BRIDGING THE ENVIRONMENTAL AND ENERGY PERFORMANCE GAP IN BUILDINGS THROUGH SIMULATION, MEASUREMENT AND DATA ANALYSIS

Matthew Horrigan[1], Edward Murphy[2] and James O'Donnell[1]
[1]School of Mechanical and Materials Engineering, Electricity Research Centre,
University College Dublin, Dublin 4, Ireland
[2]Ollio Consultancy, United Kingdom
Corresponding Author: matthew.horrigan@ucdconnect.ie

ABSTRACT

Buildings during operation are dynamic environments where changes to control strategies and space usage regularly occur. As a result of these and other issues a performance gap between design intent and actual building performance emerges. This paper seeks to address the operational performance gap and optimise building performance by identifying inconsistencies in building operation. This is achieved through the combined use of a calibrated simulation model output data and the implementation of a breakout detection algorithm based on measured time-series building data. A number of alternative breakout detection algorithms are reviewed in comparison to the chosen technique in this paper. This paper outlines a statistically robust methodology to identify breakouts for measured time-series building performance data. Time-series data from the sports centre within University College Dublin (UCD) swimming pool hall were analysed. Unintended shifts in swimming pool return humidity levels above an acceptable level of 70% were identified by the breakout detection algorithm. A calibrated simulation model is referred to and the building system or zone set-point is reset to the recommended optimal value. The engineering value of this process is that it can be run in real time, as time-series data is produced, to accurately identify shifts in performance for building managers and reduce the performance gap.

Introduction

Buildings consume 40% of global energy, 25% of global water, 40% of global resources and emit 33% of global GHG emissions (United Nations Energy Programme, 2011). It is widely recognised that buildings do not perform efficiently with European and American legislators commissioning directives such as the Energy Performance of Buildings Directive (2002/91/EC) (recast 2010) (European Union, 2010) and ASHRAE Standard 90.1 (ASHRAE, 2005) to combat the issue. In spite of these directives, indoor environmental needs of occupants are not always satisfied. Overall, buildings do not perform adequately in terms of environmental performance and associated energy consumption (Scofield, 2012). A solution to building performance inefficiency may lie in the vast quantities of data now being produced by buildings (RIBA and ARUP, 2013)

Buildings are becoming more intelligent, by 2020 an estimated 19.5% of the global building stock will be "Smart Buildings" containing 50 billion devices connected to the Internet of Things (Memoori, 2014). Commercial building managers are beginning to see the merit in analysing building performance data to inform decision making (Granderson et al., 2013). This analysis typically takes the form of visualisations, where building managers try to spot trends in time-series data (Donnelly, 2012; Aberer and Tan, 2009). However, it is difficult to evaluate the effectiveness of visualisations without considering their support in the decision making process of building managers (Khan and Hornbæk, 2011). Building managers may simply not have the time, resources and be equipped with the correct tools and skills to generate insight from building performance data (O'Donnell, 2009).

With global data production increasing at a rate of 40% annually, visual detection of performance problems by building managers is not a practical solution (IDC, 2014). A number of automated analytical processes can aid building managers in detection of performance problems. One such approach is the breakout detection algorithm which identifies problems caused by unintended shifts in building performance. Breakout detection is a form of changepoint detection that identifies when the probability distribution of a time-series changes. From a building performance perspective, the breakout algorithm has only been used once previously in a model calibration context (Miller and Schlueter, 2015). The algorithm was used to identify macro level changes in eight years of historical power meter time-series data. The algorithms focus was to determine continuous areas of performance that are similar and transition performance periods, with a minimum time-span of six months allowed between the detection of subsequent breakouts.

Dynamic building energy models are extensively used to model the environmental and energy performance of buildings. However, these models are developed during design but are typically not updated for use in operation (Coakley et al., 2016). Therefore the methodology proposed in this paper combines short term forecasts of building performance, represented by calibrated simulation output data, and measured time-series data used in a scalable breakout detection algorithm to identify unintended shifts in building performance.

The objective of this paper is to minimise the performance gap and optimise indoor environmental performance through the combined use of data analysis and simulation. A statistically robust methodology is presented to identify breakouts in time-series building data. The effectiveness of this statically robust methodology is demonstrated by means of a case study that analyses return humidity time-series data from UCD swimming pool hall. The following section examines the merits of a number of alternative breakout detection algorithms in comparison to the chosen technique in this paper.

State of the Art

Building Management Systems (BMS) often have preprogrammed alarm notifications for important criteria such as room temperature, room humidity etc. If room temperature for example goes outside a specified pre - programmed limit an alarm notification is triggered in the BMS for the building manager to check. The building manager is eventually overrun by alarms, ensuring that many building problems continue indefinitely. A building manager's focus is not optimal operation but simply keeping the building operational (O'Donnell et al., 2013).

Analytics of time-series building data is emerging, with a variety of techniques and potential application cases. Machine learning models has been used extensively to predict building energy consumption (Ahmed et al., 2011; Korolija et al., 2013). By feeding this time series building data to statistical techniques such as linear regression, neural networks, support vector machines etc., predictions of future building performance can be generated. These statistical approaches do not explicitly model the buildings' physical systems, but instead use a number of carefully selected variables such as outside air temperature, historical building performance data, occupancy etc. to develop correlated outputs such as the predicted energy consumption of a building (Li et al., 2009). However, these statistical models need to be trained and tested on extensive quantities of time-series data, in the region of two years worth, to produce a performance prediction. Additionally, this quantity of time-series data is needed for each input (e.g occupancy) that is used in the statistical model.

Machine learning models have also been used in a number of analysis contexts for Fault Detection and Diagnosis (FDD) in buildings. West et al. (2011) developed a FDD approach for HVAC subsystems based on statistical machine learning and information theory. A probabilistic relationship was established between groups of points when the HVAC subsystem was experiencing faults and operating normally. Support Vector Machines (SVM) were used in tandem with a genetic algorithms with parameter tuning for FDD in chillers by (Han et al., 2011). Parameter tuning was used to select the best parameters to optimise SVM performance while additionally simplifying the detection and diagnosis process. The number of parameters used in SVM was reduced from 64 to 8 by parameter tuning, greatly simplifying the detection and diagnosis process. Of particular interest, when reviewing the methods described above is the combined use of a number statistically based techniques to detect and diagnose performance problems in buildings.

A simpler approach, one that works with a single time-series and does not require a combination of techniques to accurately identify building performance problems is proposed. Breakout detection, a technique that works with a univariate or single time-series and does not require the same quantity of data to run effectively. A breakout characterised by either a mean shift or a ramp up from one steady state to another in a given time-series (James et al., 2014). Since breakout detection is run using a single time-series the diagnosis of the performance problem is much easier, with methods such as parameter tuning not necessary. Given the large number of building services that are now producing data, visual detection of breakouts is not a feasible solution. An algorithmic approach one that can detect breakouts in building performance data in real-time via statistical learning will be used in this paper.

This breakout detection technique, known as the E-Divisive with Medians (EDM), developed by (Twitter, 2014), uses a modified variant of energy statistics that is more resilient to anomalies through the use of robust statistics such as the median. The concept of energy statistics is to compare the distances of means of two random variables contained within a larger time-series. A hypothesis test is used to determine if this difference is statistically significant. However the presence of outliers or anomalies would limit the effectiveness of using the mean in this process as a single outlier can have a significant effect on the mean of a time-series. To that end, EDM builds the technique on a robust statistic, the median. The EDM algorithm is robust to anomalies and able to detect multiple breakouts in a time-series.

Robust statistics perform strongly for data taken from a wide range of probability distributions, especially for non-normal distributions. The use of robust statistics for breakout detection was first proposed by (Vallis et al., 2014). EDM is a non - parametric approach, meaning is does not make any prior assumptions of the form of the data distribution.

Breakouts can appear in data that do not conform to any known regular distribution, thus rendering techniques that presume a specific distribution less effective. Pruned Extract Linear Time (PELT) is a breakout detection technique that makes a prior assumption that the data is normally distributed. A number of quantitative metrics have been used to compare the performance of EDM and PELT. These metrics include Time

To Detect, meaning the exact point of the breakout to when the algorithms identifies it and "Precision" is the ratio of true positives over the sum of true positives and false positives. The EDM algorithm outperformed the PELT algorithm in the presence of anomalies, if anomalies were smoothed or if anomalies were removed in the majority of data sets (James et al., 2014). However, the PELT algorithm executes faster than the EDM algorithm due to PELT's stronger assumption that the data is normally distributed. Other approaches to identify breakouts in time-series data such as simple moving average have been shown to be sensitive to outliers and noise in the data for small window sizes (Mathworks (accessed Novermber 2016), 2016).
To summarise, the EDM algorithm is a mechanism to determine if a new segment of time-series data is considerably different from the previous through the use of distance statistics robust to anomalies. It has shown to outperform existing approaches that make prior assumptions about the underlying time-series data and can detect multiple breakouts per time-series.

Context Based Analysis

Breakout detection analysis can be undertaken at a range of resolutions (i.e. building, system, component level etc.) if time-series data exists, for that particular analysis context or building object. Considering the number of potentially available data points produced by a building, a structure is needed to organise these data points. This structure can be viewed as a formal representation of different building objects.
The full scope of building objects for an individual building have been previously defined by (Maile et al., 2010), it includes building, floors, spaces, zones, HVAC systems and HVAC system components align with the hierarchy Industry Foundation Classes (IFC) (ISO16739). Thus, enabling, a building object hierarchy to be established from both a spatial perspective and a HVAC perspective. A HVAC building object hierarchy for an example building is illustrated in Figure 1.

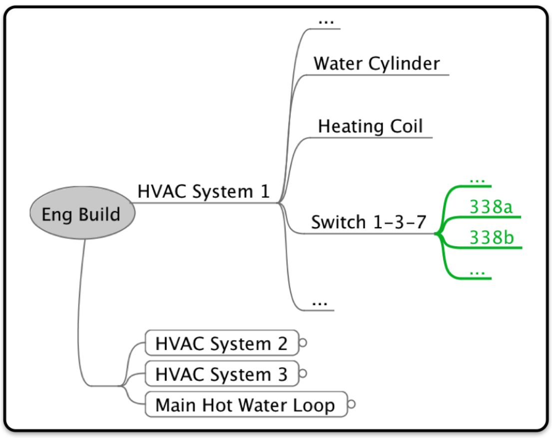

Figure 1: An example building object hierarchy (Maile et al., 2010).

A number of relationships are contained within the HVAC object hierarchy, the building - system relationship is a PartOf relationship, as is the system - component. The only exception is the component - zone object relationship which is a Is- ServerdBy relationship, this can be clearly seen in Figure 1 with the components (water cylinder, heating coil, switch) providing energy to the zone, in this case to rooms 338 within the UCD engineering building.
If time-series data exists related to the energy consumption of HVAC System 1 in Figure 1 for example, the breakout detection can be implemented and identify a ramp up in HVAC energy consumption. The same logic can be applied to any object within the building object hierarchy.

Methodology

The methodology presented in this paper combines: 1) measured time-series data used in a scalable breakout detection algorithm to identify unintended shifts in building performance and 2) short term forecasts of building performance, represented by calibrated simulation output data, to reset and guide operational performance (Figure 3).
The key inputs for the methodology are time-series data relevant to the analysis context. The key outputs from the methodology are identification of unintended shifts in building performance resulting in an enhancement in building performance.
Breakouts can occur in data that conform to any known distribution. Thus techniques that assume data conforms to a certain distribution are less effective. EDM is a non-parametric approach that makes fewer assumptions about the nature of underlying distribution. This is of paramount importance as many data do not conform to the commonly assumed normal distribution, as demonstrated by Figure 2. A density plot, is a non-parametric technique for visualising the underlying distribution of a continuous variable.

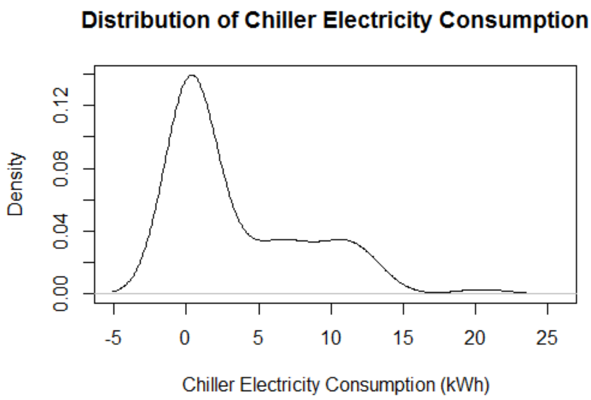

Figure 2: Example data distribution of air-cooled chiller electricity consumption, demonstrating that time-series data is not always normally distributed about the mean of the data.

The methodology comprises three key stages:

1. Assessment of the existence of seasonality in the

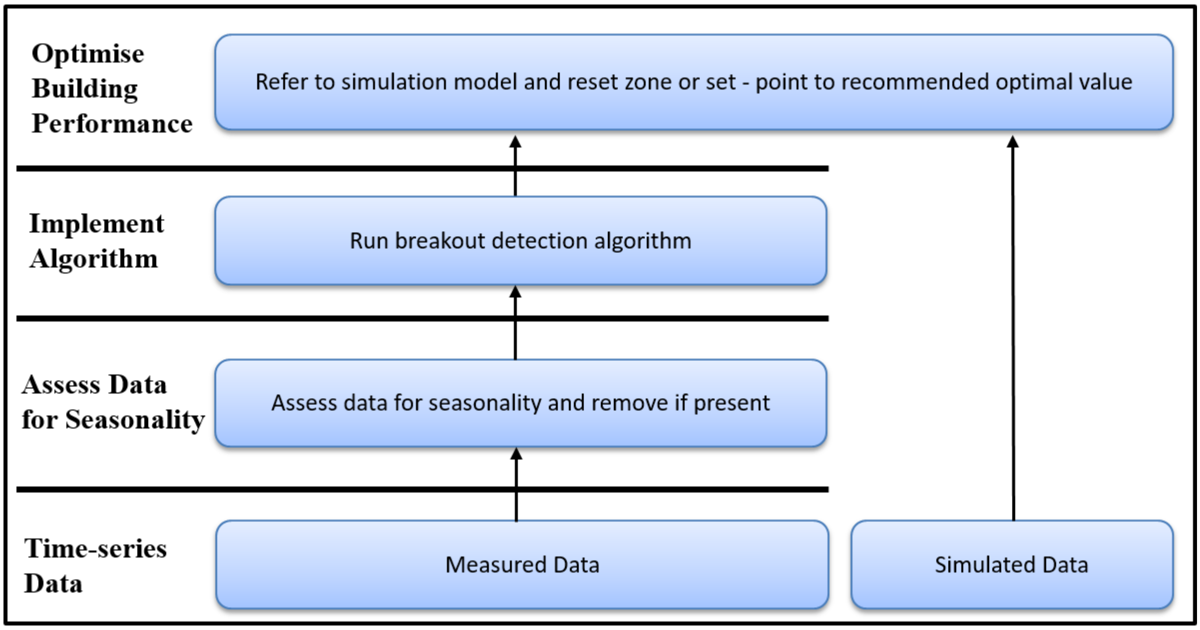

Figure 3: Methodology that combines measured time-series data used in a scalable breakout detection algorithm to identify unintended shifts in building performance and calibrated simulation model output data, to reset and guide operational performance.

data

2. Implementation of EDM breakout algorithm on the data
3. Optimisation of building performance

Assessment of the existence of seasonality in the data

The EDM algorithm has shown to be more susceptible to false positives when seasonality exists in the time-series (Magkian, 2016). A false positive can be thought as a false alarm. That is, identification of breakout by the EDM algorithm when in fact no breakout existed. Therefore the data has to be tested for seasonality and if present, it is removed. Time-series decomposition is mathematical procedure that decomposes a time-series into a number of different time-series. Seasonal Trend Decomposition using Loess (STL) algorithm was used in this paper to divide the original time-series into three components namely: seasonality, trend and remainder. The STL algorithm is available within R programming language via the stl function (Hafen, 2016)

1. Seasonal: patterns that repeat with a fixed period of time.
2. Trend: the underlying trend of the metric.
3. Remainder: Is the residuals of the time series after allocation into the seasonal and trends time-series.

Implementation of EDM breakout algorithm on data

The EDM algorithm is now run on context specific time-series to identify if building performance has shifted from one performance state to another or experienced a ramp up. A hypothesis test is used to determine if this difference is statistically significant. Specifically, within the EDM technique used in this paper, the hypothesis test seeks to determine if a significant changepoint exists or not in the time-series.

The null hypothesis (H_0) is that no changepoint exists in the time-series. The alternative hypothesis (H_1) is that one significant changepoint exists in the time-series. The calculation of a precise critical value to test the hypothesis requires knowledge of the underlying distribution, which is generally unknown. Therefore, a permutation test is used to determine if the distance in means of the two random variables is statistically significant. Data from the two time-series is permuted a finite or limited number of times to ensure the process of comparing permuted time-series computationally manageable. EDM tests at a significance level of 5% and completes 199 random permutations. It should be should be noted that EDM does not provide significance values for each breakout detected.

Optimisation of building performance

If a shift in the context specific time-series is deemed statistically significant by the hypothesis test contained within the breakout algorithm, a breakout will be detected. The Null hypothesis may be that a breakout does not exist in the time-series. The Alternative hypothesis would be a breakout does exist in the time-series. If the breakout detection algorithm deems the difference in the time-series to be statistically significant, the null hypothesis will be rejected and a breakout will be identified. This breakout algorithm can identify either a mean shift or a ramp up from one steady state to another in a given time-series.

Once a breakout is identified, the output data from the

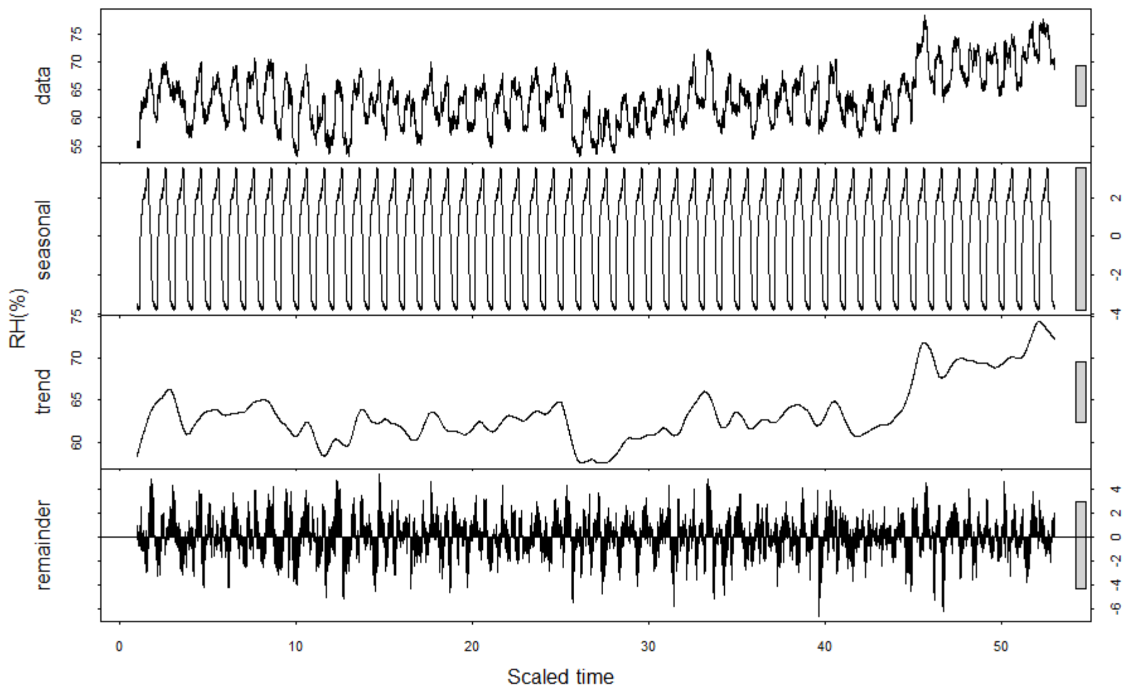

Figure 4: Output from STL algorithm implemented on ACC electricity consumption data. The STL algorithm divided the original time-series into three components. With a seasonality of one day clearly seen in the ACC electricity time-series.

calibrated simulation model is referred to. This enables identification of the optimal value, for the specific analysis context, at that particular moment in time. The building system or zone set-point is reset to the recommended optimal value, represented by calibrated simulation model output data. The building manager is not expected to update the simulation as a typical building energy manger would not possess energy modelling skills. The simulation model is instead maintained by a building energy modeller. Malfunctioning building objects such as boilers and chillers can also be identified by the breakout detection algorithm if time-series data exists for such building objects.

The time-series building data that is fed into the breakout detection algorithm and simulated model output data must have the same resolution (e.g. 15 minute, 30 minute etc.). This ensures that the measured data resolution and the simulated model output data align, when the simulated model is referred to.

Dependent on the analysis context in question, this will result in an improvement in indoor environmental conditions or a reduction in building energy consumption. The energy performance gap between measured and simulated energy consumption will narrow, as measured building performance is aligned with simulated building performance. The EDM can be run in real time as new data becomes available to identify changes in building performance.

Case Study

The case study analysed time-series data from the sports centre within UCD swimming pool hall in 2015. The pool air is typically kept at 30°C and relative humidity should kept in the region of 50%-70%. With this air temperature and the warm temperature of the water (29-30°C), there will be a lot of evaporation from the pool. Humidity can rise quickly and if humidity levels in the swimming pool rise above 70% chemicals from the pool water become airborne.

The decomposition of the seasonal trend with the return humidity data is clearly demonstrated in the seasonal time-series of Figure 4. The removal of seasonality from the time-series will mitigate the effect of external air moisture content and solar gain on pool humidity levels. The underlying trend in the data is also clearly seen in Figure 4, with an increase in RH trend coinciding with the detection of a breakout in Figure 5.

Simulation model development

The whole building energy simulation model was developed in EnergyPlus, while the HVAC and control system of UCD swimming pool were developed in Modelica. This was facilitated by the Building Controls Virtual Test Bed (BCVTB).

The BCVTB is a software environment that allows users to couple different simulation programs for co-simulation, and to couple simulation programs with actual hardware. For example, the BCVTB enables a building to be simulated in EnergyPlus and the HVAC

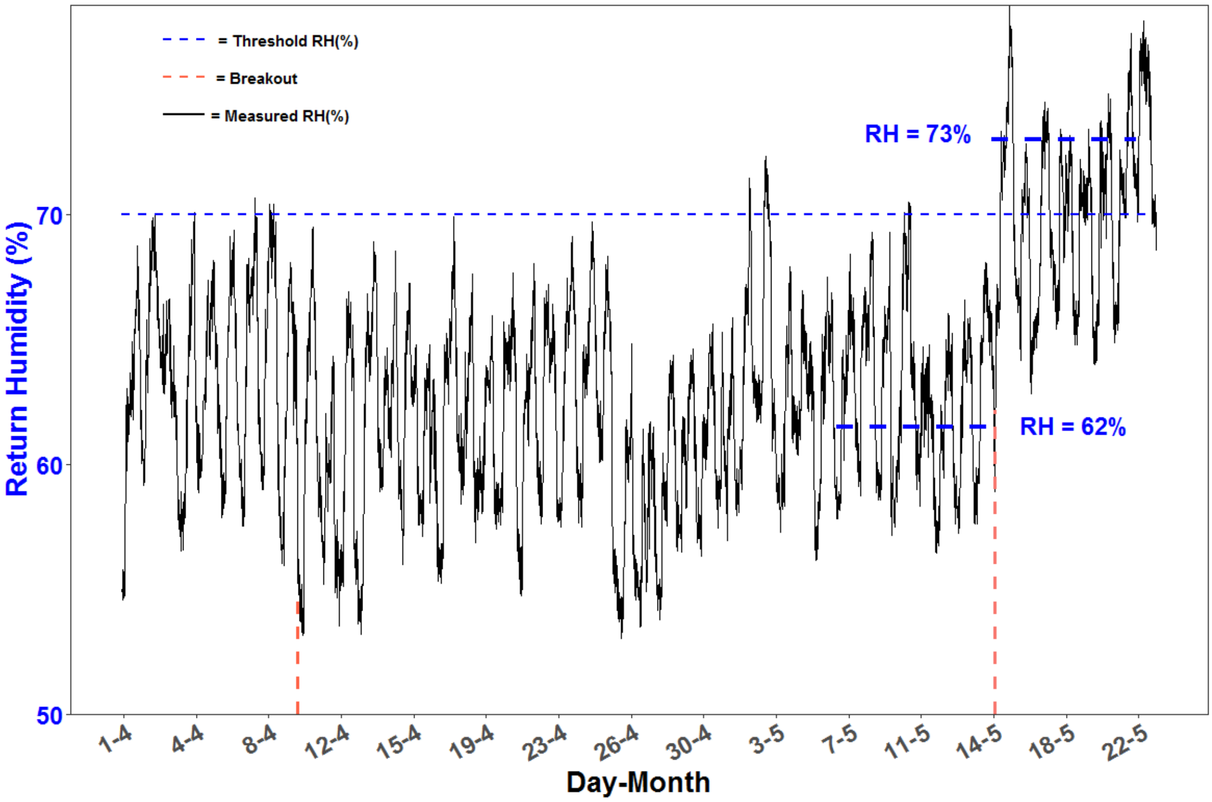

Figure 5: Breakout detection algorithm illustrating the detection of a steady state shift in RH above an acceptable value of 70%.

and control systems in Modelica, while exchanging data between the software as they simulate.

The International Performance Measurement and Verification Protocol (IMVP) provided validation criteria for simulation models to be calibrated to hourly building data with a 5% difference between measured and simulated data deemed acceptable (IPMVP, 2002). Other validation criteria exist based on if the simulation model is calibration to monthly measured building data (ASHRAE, 2002; Nexant, 2008). Given the real-time nature of this analysis, the acceptable difference between measured and simulated is 5% based on the validation criteria provided by IMVP.

Ideally real-time simulation modelling would be used as the reference point for when a breakout is detected and the building object has to be reset to the recommended optimal value. In practice generally, and specifically within this paper an existing model will was used with historical building data.

Results

Time-series data at 15 minute resolution were analysed from the sports centre within UCD swimming pool hall. Traditionally, an alarm notification would have been set off in the building management system, notifying the Building Manager (BM) if RH levels went above 70%. This is the case, as illustrated by the RH time-series venturing above 70% RH (represented by the dashed blue line) multiple times in Figure 5. Numerous alarms would have been sent to the BM from the building management system. The BM may simply ignore these alarms due their velocity or she/he would have spent considerable time and resources checking the environmental conditions within the pool hall and HVAC system settings to ensure everything was satisfactory.

Maintaining relative humidity below 70% reduces condensation on the structure and finish materials, which can lead to rapid deterioration of structural elements and finish materials. Additionally, high humidity levels provide uncomfortable conditions for occupants.

The breakout detection algorithm was run on the pool hall RH time-series data with two separate breakouts detected (represented by the dashed red line) by the algorithm in Figure 5. It should be should be noted that EDM does not provide significance values for each breakout detected. EDM automatically detects based on a significance value of 5%, as discussed in the above section. A basic visualisation is provided by the EDM algorithm of the time-series with the location of the detected breakouts. The computation time to run the EDM algorithm on the RH time-series of approximately five thousand observations was just sixteen seconds on a Dell laptop with a intel core i7 processor and sixteen gigabytes of RAM.

The first breakout was detected on about 10th of April, alerting the BM to a potential change in operating con-

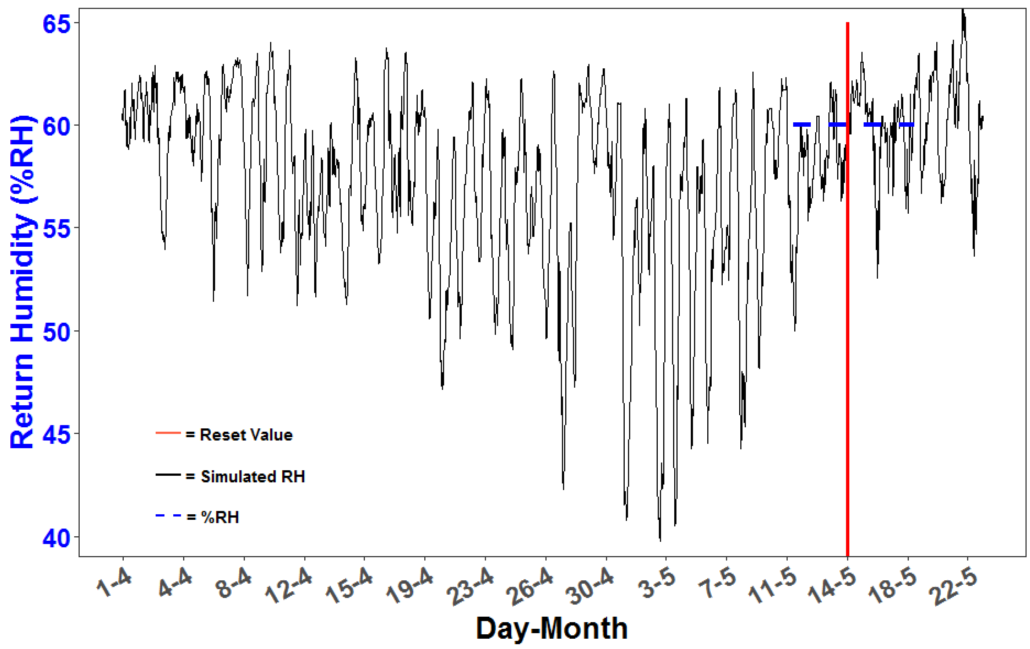

Figure 6: The calibrated simulation output data is referred to, enabling identification of the optimal value of pool hall RH at that particular moment in time. In this instance it is approximately 60% RH.

ditions. However, the detected breakout may not be relevant from an indoor environment perspective to the BM, meaning that RH levels within the pool hall may not have gone above 70%. This is the case, as pool hall RH is approximately 55%. The BM should refer to the output data from the calibrated simulation model (Figure 6) to cross reference what the optimal value should be for the pool hall RH at that particular moment in time. The output data from the calibrated simulation model ensures the BM does not waste valuable time and resources rechecking the settings of the HVAC system that maintain the environmental conditions within the pool hall.

The second breakout detected by the algorithm was detected on the 14th of May. In this instance, pool hall RH levels ramp up considerably and RH levels appear to be operating on a new steady above 70%. The BM should again refer to the output data from the calibrated simulation model. A value of 60% RH would be obtained from the simulation model, illustrated by the solid red line in Figure 6. The BM would then reset the operating of HVAC system that maintain the environmental conditions within the pool hall until optimal operating conditions are reached. In this instance 60% RH. The difference between measured and simulated RH is 17%. This is quite a big difference and substantially larger than the validation criteria percentage set out by IMVP of 5%. However, the operating conditions within the swimming pool hall are outside specified acceptable limits as detected by the breakout detection algorithm. The calibrated simulation model output value for RH is within specified limits of between 50% and 70%.

The RH time-series ventured above 70% RH (represented by the dashed blue line) a multitude of times in the Figure 6. Given that the RH data is at a resolution of 15 minutes over a hundred alarms would have been triggered in the BMS for the building manager to review, which would take considerable time. Due to the volume triggered in the BMS the building manager will more than likely deem the alarms redundant and move onto the latest problem or fire that needs to be dealt with in the building. When an unintended shift in building performance was detected (e.g. RH) by the breakout detection algorithm, actual time-series data from UCD pool hall were compared with output data from the simulated model. The combined use of the breakout detection algorithm and calibrated simulation output data ensures the BM is not overrun by alarms, enabling he/she to do more work with available resources.

Conclusion and Future Work

This paper proposes the idea that data analysis and simulation can be be used in tandem to optimise building performance and reduce the environmental and energy performance gap in buildings.

This work presents a novel combined approach to address the performance gap. A scalable methodology was implemented that identifies shifts in building performance, with calibrated simulation used to reset and guide operational performance. The engineering value of this process is that it can be run in real time, as time-series data is produced, to identify shifts in performance and reduce the performance gap.

The combined approach can be viewed as a tool to improve building performance and reduce the environmental and energy performance gap, in the face of increased regulation and specifications regarding building performance.

The methodology can be applied generally to any building context, as long as time-series data is produced and a calibrated simulation model system exists related to the context in question. Specifically, this context based analysis approach generates efficiencies within the building performance sector, by enabling building managers to identify the root cause of the problem within the building and rectify it.

This work automates manual visual detection of performance problems through trend analysis that is currently undertaken by building managers. Thus enabling building managers to do more work with the resources available to them. This approach is scalable across the building stock if time-series data exists.

Future work in this research process will look into a context based analysis of building performance. This analysis will take place on the building, system, component and zone level of the building object hierarchy. The delivered outcome from this process will be the production of statistically based prediction models to represent building performance.

Acknowledgements

This work is supported by a Marie Curie FP7 Integration Grant within the 7th European Union Framework Programme project title SuPerB, project number 631617.

REFERENCES

Aberer, K. and Tan, K.-l. 2009. Towards Integrated and Efficient Scientific Sensor Data Processing: A Database Approach. *EDBT '09 Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, pages 922–933.

Ahmed, A., Korres, N. E., Ploennigs, J., Elhadi, H., and Menzel, K. 2011. Mining building performance data for energy-efficient operation. *Advanced Engineering Informatics*, 25(2):341–354.

ASHRAE 2002. ASHRAE.Guideline14-2002: Measurement and Energy Demand Savings. Technical report, American Society ofHeating, Refrigeratingand Air- Conditioning Engineers.

ASHRAE 2005. ASHRAE Guideline 0-2005, The Commissioning Process. Technical report, ASHRAE.

Coakley, D., Aird, G., Klebow, B., Earle, S., and Conaghan, C. 2016. Development of Calibrated Operational Models of Existing Buildings for Real-Time Decision Support and Performance Optimisation.

Donnelly, M. 2012. Building Energy Management : Using Data as a Tool Issue Brief. Technical Report October, Johnson Controls.

European Union 2010. Directive 2010/31/EU of the European parliment and the council on the energy performance of buildings. Technical report, European Union.

Granderson, J., Lin, G., Ann, M., and Lawrence Berkeley National Laboratory 2013. Energy Information Systems (EIS): Technology costs, benefits, and best practice uses. Technical Report November, Lawrence Berkeley National Laboratory.

Hafen, R. 2016. Enhanced Seasonal Decomposition of Time Series by Loess. Technical report.

Han, H., Gu, B., Wang, T., and Li, Z. R. 2011. Important sensors for chiller fault detection and diagnosis (FDD) from the perspective of feature selection and machine learning. *International Journal of Refrigeration*, 34(2):586–599.

IDC 2014. The Digital Universe Is Huge And Growing Exponentially. Technical report, International Data Corporation.

IPMVP 2002. International Performance Measurement & Verification Protocol. Technical Report DOE/GO-102002-1554, United States Department of Energy.

James, N. A., Kejariwal, A., and Matteson, D. S. 2014. Leveraging Cloud Data to Mitigate User Experience from "Breaking Bad" arXiv preprint arXiv:1411.7955.

Khan, A. and Hornbæk, K. 2011. Big Data from the Built Environment.

Korolija, I., Zhang, Y., Marjanovic-Halburd, L., and Hanby, V. I. 2013. Regression models for predicting UK office building energy consumption from heating and cooling demands. *Energy and Buildings*, 59:214–227.

Li, Q., Meng, Q., Cai, J., Yoshino, H., and Mochida, A. 2009. Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, 86(10):2249–2256.

Magkian, M. 2016. Change Point detection with seasonal time series.

Maile, T., Fischer, M., and Bazjanac, V. 2010. A Method to Compare Measured and Simulated Data to Assess Building Energy Performance.

Mathworks (accessed Novermber 2016) 2016. Signal Smoothing.

Memoori 2014. Big Data for Smart Buildings: Market Research 2015 to 2020. Technical report, Memoori.

Miller, C. and Schlueter, A. 2015. Forensically Discovering Simulation Feedback Knowledge from a Campus Energy Information System.

Nexant, I. 2008. M & V Guidelines: Measurement and Verification for Federal Energy Projects. Version 3.0. Technical report, United States Department of Energy.

O'Donnell, J. 2009. *Specification of Optimum Holistic Building Environmental and Energy Performance Information to Support Informed Decision Making*. PhD thesis, University College Cork.

O'Donnell, J., Corry, E., Hasan, S., Keane, M., and Curry, E. 2013. Building performance optimization using cross-domain scenario modeling, linked data, and complex event processing. *Building and Environment*, 62:102–111.

RIBA and ARUP 2013. Designing with data : Shaping our future cities - Royal Institute of British Architects. Technical report, Royal Institute of British Architects & ARUP.

Scofield, J. 2012. The Science Behind Green Building Rating Systems, Written testimony to the U.S. House of Representatives. Technical report, Oberlin College.

Twitter 2014. github.com/twitter/BreakoutDetection.

United Nations Energy Programme 2011. What offers governments the greatest potential for reducing GHG emissions at the least cost while creating jobs and improving energy security? Technical report, United Nations.

Vallis, O., Hochenbaum, J., and Kejariwal, A. 2014. A Novel Technique for Long-Term Anomaly Detection in the Cloud. Technical report, Twitter Inc.

West, S., Guo, Y., Wang, X., and Wall, J. 2011. Automated fault detection and diagnosis of HVAC subsystems using statistical machine learning. In *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association*, pages 2659 – 2665.