Machine level energy data analysis – development and validation of a machine learning based tool

Samuel Carvalho
Research Engineer, Department of Electrical and Electronic Engineering
Limerick Institute of Technology, Moylish Campus
Limerick City
Ireland
Samuel.Carvalho@lit.ie

Julio Rezende
Research Engineer, Department of Electrical and Electronic Engineering
Limerick Institute of Technology, Moylish Campus
Limerick City
Ireland
Julio.Rezende@lit.ie

John Cosgrove
Section Head, Department of Electrical and Electronic Engineering
Limerick Institute of Technology, Moylish Campus
Limerick City
Ireland
John.Cosgrove@lit.ie

Frank Doyle
Assistant Lecturer, Department of Electrical and Electronic Engineering
Limerick Institute of Technology, Moylish Campus
Limerick City
Ireland
Frank.Doyle@lit.ie

Keywords
industrial energy saving, data monitoring, machine learning, energy profiles

Abstract
Industrial energy consumption is known to be significantly high worldwide, reaching up to one half of the total energy in some countries and almost one third of the world's consumed energy. Consequently, the industrial sector is also responsible for a large share of greenhouse gases (GHG) emissions. Many recent manufacturing standards and methodologies have efficiency, environmental and social impacts as key aspects and concerns, in order to meet the demands of international agreements and regulations on these subjects. The ongoing development brought by Industry 4.0 and Smart Factories are the main demonstrations of the growing awareness of the importance of efficient and intelligent production. Data gathering, processing, analysis and benchmarking play a key role in this scenario, enabling a smarter and well informed decision-making process. In this context, energy consumption data extracted directly at machine level inside factories also show a significant potential for being a reliable source of process-related information, such as automatic production counting, consumption analysis, Overall Equipment Effectiveness (OEE) and costs identification. This paper presents a pilot implementation of a machine learning based application for the automatic extraction of useful insights from machine level energy data sets. The developed tool uses a K-means clustering algorithm in order to categorise energy profiles into production or idle periods, from which relevant Key Performance Indicators (KPIs) are calculated. A controlled experiment using Non-Invasive Load Monitoring (NILM) equipment connected to a CNC milling machine was performed in order to validate the approach, and the results are presented in this paper. The method has identified potential savings of up to 30 % in an Injection Moulding machine and up to 25 % in a Precision Engineering company, besides providing deeper understanding of the consumption profiles of machines in these sectors.

Introduction
The use of electricity within manufacturing operations was a key enabler for the second industrial revolution, as new equipment and tools started being powered by a better alternative to steam. Therefore, new production approaches such as assembly lines and mass production were made possible, and the utilisation of electricity in the industrial sector has escalated ever since. According to recent reports from the U.S. Energy Information Administration (EIA 2017) the industrial sector is responsible for almost 50 % of the overall energy consumption in the world, and initiatives from governments and environmental agencies have been setting improvement targets related to the emissions of CO_2 and greenhouse gases.

In order to meet the exigencies from stricter regulations the industrial sector has developed strategies and methodologies of continuous improvement, aiming to reduce energy and material utilisation while keeping productivity and quality. Methodologies such as Six Sigma, Lean Manufacturing and Kaizen, which take inspiration from the Japanese car manufacturing industry, are some examples of successfully implemented improvement-driven company cultures (Alhuraish, Robledo, and Kobi 2016).

With the upcoming trends from Industry 4.0 new ways of dealing with data from the shop floor are being made possible through Key Enabling Technologies (KETs) such as Big Data, Cyber Physical Systems, Internet of Things, Cloud Computing and Machine Learning, for example. Therefore, data related to the energy consumption at machine level can be extracted and analysed in order to provide useful production-related information that can help the decision-making process at different business levels and validate human-input data that normally goes into Enterprise Resource Planning (ERP) systems.

This paper presents the development and validation procedures of an energy data analysis tool. The developed web application provides an "installation wizard style" interface to a series of analysis steps, in which machine learning algorithms such as Principal Component Analysis (PCA) and K-Means clustering are used in order to automatically identify production and idle periods from the analysed energy profiles. Information such as the number of produced parts, the energy costs related to each part and the running and idle costs are then calculated from the obtained classification between production and idle. This paper also presents the results from a controlled lab experiment and two case studies in which the developed application was used in the analysis of data from precision engineering and injection moulding machines.

Methodology

DATA ACQUISITION PROCEDURES

The growing number of applications within manufacturing companies that rely on process-related data has boosted the development of new technologies, equipment and methodologies for the acquisition, storage and processing of data. The concepts of Industrial Internet of Things (IIoT) and Cyber Physical Systems (CPS) play a key role in the data-driven factories proposed by Industry 4.0, as these technologies promote the connection and mutual feedback between the computational and the physical world (Jazdi 2014; U.Farooq et al. 2015; Papakostas, O'Connor, and Byrne 2016).

Since the analysis reported in this paper is focused in energy data extracted at the machine level, a strategy for the consistent acquisition and storage of these data was stablished. Non-Invasive Load Monitoring (NILM) equipment were used in order to log three-phase current values at the power supply of the analysed machines, and a "Tidy Data" (Wickham 2014) structure, shown in Figure 1, was adopted for the storage of the values. The chosen structure assures data integrity and was used as a standard input format for the implemented analysis.

The chosen data structure always had the first column filled with timestamps for each observation, while the second to fourth columns contained the RMS current values measured in each phase at a given time. Once a data set was uploaded to the application for analysis the current values were averaged and converted into kW values according to Equation 1.

$$P_{(kW)} = \frac{\sqrt{3} \times PF \times I \times V}{1000} \quad (1)$$

Equation 1. Power, in kW, for a three-phase machine, given the power factor PF, current I and voltage V.

ENERGY PROFILES' ANALYSIS

The typical process cycle for discrete manufacturing machines takes a number of stages, such as material feed, tool changes, heating and cooling, finishing, etc. As different subsystems or auxiliary circuits of the machine are used in each of these stages, a consistent sequence of kW values corresponding to each of the stages can be seen in the measured data as a machine produces parts (Geramifard et al. 2016). Figure 2 depicts the energy profiles obtained from the production of a batch of parts in a precision engineering machine.

The time series energy data obtained from a machine can provide a clear picture on when it was producing parts or not, as production cycles will present significantly higher energy consumption. However, as the task of identifying these two operation modes in a graph may be trivial for a human, the definition of generic and accurate enough mathematical rules to correctly assign each data point to a category can be a challenge. For this reason, a machine learning approach was implemented to automatically identify production and idle periods in the analysed energy data sets.

The solution implemented uses a K-means clustering algorithm for the grouping of similar data points. However, as the clustering of time series is a computational costly operation, a dimensionality reduction using Principal Component Analysis (PCA) is performed on the data before the clustering (Hyvärinen and Oja 2000). The PCA algorithm takes in account the variance presented by the analysed data to establish new coordinate systems in which most of the variance is represented. Thus, a simpler representation of data is obtained in a lower dimension hyperspace, but with the minimum possible loss of

	time	AIRMS	BIRMS	CIRMS
1	2016-09-05 13:42:28	1.812	0.611	0.164
2	2016-09-05 13:42:13	1.816	0.607	0.160
3	2016-09-05 13:41:58	1.828	0.612	0.162
4	2016-09-05 13:41:43	1.808	0.604	0.164
5	2016-09-05 13:41:28	1.815	0.609	0.163
6	2016-09-05 13:41:13	1.818	0.606	0.164

Figure 1. An example of the utilised tidy data structure.

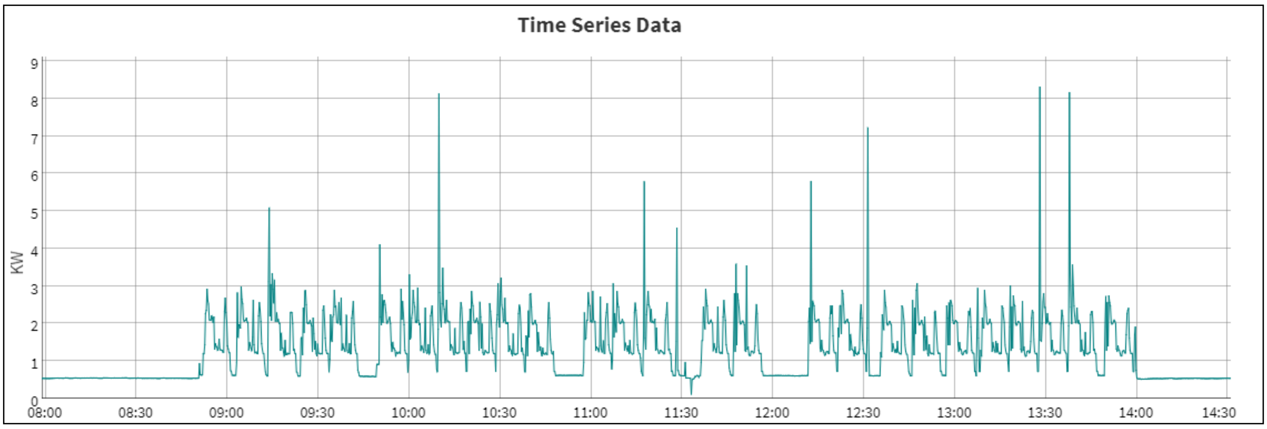

Figure 2. Energy profiles from a batch of produced parts in a precision engineering machine.

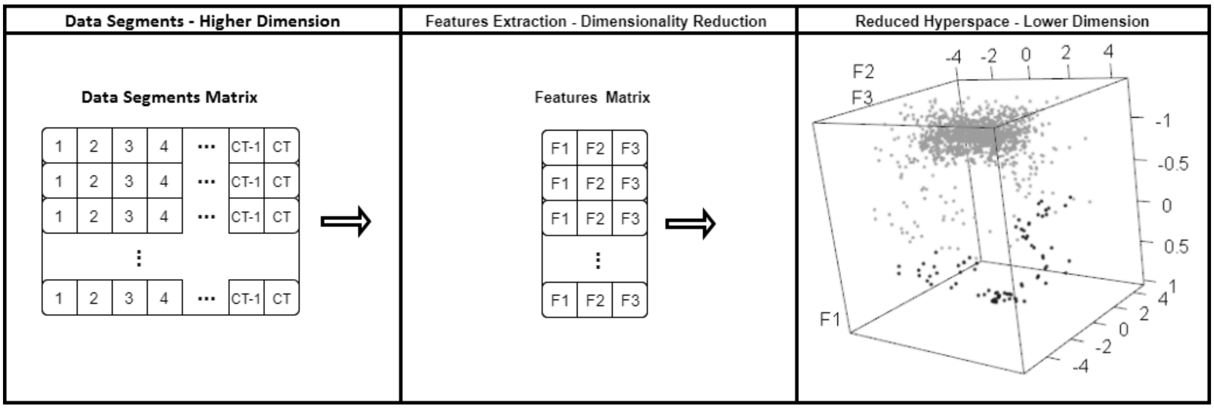

Figure 3. The process of data segmentation and features extraction from the time series data, adapted from Hyvärinen and Oja (2000).

information. In different terms, the PCA algorithm solves an optimisation problem for a set of n data points with p features which yields "a line in p-dimensional space that is closest to the n observations" (James et al. 2017).

Before the PCA features are extracted, a sliding window is used to segment the time series data into temporal data segments with the same length of the product cycle time (CT). Then, a set of features is extracted from each of these data segments. The choice for using one cycle time as the length for the data segments enables that eventually the energy profile of exactly one product is contained by the segments. Figure 3 presents a schematic representation of the dimensionality reduction process into a three-dimensional hyperspace.

Once the features are extracted the K-means algorithm is used in order to group them in two clusters, corresponding to production and idle. The K-means, described below, is an unsupervised machine learning algorithm that uses the Euclidean distance to categorise a set of data points in K different clusters from an initial random guess.

After the feature clustering is performed, the grouping information is mapped back to the temporal data segments, which can be visualised in a scatter plot, for example.

ANALYSIS OUTCOMES

Once the clustering of the energy data is finished, useful outcomes and Key Performance Indicators (KPIs) can be extracted from the resulting analysis if a few parameters are known, such as the products' cycle time (CT), the energy price for each kWh consumed and the sampling time of the data (Ts). These outcomes can provide a better understanding of the production process, its energy consumption and Overall Equipment Effectiveness (OEE) figures. Table 1 presents a summary of the possible outcomes obtained from the analysis.

Box 1. K-means Clustering Algorithm.

1. Each observation from the data set is randomly associated to one of the K clusters.
2. The following operations are iterated until changes in the clusters cease:
   a. Each centroid for the K clusters is calculated. The centroid is the mean value of the features from all observations assigned to a cluster.
   b. All observations are re-assigned to the closest calculated centroid, using the Euclidean distance as measure.

Algorithm 1. K-means Clustering algorithm (James et al., 2017).

Table 1. Possible outcomes from the developed analysis tool.

Outcome/Indicator	Formula	Unit
Total Energy Consumed	$\sum_{t=A}^{B-1}[T_s\times(\frac{kW_{(t)}+kW_{(t+1)}}{2})]$	$[kW]\times[h] = kWh$
Total Energy Cost	$(Total\ kWh)\times(kWh\ Price)$	$[kWh]\times\left[\frac{€}{kWh}\right] = €$
Total Energy Consumed in Production	$Inside\ the\ Production\ cluster$: $\sum_{t=A}^{B-1}[T_s\times(\frac{kW_{(t)}+kW_{(t+1)}}{2})]$	$[kW]\times[h] = kWh$
Total Energy Consumed in Idle	$Inside\ the\ Idle\ cluster$: $\sum_{t=A}^{B-1}[T_s\times(\frac{kW_{(t)}+kW_{(t+1)}}{2})]$	$[kW]\times[h] = kWh$
Total Production Energy Cost	$(\text{Production Energy})\times(kWh\ Price)$	$[kWh]\times\left[\frac{€}{kWh}\right] = €$
Total Idle Energy Cost	$(\text{Idle Energy})\times(kWh\ Price)$	$[kWh]\times\left[\frac{€}{kWh}\right] = €$
Total Number of Products	$\frac{(Total\ Production\ Time)}{(Cycle\ Time)}$	$\left[\frac{h}{h}\right] = Count$
Average Energy Consumed per Product	$\frac{(Production\ Energy)}{(Number\ of\ Products)}$	$\left[\frac{kWh}{Product}\right]$
Average Cost per Product	$\frac{(Production\ Energy\ Cost)}{(Number\ of\ Products)}$	$\left[\frac{€}{Product}\right]$
Average Energy Cost When Producing	$\frac{(Production\ Energy\ Cost)}{(Production\ Time)}$	$\left[\frac{€}{h}\right]$
Average Energy Cost When Idle	$\frac{(Idle\ Energy\ Cost)}{(Idle\ Time)}$	$\left[\frac{€}{h}\right]$
Total Analysed Time	$(Final\ Timestamp)-(Initial\ Timestamp)$	$[h]$
Production Time	$T_s\times(Production\ Cluster\ Size)$	$[h]$
Idle Time	$T_s\times(Idle\ Cluster\ Size)$	$[h]$
Average Time Spent per Product	$\frac{(Production\ Time)}{(Number\ of\ Products)}$	$[s]$

Validation Experiments

SETUP AND PROCEDURES

Before its application in real industrial environments the developed energy data analysis tool was tested and assessed in a controlled lab experiment, in order to validate its outcomes and assure the accuracy of its results. A batch of mock-up metal parts was produced in a Fanuc Robodrill CNC milling machine while a CPS logged its energy consumption. The equipment setup and the drawing of the produced part are shown in Figure 4 (a) and (b), respectively. The experiment was planned to simulate a real shift of work inside a factory, comprising machine setup and warm-up procedures as well as code execution checks, for example. Also, the part was intentionally designed to demand several tool changes and machining processes by mixing round, square and hexagonal shapes with drilled holes and surface finishing. The resulting energy profiles for the performed experiment can be seen in Figure 5.

From the performed operations a data set containing 24 hours of consistent production was generated from the replication of a 1 hour period, in order to assess the performance of the application regarding data granularity and length. The replicated period was known to have 5 parts produced in it, yielding a total of 120 products on the 24 hours, and further details about the performed tests are presented below.

INVESTIGATION OF THE RELATIONSHIP BETWEEN DATA LENGTH AND ACCURACY

In order to investigate how the application performed in relation to the amount of data analysed a series of tests was conducted using the 24 hours replicated data set. As the amount of produced parts, production and idle times were known for the original 1 hour replicated period, segments of increasing duration from 1 to 24 hours were uploaded to the application and analysed. Thus, the expected values for these three outcomes were compared with the outputs from the application through the calculation of the Absolute Percentage Error (APE). The successive APE values for segments of data with increasing duration are given in Figure 6, from which a decreasing behaviour in the error can be seen as the length of the analysed data set grows.

INVESTIGATION OF THE RELATIONSHIP BETWEEN DATA GRANULARITY AND ACCURACY

Another series of tests regarding the performance of the developed application investigated how the sampling time of the analysed data set would interfere in the accuracy of its outcomes. For this purpose, the same 24 hours replicated data set was manipulated and resampled with increasing sampling times, ranging from the original 1 second up to 32 seconds. Therefore, the resulting under sampled data sets would cover the same 24 hours span as the original one, but with decreasing number of samples.

The same three outcomes from the application were again compared with its expected values: products count, production time and idle time. The evolution of the APE for each of the used sampling times is presented in Figure 7.

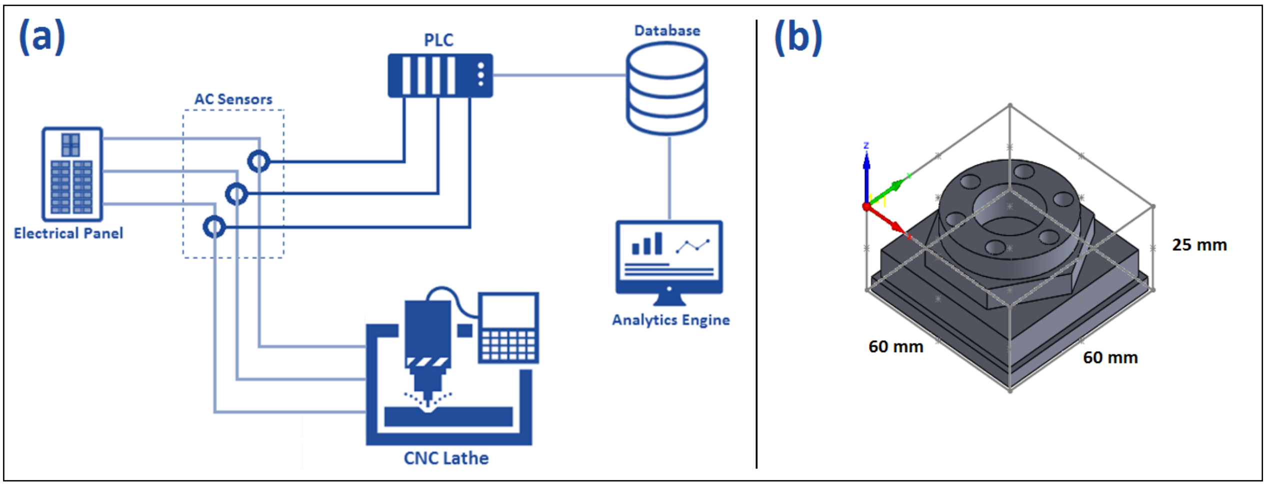

Figure 4. a) The setup utilised in the validation tests, and b) the mock-up part produced in the tests.

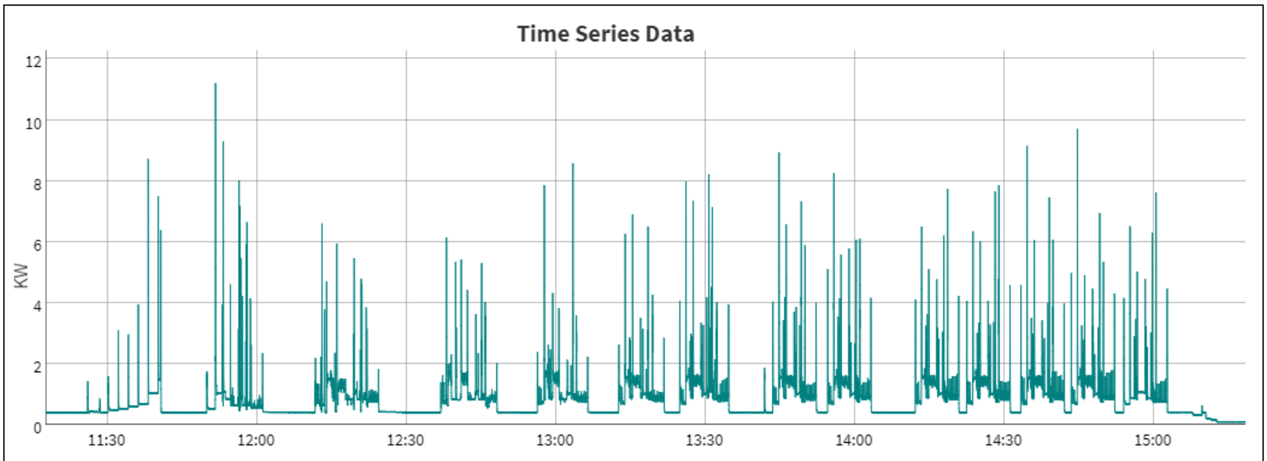

Figure 5. Energy profiles obtained from the validation experiment.

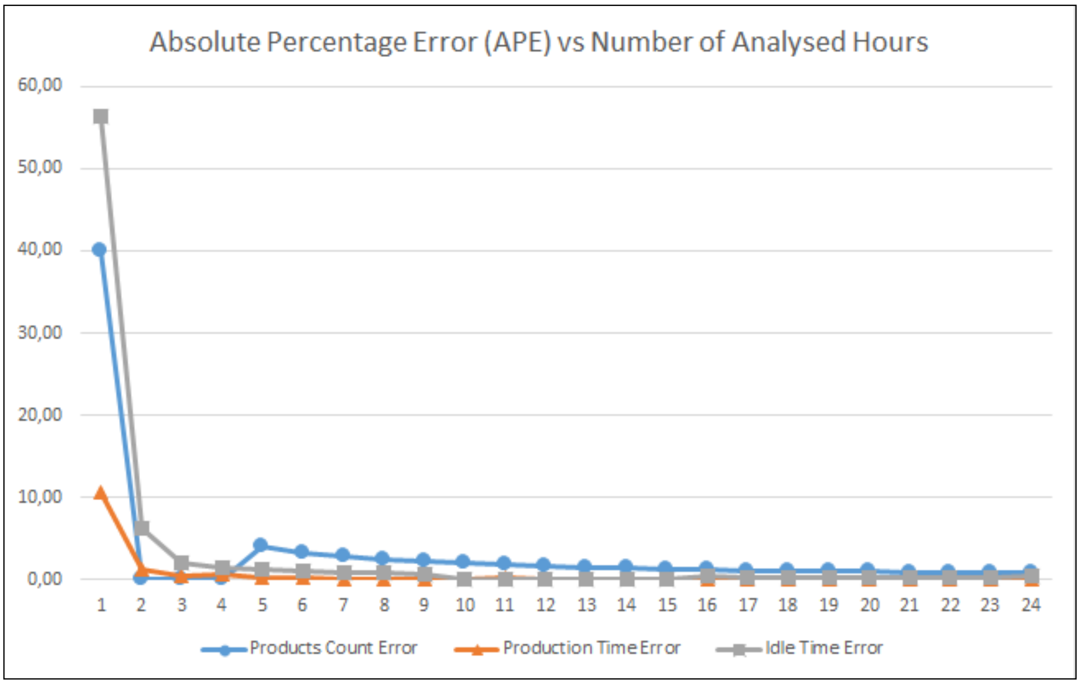

Figure 6. APE between the outcomes and the expected values.

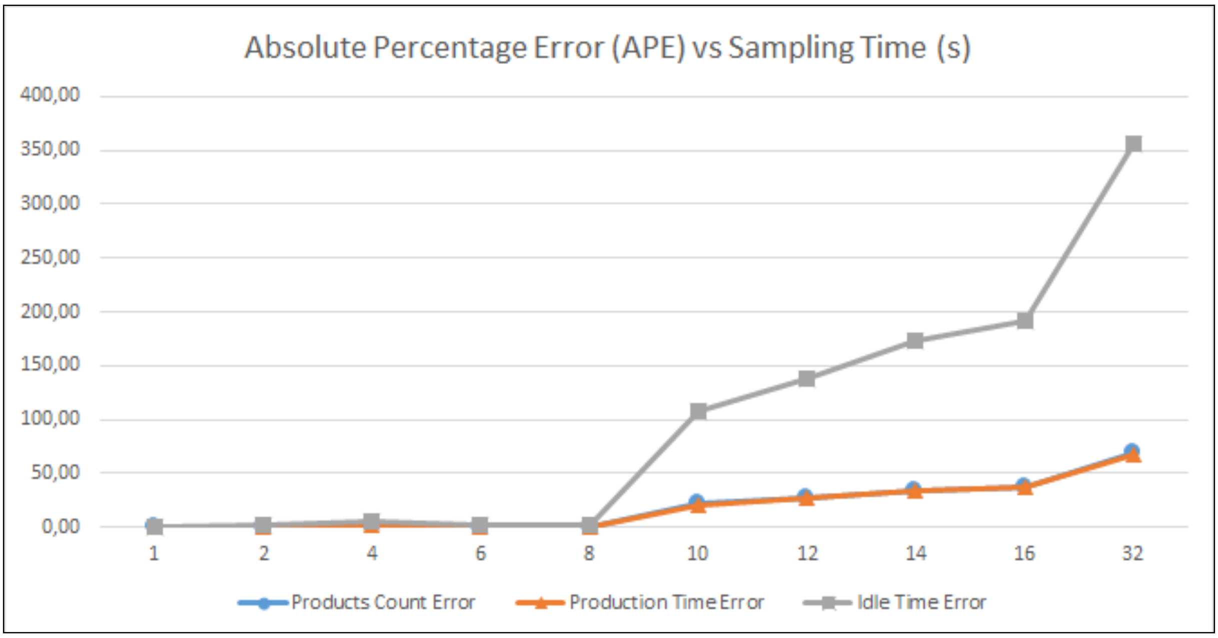

Figure 7. Evolution of the APE for growing sampling times.

Case Studies

CASE STUDY 1: PRECISION ENGINEERING COMPANY

Introduction

"TPE" is a precision engineering SME from the west of Ireland. Employing approximately 60 people, the company currently has around 30 CNC machines in its facility, from different makes and models. Its main customers are manufacturers from the aerospace, automotive and medical devices sectors, to which the company sells components made of steel, aluminium and plastic. The machines used in the production of these parts include lathes, milling and grinding equipment, for example.

With a fast-paced growth in the last few years, the company has identified some bottlenecks in its production, management, planning and logistics that may impede the upscale of the business in an organised and efficient fashion. One of the main reported bottlenecks is the excessive dependency on paper-based processes for the management, scheduling and execution of tasks in its shop floor. Also, the company has identified that a direct and efficient interface between its machine operators and the ERP system does not exist at the moment, and valuable production-related information such as exact products count, scrap quantities and quality issues are not being properly logged and analysed.

The company recognises that a digitisation strategy is needed for the short term, in order to better organise its operations and to implement better ways of extracting, analysing and presenting production-related data. Therefore, the use of energy profiles for obtaining these relevant shop floor data and turning it into useful insights from the processes was investigated through a pilot study where the analysis tool described in this paper was used.

Procedures and Results

In order to gather data for the proposed analysis, a Cyber Physical System was deployed to one of the company's Myiano CNC milling machines using the same setup as the one depicted by Figure 4. Energy consumption at the referred machine was logged for a period of 19 days, in which three different types of products have been made. The collected data set was analysed, and the results in Table 2 were obtained.

A quick analysis of the whole data reveals a poor utilisation percentage for the machine. In fact, the machine was reportedly left in idle state for several days within the logged period, like weekends and bank holidays, for example. This behaviour was responsible for a lower than 40 % percentage of time being used in actual production, while approximately one third of the total energy was wasted in idle periods in the analysed data.

Figure 8 brings pie charts depicting the percentages of time and energy associated to either production or idle periods, from which the poor utilisation strategy for the machine is made explicit. This overall analysis of the machine's utilisation has demonstrated a relevant potential for savings through the implementation of a better scheduling strategy, for example. Using the behaviour changes methodology proposed by Cosgrove et al. (2014), it was estimated that energy savings of up to 25 % would be achievable for the analysed machine. As these changes could be certainly applied to the other machines in the shop floor the company would certainly be able to significantly reduce its overall energy consumption with little or no investment needed.

Also, the analysis tool was tested regarding its capability of automatically identifying the number of products made in a certain period. For this purpose, subsets of data containing en-

Table 2. Outcomes from the analysis of the Myiano CNC machine data.

Indicator	Result
Total Energy Consumed	486.96 kWh
Total Energy Cost	€73.04
Total Energy Consumed in Production	325.89 kWh
Total Energy Consumed in Idle	161.07 kWh
Total Production Energy Cost	€48.88*
Total Idle Energy Cost	€24.16*
Total Number of Products	1,611**
Average Energy Consumed per Product	0.20 kWh**
Average Cost per Product	€0.03 **
Average Energy Cost When Producing	€0.28/h*
Average Energy Cost When Idle	€0.08/h*
Total Analysed Time	462.71 h
Production Time	174.55 h
Idle Time	288.16 h
Average Time Spent per Product	390.06 s**

** Costs estimated using the price of €0.15/kWh.*

*** Values obtained using the cycle time of 390 seconds, and may not be accurate due to the presence of different types of products in the analysed period.*

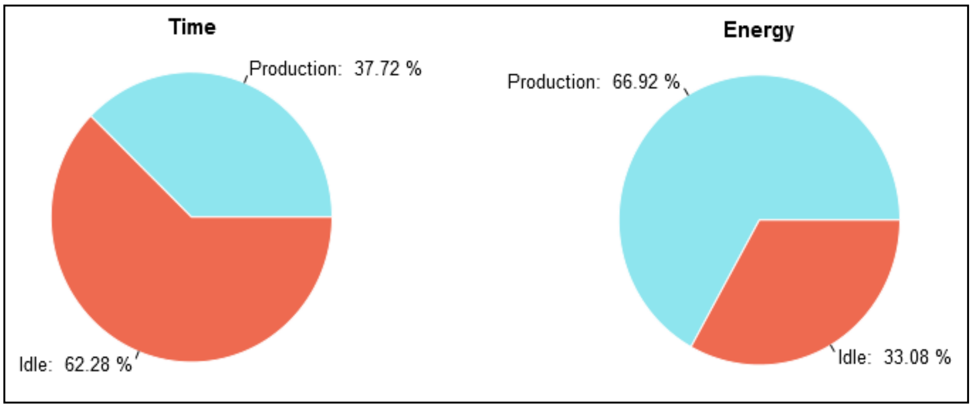

Figure 8. Percentages of time and energy addressed to production and idle periods in the analysed data.

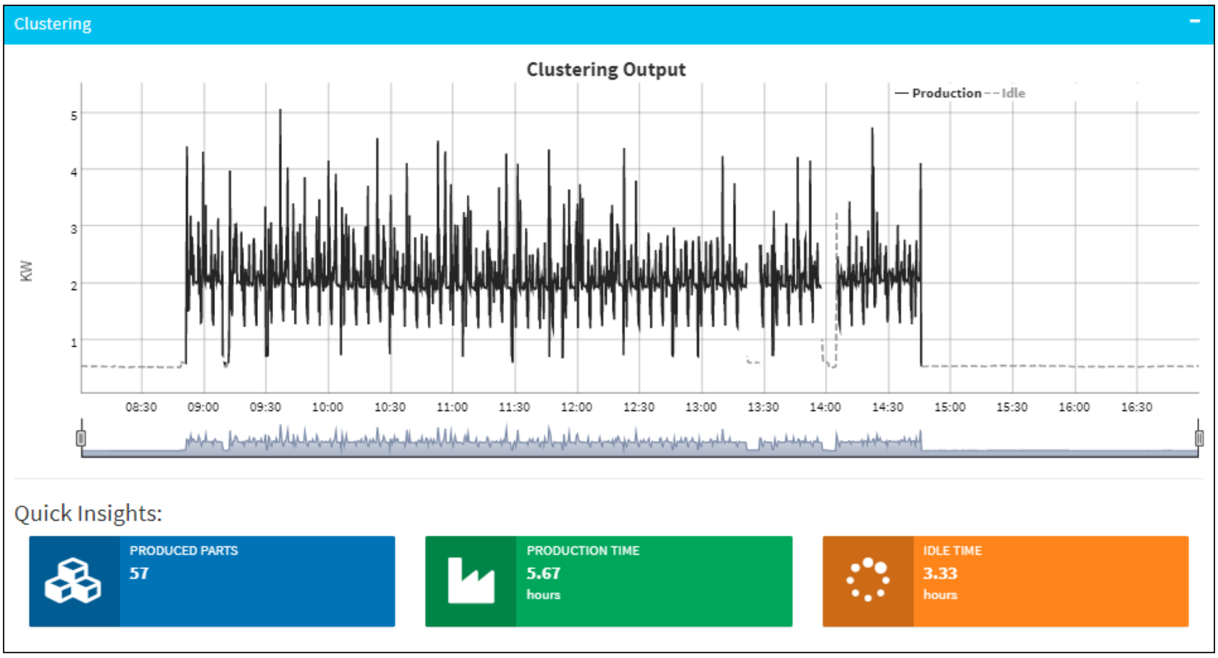

Figure 9. Products counting assessment of one shift from 8:00 am to 5:00 pm, comprising 57 products.

ergy profiles of only a single type of product were individually analysed, such as the one depicted in Figure 9.

The tool was capable of correctly identifying the number of parts in subsets of data containing energy profiles of the three reported different products. Also, the resulting colour coding on the graph outputs, such as the one shown in Figure 9, indicated an accurate estimate of the production and idle times for all the analysed subsets. Therefore, as the application was capable of correctly counting and suggesting to the operators the number of parts made at the end of a batch in an automatic way as well as identifying energy saving opportunities, the implemented analysis was considered a potential solution for some of the digitisation challenges of the company.

CASE STUDY 2: INJECTION MOULDING COMPANY

Introduction

Company "IM" is an Ireland-based SME from the Shannon area. The company currently has 18 different machines ranging from 50 to 2,700 tonnes of weight, and is known for its expertise in the production of plastic parts using processes such as injection moulding, gas assisted injection moulding, silicone moulding and product expansion. As most of these production techniques heavily rely on thermal processes, the injection moulding sector has significant concerns regarding the running costs related to energy consumption in production operations. Also, as some quality and aesthetic characteristics of plastic parts directly relate to the temperature in which they are produced, injection moulding and similar manufacturing techniques typically demand continuous material heating and high pressure clamping, for example.

Given this scenario, the company currently spends nearly €180,000.00 in energy bills every year, or €10,000.00 per machine in average. Therefore, there is a wish from the company to better understand the consumption profile of its production operations and machines, with plans to install real-time monitoring equipment on its machines in the near future. There is a belief that energy consumption information can help the company to better plan its operations, as well as to achieve smarter and ecological friendly operation strategies. Therefore, this case study had the objective of demonstrating the potential insights that the company could extract from energy data and what value this information could add to the company.

Procedures and Results

Data logging equipment was deployed to a Sandretto 110T injection moulding machine, and energy data was collected throughout a 12 hours shift. A similar setup to the one depicted in Figure 4 was used. The machine was producing a symmetrical part during the data logging activities, in which each of its halves would take 25 seconds to be made, resulting in a cycle time of 50 seconds for each complete product. The obtained data set was uploaded and analysed by the application. Table 3 summarises the outcomes from the performed analysis.

As the company expected, the analysed injection moulding machine presented a high energy consumption even when in idle states, consuming more than 15 kW even when not in production, as Figure 10 shows. This consumption profile is highlighted by the small difference seen between the average energy costs when the machine is producing, which was found to be €2.96/h, and the average energy costs when the machine is left in idle state, which was found to be €2.43/h. In percentage, the machine consumes only 9.66 % more energy when producing a part when compared to its idle state, what makes explicit the high costs of keeping the machine on when not in production.

Figure 10 also depicts the obtained categorisation of the analysed time series data into production and idle periods. The application was able to correctly identify the number of parts produced during the shift, which was 450 complete parts taking 50 seconds each. This numbers have matched the information reported by operators to the production manager.

As Figure 10 suggests, the percentages of time and energy utilisation for the machine were found to be significantly poor, with nearly only half of both energy and time being linked to production operations. The pie charts presented in Figure 11 bring the exact percentage numbers for these outcomes.

As the scenario seen in the analysed machine may be the same for the other ones in the shop floor, there is significant room for improvements in the utilisation strategy for the company. An increase of 30 % in both the energy and time addressed to production would certainly bring the company closer to what is considered a "world-class" OEE value (Vorne 2017). These improvements could be achieved by turning machines into a lower consumption state when not in production, such as emergency stop or stand-by. Therefore, besides a better understanding of the energy consumption profile of the analysed machines, the monitoring of the energy usage inside the company could also lead to significant savings by better scheduling behavioural changes. A competitive advantage would also be possible by the reduction of prices to the final customer once the company achieves a more accurate addressing of production costs to each produced part.

Conclusion

Energy efficiency is a growing concern in Industry worldwide, and several studies and strategies have been developed on this field. This paper has introduced a methodology for the acquisition and analysis of energy data from discrete manufacturing machines and its validation procedures, using the K-means unsupervised machine learning algorithm to categorise the col-

Table 3. Outcomes from the analysis of the Sandretto 110T machine data.

Indicator	Result
Total Energy Consumed	205.5 kWh
Total Energy Cost	€30.75
Total Energy Consumed in Production	113.3 kWh
Total Energy Consumed in Idle	91.75 kWh
Total Production Energy Cost	€16.99*
Total Idle Energy Cost	€13.76*
Total Number of Products	450
Average Energy Consumed per Product	0.25 kWh
Average Cost per Product	€0.038*
Average Energy Cost When Producing	€2.69/h*
Average Energy Cost When Idle	€2.43/h*
Total Analysed Time	12 h
Production Time	6:15 h
Idle Time	5:45 h
Average Time Spent per Product	50 s

** Costs estimated using the price of €0.15/kWh.*

Figure 10. Graph output for the Sandretto 110T energy data analysis.

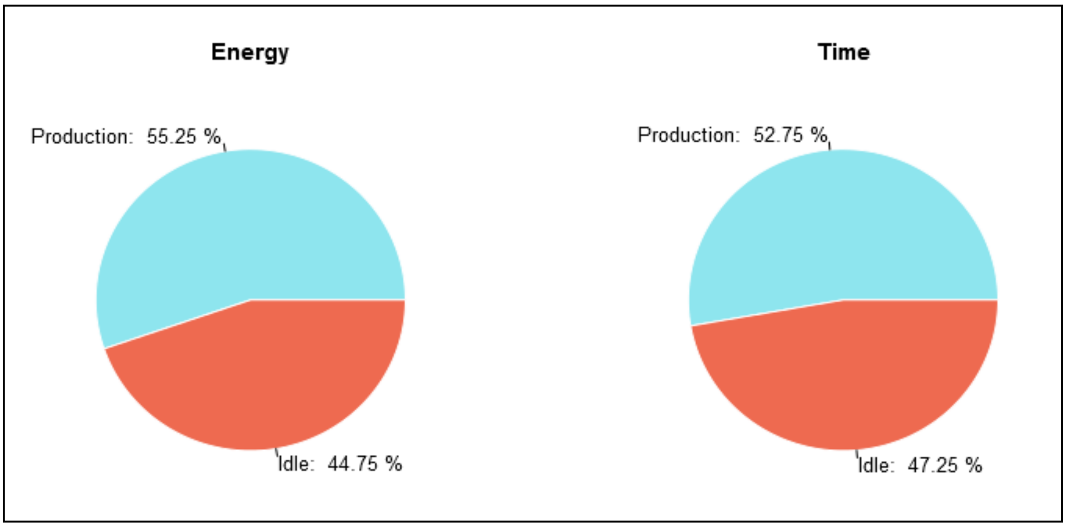

Figure 11. Percentages of time and energy addressed to production and idle periods in the analysed shift.

lected data into production and idle periods. The ability to accurately address the energy spent in these two states can enable a better understanding of machines and processes, allowing for example the implementation of smarter scheduling strategies and the monitoring of Overall Equipment Effectiveness (OEE).

The presented case studies, in which the described methodology was successfully used in the analysis of data from two different companies, have shown potential improvements to the overall operations management and machines' utilisation, achievable through operator engagement, behavioural changes and smarter scheduling. The developed solution was also capable of correctly identifying the number of parts produced in the analysed data sets. Future implementations of the described methodology could be directly linked to Enterprise Resource Planning (ERP) systems in order to provide historical data for targets generation and live monitoring of Key Performance Indicators (KPIs). An ideal solution should also be able to automatically extract all relevant process-related parameters and information exclusively from the energy data sets, in order to empirically validate the production data reported to the production manager, which is known to contain inaccuracies.

References

Alhuraish, Ibrahim, Christian Robledo, and Abdessamad Kobi. 2016. "The Effective of Lean Manufacturing and Six Sigma Implementation." *Proceedings of 2015 International Conference on Industrial Engineering and Systems Management, IEEE IESM 2015*, no. October: 453–60. doi:10.1109/IESM.2015.7380197.

Cosgrove, John, Mike O Neill, Paul Wilgeroth, Frank Doyle, and John Littlewood. 2014. "A Methodology for Verified Energy Savings in Manufacturing Facilities through Changes in Operational Behaviour," 223–32.

EIA. 2017. "International Energy Outlook 2017 Overview." *U.S. Energy Information Administration* IEO2017 (2017): 143. doi:www.eia.gov/forecasts/ieo/pdf/0484(2016).pdf.

Geramifard, Omid, Zhao Yi Zhi, Chua Yong Quan, Hian Leng Chan, and Xiang Li. 2016. "Power-Signature-Based Bayesian Multi-Classifier for Operation Mode Identification." *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA* 2016–Novem. doi:10.1109/ETFA.2016.7733530.

Hyvärinen, Aapo, and Erkki Oja. 2000. "Independent Component Analysis: Algorithms and Applications." *Neural Networks* 13 (45): 411–30. doi:10.1016/S0893-6080(00)00026-5.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2017. *An Introduction to Statistical Learning. Performance Evaluation*. Vol. 103. Springer Texts in Statistics. New York, NY: Springer New York. doi:10.1007/978-1-4614-7138-7.

Jazdi, N. 2014. "Cyber Physical Systems in the Context of Industry 4.0." *2014 IEEE Automation, Quality and Testing, Robotics*, 2–4. doi:10.1109/AQTR.2014.6857843.

Papakostas, Nikolaos, James O'Connor, and Gerald Byrne. 2016. "Internet of Things Technologies in Manufacturing: Application Areas, Challenges and Outlook." In *2016 International Conference on Information Society (I-Society)*, 126–31. IEEE. doi:10.1109/i-Society.2016.7854194.

U.Farooq, M., Muhammad Waseem, Sadia Mazhar, Anjum Khairi, and Talha Kamal. 2015. "A Review on Internet of Things (IoT)." *International Journal of Computer Applications* 113 (1): 1–7. doi:10.5120/19787-1571.

Vorne. 2017. "Use OEE to Measure Improvements in Manufacturing Productivity." Accessed September 30. http://www.leanproduction.com/oee.html.

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (1): 1–23. doi:10.18637/jss.v059.i10.

Acknowledgements

The Authors would like to acknowledge the help from the partner companies for hosting the data collection activities and for providing crucial feedback regarding the analyses performed. The funding by the Technological Higher Education Association (THEA) and the Limerick Institute of Technology, scholarship providers for the primary Author, is also much appreciated.