it is necessary to highlight missing data automatically. AMR systems can rely on batteries and cables, these are prone to fail or be unplugged over time leading to periods of missing data. Systems often provide import alarms as a feature whereby if a dataset doesn't provide the expected readings they are highlighted. This allows the analyst to trigger an investigation to get the data flowing again.

Extending this somewhat is the ability to set 'alarm limits'. The user will be warned if consumption for a given meter falls outside of these limits. High consumption may indicate there is waste occurring in the building, low consumption could indicate a fault in equipment or in the meter reading system itself. As an energy management tool these alarm systems can only really provide a 'sanity check' and a fallback to catch extreme values. In practice, these systems produce such large numbers of warnings and many of these warnings prove to be unhelpful. Consumption can vary a lot in some buildings so the alarm limits must be set at the most extreme expected value. This reduces the number of alarms to a manageable level but also renders the alarms fairly useless, only triggering alarms in extreme circumstances. Users will either widen the limits until false positives are virtually eliminated or will learn to ignore the alarms. In any case, a small but persistent increase in consumption will be ignored completely by these alarms even though it may represent a very large amount of total additional consumption.

With the new high quality data many new kinds of analysis are possible and a new set of standard techniques is slowly emerging. A key aspect of this is looking at time of use. Consumption during unoccupied periods is a tell-tale signal that something is wrong. High resolution data can easily reveal problems such as poorly configured control systems. Highlighting this kind of problem is very easy by simply looking at raw data.

One of the reviewed systems generates a report based on the proportion of daily consumption which occurs in unoccupied periods. Calculating this percentage every day and presenting the results in a daily 'league table' of buildings is useful as a guide to highlight control problems. The analyst can use the league table to allocate their time disproportionately towards buildings which are likely to provide opportunities for action. If consumption occurs in a building overnight then it will rise to the top of the list and the problem can be investigated and potentially resolved straight away.

Analysis such as this, applied across multiple buildings, has enormous potential to provide the analyst with a guide as to which datasets should be prioritised and which datasets can be safely ignored. Thus expanding the time between analyses for some buildings (where nothing has changed since the last analysis) and shrinking the time between analyses for others (where things are changing). It also, depending on the data collection regime and the software in place, has the potential to bring the analysts attention directly to buildings where faults have occurred within a few hours or minutes of the fault being recorded.

The EDI-Net approach

The approach described in this paper has been developed in the current EU EDI-Net project. The analysis approach is an evolution of the standard practice of building consumption models and analysing divergence from those models. However, it takes a 'whole portfolio' view by producing consumption models for all available datasets and continuously updating the analysis automatically. In this way, the analyst can easily view a report of all available datasets highlighting those which may need further attention drawing attention away from those with steady performance.

The approach is based on work done by the EU smartspaces project (Stuart et al 2013). In particular by the Leicester pilot site of the smartspaces project. The smartspaces project introduced a different approach by automating the data analysis. As new data were collected they were automatically analysed to generate baseline models, predictions and performance indicators. The results of this analysis were then processed further to generate summary reports. This process of analysing the results of analysis allowed for an extremely simple user interface accessible to non-experts which also helped experts to navigate to the areas of their data which required further investigation. The final aspect of smartspaces was the ability to consult with stakeholders, sharing ideas and knowledge. An online forum was used to facilitate this, its operation is described in (Stuart et al 2016).

The EDI-Net project aims to take this pilot system from the proof of concept stage (the smartspaces system included only 25 buildings) to a fully developed, scalable system capable of delivering services to thousands of buildings. The primary output of the analysis is identifying buildings and datasets which require further investigation. Secondary to this is the provision of diagnostic reports which indicate why the dataset in question was considered noteworthy.

The approach also expands the audience to a wider group of stakeholders including building users, financial managers and decision makers. It does this by using simple visualisations in summary reports and by employing an information architecture which hides details under layers of user interface. Though the technical experts (energy managers and building operators) are still the primary audience and can access the most detailed technical reports whenever they wish, it is still possible to access summary information without ever looking at the detailed diagnostic reports. For technical experts, this simple user interface helps direct their attention to the most appropriate datasets, for the non-technical audience it provides a comprehensible overview.

THE EDI-NET MODEL

The focus of the EDI-Net approach is to provide energy professionals with the ability to track energy performance across a large portfolio of buildings and to enable sharing of this information with stakeholders in a non-technical, user-friendly manner. To do this the EDI-Net software employs a sophisticated consumption model and a systematic analysis methodology.

The core of the approach is a data analysis which relies on a statistical, data-driven model of consumption. The main areas where the analysis departs from standard practice is in the sophistication of the model used, the systematic application of that model to meet the requirements detailed above and the careful information design and information architecture which allows the user to choose the level of complexity they are comfortable with and to easily understand what they are being shown.

The consumption model is applied at a half-hourly resolution, accounting for both the weekly occupancy pattern (the so-called 'profile') and the effects of outside air temperature. So, the system is able to update the model and outputs every half hour and can produce half-hourly predictions. The model is fitted to each dataset automatically. Rather than viewing datasets one at a time, we attempt to aggregate many many datasets and models to produce a detailed model of an entire building portfolio at building-by-building granularity.

The model itself is composed of a series of piece-wise regression models of consumption against outside air temperature. One of these sub-models is fitted for each half-hour in the week. In this way we generate 336 sets of model parameters and when predicting consumption we combine outside air temperature with the appropriate model parameters. When fitting models we calculate and store the model residuals as a measure of the 'scatter' around the model.

The system currently has two primary modes of operation. In the fixed baseline mode a consumption model is fitted to a fixed baseline period (e.g. one year). Predictions generated by this baseline model are generated for the whole dataset and compared to the actual measured consumption. In this case we can do calculations such as cumulative difference since the baseline period. We can also look at the variation in consumption within the baseline period to gauge whether savings are significant or whether they fall into the usual range. Larger buildings have more chance of being recognised in this mode of operation because they have greater potential for changes in consumption levels. This analysis is the basis of the evaluation of energy savings achieved by the EDI-Net project.

The system can also run with a rolling baseline model. In which case the baseline period is updated continuously (the baseline model always uses 12 months of data, ending at midnight on the most recent Sunday) and each model is used to generate predictions of consumption for a single week. The results are aggregated into a time series which is again compared to current consumption. The model residuals are also calculated each week and compared to the prediction residuals to determine where consumption in the current week falls within the spread of consumption predicted by the model. In this case we calculate a half-hourly performance indicator as the main output. The detailed calculation is described in (Stuart and Fleming 2014).

The indicator always falls between zero and 100. An indicator of 50 implies consumption is in the middle of the expected range (based on the most recent 12-months of data). In fact it implies that – after correcting for outside air temperature – half of observed consumption levels in the baseline period were above this value and half were below. Similarly, a value of 10 or 90 imply that 90 % or 10 % of equivalent consumption values in the baseline period were above the current level. In this way, we can use the value to represent energy performance. A value of 0 % is lower than any in the baseline period. A value of 100 % is higher than any in the baseline period.

This approach has multiple advantages. Firstly, it avoids the need to present results in units such as kWh which may be unfamiliar to our chosen audience. This allows the results to be interpreted by those stakeholders without specialist interests in energy efficiency in buildings. We can easily convert the simple scale into good, neutral and bad 'zones' of consumption (e.g. less than 25 % is good, greater than 75 % is bad, intermediate values are neutral). These zones can also be visualised as simple visual scales (such as smiley/sad faces) to remove the need for presenting numbers in the user interface.

We can also summarise these values by taking averages over time and even across meters and buildings. The raw indicator is calculated at half-hourly intervals but can be aggregated at daily or weekly resolution as necessary. The method of aggregation (e.g. mean verses median) will affect the result and can also be tuned to identify extreme short term divergence or small, persistent changes. Summaries of the indicator can be used to order league tables and highlight those buildings whose energy consumption is currently high or low relative to the norm.

Discussion

The EDI-Net project moves the approach developed in the smartspaces pilot from 25 buildings in one city to 40 entire building portfolios with an expected total of up to 5,000 buildings. The major benefits of the approach are more obvious at this larger scale. The approach expands beyond looking at individual buildings and places the focus on the portfolio as a whole. Having access to the entire portfolio allows the system to optimise analyst time, ensuring that time spent looking at data is mostly focussed on looking at the right data from buildings where there are opportunities to reduce waste. Sharing this information across multiple stakeholders multiplies the benefits further.

The EDI-Net software automatically fits a complex consumption model to a large volume of data. It stores model parameters, predictions of consumption and other statistics so they are available for further analysis. Summary reports are produced at individual dataset level and aggregated across buildings. This has profound implications on the system architecture. As data are continuously input into the system, analytical services constantly update model parameters, generate predictions and calculate model residuals and performance indicators. This results in a large volume of half-hourly results. In fact, the results generated for each dataset are many times the size of the original dataset. Maintaining an up-to-date complex model of an entire building portfolio and this huge pool of detailed analysis results is a new paradigm in energy management analytics requiring a scalable and high performance architecture.

The model is not useful on its own. It can provide a certain type of technical information in terms of a comparison of model parameters between and within building type categories. However, the model becomes truly useful when it is applied as a tool to meet the requirements of energy managers. The approach taken is to generate the most detailed and complex results on a dataset-by-dataset basis and then summarise them across time and across buildings to produce a simple summary of energy performance across a portfolio.

These results contain crucial information which can help identify datasets which are more likely to yield opportunities. For example, we can easily identify datasets where consumption has increased beyond historical patterns. To help direct the user to the most interesting datasets a summary of these results is provided as the primary report. A league table showing the best performing buildings at the top and the worst at the bottom. The user can then use this report to navigate to a dataset of interest and see the detailed results. In this way, the detailed results are made available to the user interface but only some will be viewed, typically those with significant increases or decreases in consumption.

Buildings are listed in league tables and ordered in various ways according to the recent or long-term values of these performance indicators. As such, the user is given a very simple way to identify from amongst their large building portfolio, those buildings whose consumption is increasing beyond expected levels.

Conclusion

Automating data analysis can complement the automation of data collection. Whether analysing monthly or half-hourly data the standard approach to analysis only allows a small proportion of the available datasets to be analysed per day. Thus, reviewing a large portfolio can take a long time. Using software to help manage data and to generate automated reports can improve this rate and can improve the accuracy and consistency of the analysis but traditional, ad-hoc analysis is slow and labour intensive which makes it prone to human error. With this approach, much of the analyst's time is spent confirming that no action is required.

Introducing systems capable of automatically generating analyses of individual datasets and automatically analysing the combined results has huge potential. Rather than conducting analysis on demand, these calculations are conducted automatically and continuously. This enables the results to be analysed as a whole and so an analysis of an entire building portfolio can be presented to the user on demand. This analysis at the portfolio level is the key to increasing the effectiveness of analyst time since they no longer need to trawl through hundreds of datasets which have not changed since the last time they checked and which offer no new information.

Summary reports help to prioritise the detailed analyses such that buildings with potential problems are investigated earlier. This has the effect of reducing the time between reviews for problematic buildings and improves the efficiency of the monitoring system as a whole. A side benefit is that pre-calculated diagnostic reports can be presented at the click of a button. A convenient interface can be designed by making the league table items link directly to detailed reports.

The EDI-Net project, building on the findings of the smartspaces project is implementing this approach as working software. Moving the analysis from an ad-hoc, on-demand process into an ongoing, automated process. The intelligent and systematic monitoring and targeting process allows those datasets which are unremarkable and unchanged to be effectively ignored in the day-to-day search for opportunities to reduce waste.

By formalising and fully automating the modelling step in the analysis it is possible to produce a continually updating statistical model for the entire building portfolio. As new data are collected they can automatically be compared to a prediction of consumption. Reports can then be produced which summarise these findings in convenient forms. The analyst can use these reports to navigate the huge volumes of data, only looking in detail at those buildings where problems have already been detected.

The opportunity presented by this approach is to design further layers of processes which analyse the results of analysis and produce notifications to the analyst, pointing at datasets and pre-calculated analysis which demonstrates the reasons for the notification. Automatically generating further analyses when problems are identified. This represents an entirely new paradigm for 'smart' energy management.

References

Carbon Trust (2012), CTG077: Monitoring and targeting techniques to help organisations control and manage their energy use. The Carbon Trust, March 2012.

Ferriera, V., Alves, L., Fleming, P., Stuart, G., Patel, P., Webber, P. and S. Conway, 2007. "Low hanging fruits" or cost-effective energy and water savings using intelligent metering and monitoring systems? Proceeding of eceee 2007 Summer Study. Presented at the eceee 2007 summer study, eceee.

Stuart, G. and Fleming, P. (2014). Smart energy performance indicators for live historical and normative feedback systems. Proceedings of 8th International Conference Improving Energy Efficiency in Commercial Buildings (IEECB'14), Frankfurt, DOI: 10.2790/32838, pp. 400–414.

Stuart, G., Snape, J. R. and Fleming, P. (2016). Closing the Feedback Loop: A Systems Approach to Supporting Community-wide Behaviour Change in Non-domestic Buildings. Proceeding of ACEEE 2016 Summer Study. Presented at the ACEEE 2016 summer study, ACEEE, Pacific Grove, USA.

Stuart, G., Wilson, C., Bull, R., Irvine, K.N., 2013. Designing live energy performance feedback for public buildings in Leicester, in: Energy Efficiency First: The Foundation of a Low-Carbon Society – Proceeding of eceee 2013 Summer Study. Presented at the eceee 2013 summer study, eceee, Belambra Presqu'île de Giens, France, pp. 839–844.

Stuart, G., Fleming, P., Ferriera, V. and P. Harris (2007) Rapid analysis of time series data to identify changes in electricity consumption patterns in UK secondary schools. Building and Environment. Volume 42, Issue 4, April 2007, Pages 1568–1580.

Stuart, G. (2011) Monitoring energy performance in Local Authorities. PhD thesis. De Montfort University. Available at https://www.dora.dmu.ac.uk/handle/2086/4964 (Accessed: 24 January 2017).

Acknowledgements

The authors would like to recognise the contribution of the European Union to this work under the EDI-Net project. The EDI-Net project has received funding from the European Union's Horizon 2020 framework programme under grant agreement No 695916.