Investigating Robustness of Energy Management Maps for SMEs

1[st]Danica Vukadinović Greetham
Knowledge Media Institute
The Open University, UK
Milton Keynes, UK
danica.vukadinovic-greetham@open.ac.uk

2[nd] Georgios Giasemidis
CountingLabLtd.
Reading, UK
g.giasemidis@gmail.com

***Abstract*—Using the data from three small businesses, we are investigating robustness of the recently proposed Recurrence Quantitive Analysis (RQA) based method for energy management of small and medium enterprises. The method consists of two phases, the training phase where the map or maps of 'usual' behaviour is obtained, and the operational phase where the new data is tested against the existing map(s). We measure how the output changes when there is a small change in input, with respect to the sampling rate, missing data and noise. Our results over three qualitatively different datasets show that the method is relatively robust and can be used for different SMEs.**

***Index Terms*—recurrence quantification analysis (RQA), energy management, energy profiles, smart meters, small and medium enterprises, visualisation, robustness, missing data**

I. Introduction

Electric energy demand in developed countries is predicted to grow due to emergent low carbon technologies. To tackle this challenge, different methods for efficient energy management of households and large systems are being developed, often based on machine learning and artificial intelligence. In research literature, a scarcity of methods is noted for solution for small and medium enterprises (SME). This is mostly due to non-existence of smart meter data and related complexity/lack of economy of scale. Namely, a wide range of different and/or customised appliances makes training of specialised algorithms much more difficult. Hence, novel, very robust solutions are sought after. On the other hand, benefits that would come from successful solutions for SMEs are large. Efficient management of energy usage for SMEs would allow them to identify unusual behaviour, optimise usage, and made informed choices about installing batteries and micro-generation. As the SMEs are numerous and their number is growing in developed countries, this is a promising area of research.

In this article, we are investigating robustness of the recently proposed Recurrence Quantitive Analysis (RQA) ([1], [2]) based method (for further details see [3], [4]) for energy management of SMEs. The method consists of two phases, the training phase where the map(s) of 'usual' behaviour is obtained, and the operational phase where the new data is tested against the existing map. Here, the robustness of this method is tested, i.e. measuring how the output changes when there is a moderate change to the input. If the output changes dramatically for the slight changes of the input, that means that the method is not very robust and it might be too sensitive to the input data. We would obviously expect that the output changes more as the changes in the input getting bigger, as one of the main features of this method is to alert to differences in behaviour.

AND Technology Research

To investigate robustness, we quantify the impact on the output of missing and noisy data in the input. For both cases we test when the changes in data occur in both training and operational phase. We then look at the different sample rate and different lengths of the training period. The method seems to be robust regarding our tests.

The article is structured as follows: in the Section II a brief overview of the method and the two datasets is given. The processes to simulate missing and noisy data in the input are explained and the results presented for two datasets and for training and operational phases when missing or noisy data is introduced in Section III. The Impact of sampling rate and the length of training period using another dataset are described in Section IV. We conclude with the summary of the main results in Section V.

II. Data and Methodology

A. Method

Energy usage time-series from small and medium businesses consist of stochastic, complex and non-linear components. Recurrence, one of the fundamental properties of dynamical processes can help analyse it. A recurrence plot [1], [5] is a visual tool that facilitates an investigation of the system by computing time dependent RQA variables [6], [7].

Only recently, RQA was applied to energy data obtained from high resolution monitoring of electricity usage [3] and was used to map different appliances, to detect faulty devices and identify unexpected usage patterns.

Here, we give a short description of the method. Given a time-series of n observations $X = \{x_1, \ldots, x_n\}$, (e.g. energy usage readings, usually at 5 min or 30 min resolution, over a period of time) the phase space is defined by transforming the readings into time-delayed vectors at each time-step, i.e. $\mathbf{Y} = \{Y_1, \ldots, Y_m\}$, where $Y_i = (x_i, x_{i-\tau}, \ldots, x_{i-(D-1\tau)}) \in \mathbb{R}^D$,

τ is the delay and D is the dimension of the phase-space, and $m = n - (D-1)\tau$.

The distance matrix $DM(i,j) = ||Y_i - Y_j||_2$, $i, j = 1, \dots, m$, is the Euclidean distance between vectors Y_i and Y_j in the phase-space. The recurrence plot follows from the distance matrix, defined as

$$R(i,j) = H(\epsilon - DM(i,j)), \tag{1}$$

where $H(x)$ is the Heaviside step function.

The entry $R(i,j)$ equals one and the states Y_i and Y_j are considered recurrent, when the distance between Y_i and Y_j is within an ϵ-radius in phase-space. As a state Y_i in phase space corresponds to a time-step of the original time-series X, recurrence plots inform us of recurring patterns within our current time-series.

The recurrence plot $R(i,j)$ is created from the input data - energy usage readings over a period of time, and 5 RQA variables are computed (see [3] for further details), using a sliding window over the diagonal of the recurrence plot. The RQA variables are then projected onto 2-d space using Principal Component Analysis, and a training map is created using density of the points in different areas to define its boundary. An example can be seen in Fig 16. New data is then tested against the map, (again computing RQA variables and projecting them onto the 2-d space) and the number of values outside the map is recorded as alerts.

From the above description, it is clear that the recurrence plot depends on three parameters, the time-delay τ, the embedding dimension D and the radius ϵ, to capture the correct dynamics of a system with noise. The parameters need to be optimised according to the observed system's characteristics and the application. In [8] several potential issues are recognised when choosing parameters. This is especially true with emerging areas of applications such as 30 min resolution electric energy data. The optimal selection of parameters for energy usage data was discussed in [4].

We will now describe the datasets that we have used to test the robustness of this method.

B. Dry cleaners

The dry-cleaners, see [3] for further details, has valid data for a period of 6 weeks from 11-th of September to 22-nd of October 2017. Figure 1 shows a typical weekly profile.

As we are interested in the dynamics while the business is open, we include only the operational times in our analysis. To identify the operational times, we searched for the maximum reading in the data between 10pm and 6am at night, where the business is closed, see Figure 1. The maximum value is $1.4A$ and any values greater than this are treated as "on" states, i.e. operational times.

In Figure 2a, we plot the distribution of readings for each day of the week. We observe that all days have similar medians and quartiles. In Figure 2b, the distribution of readings for each hour of the day is plotted. The distributions of the readings are similar for the hours between 8am to 5pm, the current usage drops between 5pm and 6pm, however after 6pm the median and quartiles of the electrical current are much lower, but higher than the "off-state". This is because most devices have been turned off and not in use, whereas one or two devices still operate until the closure of the business.

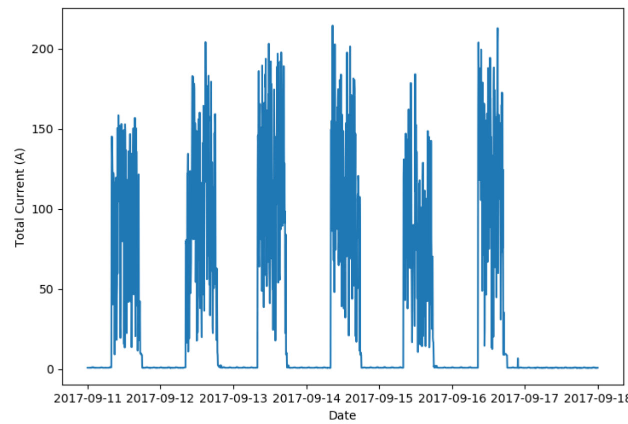


Fig. 1: A weekly profile of the total current from a dry-cleaners business. The business is closed on Sundays.

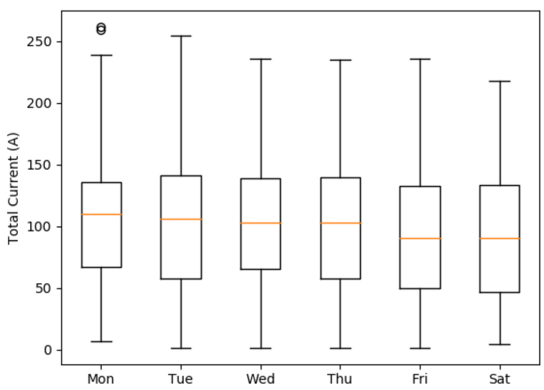
From Figure 2, we observe no daily or hourly patterns. This is expected as this is data from an SME, which operates in full demand during its operational hours, in contrast to household data which has daily and weekly behavioural patterns. The lack of strong periodicity patterns and the high standard deviation of the distributions indicate that the readings have a strong stochastic component, which is also unveiled from our analysis below.

For the dry-cleaners data, the total demand (measured in total current (A)) in the 5 min resolution from 11/09/17 until 09/10/17 was used for the training phase to create the map. The following week (from Tuesday 10/10/17 to Monday 16/10/17) was used as the operational phase, where each day was treated separately. The map's sensitivity was set to 90%. This means that we expect at least 90% of input data to be projected inside the boundary.

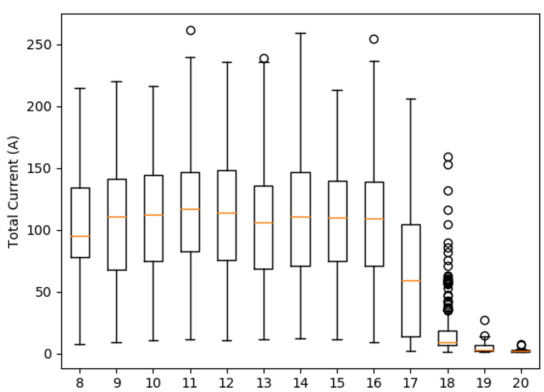
C. Butchers

A weekly profile of the butchers' data is plotted in Figure 3. In contrast to the dry-cleaners, the butchers has no obvious threshold for non-operational hours. High demand values are observed when the business is closed, due to constantly switched on fridges. For these reasons, we set the threshold for operational hours to $0A$, hence, we treat all time-steps as 'operational' from the algorithmic point of view. In Figure 4a, we plot the distribution of readings for each day of the week. We observe that all days have similar medians, but Wednesday and Thursday have wider distribution of values. In Figure 4b, the distribution of readings for each hour of the day is plotted. The median of the distributions is similar for all times of the day, but the width and standard deviation of the distributions vary from hour to hour.

For the butchers, the total demand (measured in total current (A)) at the 5 min resolution from 16/01/18 until 09/02/18 was



(a) Day of the week (the business is closed on Sundays)



(b) Hour of the day (business is closed before 8am and after 9pm).

Fig. 2: Distribution of electrical current readings for the dry-cleaners.

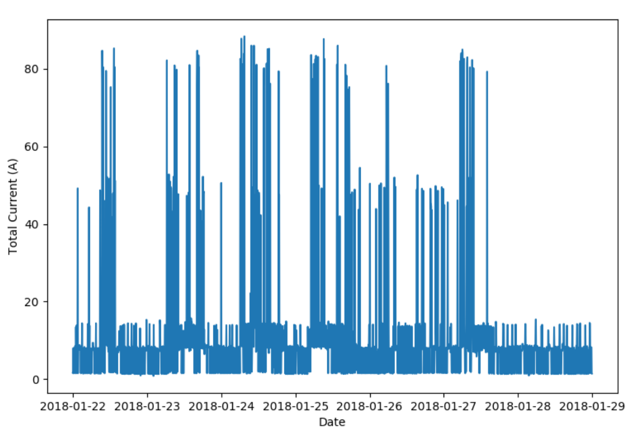


Fig. 3: A weekly profile of the total current from a butchers business.

used for the training phase to create the map. As the complete dataset corresponding to the following week (starting on 10/02/18) was not available, we used the data from Saturday 17/02/18 until Sunday 23/02/18 for the operational phase, and again each day was treated separately. The map's sensitivity was set to 90%.

III. Robustness

In order to quantify robustness of the method, we focused on two cases: missing and noisy data. In both cases due to the randomness, we repeated each experiment 10 times and reported the averages and standard deviations (reported in brackets) of the values over those 10 runs.

A. Missing data

For the missing data, we set a percentage $p\%$, where $p = 5, 10, 20, 40$, of data input to zero. For the input data of length l, we calculate $n = \lceil pl \rceil$. Then uniformly at random, a sample of n entries out of l is chosen and those entries are set to zero.

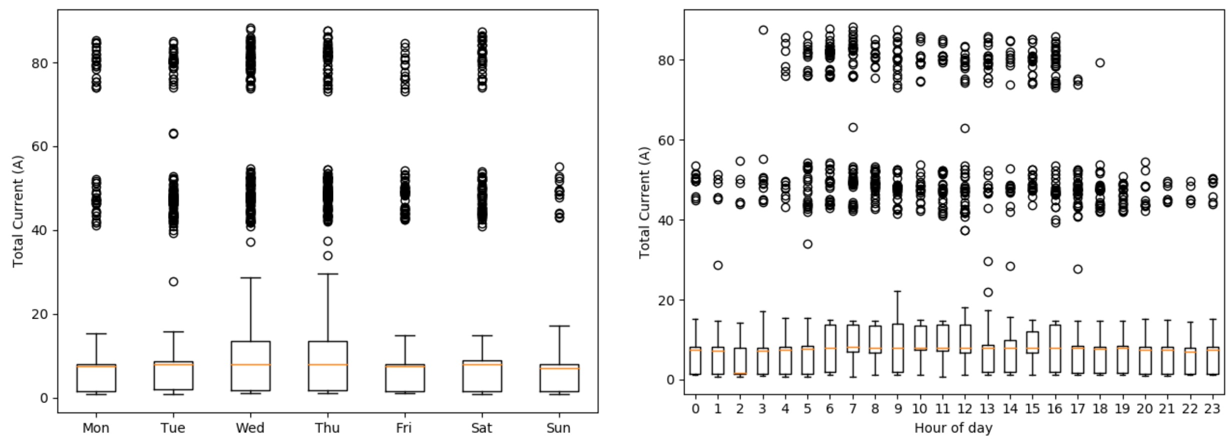
1) Missing data in the training phase for Dry cleaners: The examples of resulting maps and the map obtained from the original input are given on the Fig 16, and the corresponding optimal parameters are given in the Table I. The maps obtained with 5% and 10% look relatively similar between themselves, 20% is slightly different, while 40% looks more different. The alerts for the operational phase obtained for the original and the missing data cases are given in the Fig 5. From the figure, one can see that the percentage of alerts stays in limits of 10% which is the sensitivity of the map, therefore the results stay robust even with higher percentages of missing data in training.

TABLE I: The optimal parameters for the original training data and the training with missing data for the dry-cleaners. For the missing training data, the averages over 10 instances are reported, and the sample standard deviation is shown in brackets.

Dataset	Radius	Embedding dim.	Time delay
The original set	216.23	10	1
Missing 5%	193.38 (61.49)	8.1 (1.97)	1(0)
Missing 10%	213.7 23.67	8.6 (1.43)	1(0)
Missing 20%	220.53 (46.41)	9.1 (1.87)	1(0)
Missing 40%	187.98 (41.29)	8.1 (1.92)	1(0)

2) Missing data in the training phase for the Butcher: We are repeating the analysis for the butchers dataset. Similarly to the dry-cleaners, maps look quite similar (see Fig 17), despite the correlated movement of radius and embedding dimension, as the optimisation of the former depends on the latter (see Table II). Figure 5 confirms the robustness with the alerts staying in 10% brackets for most of the time (only 5 values in total are higher than 10%).

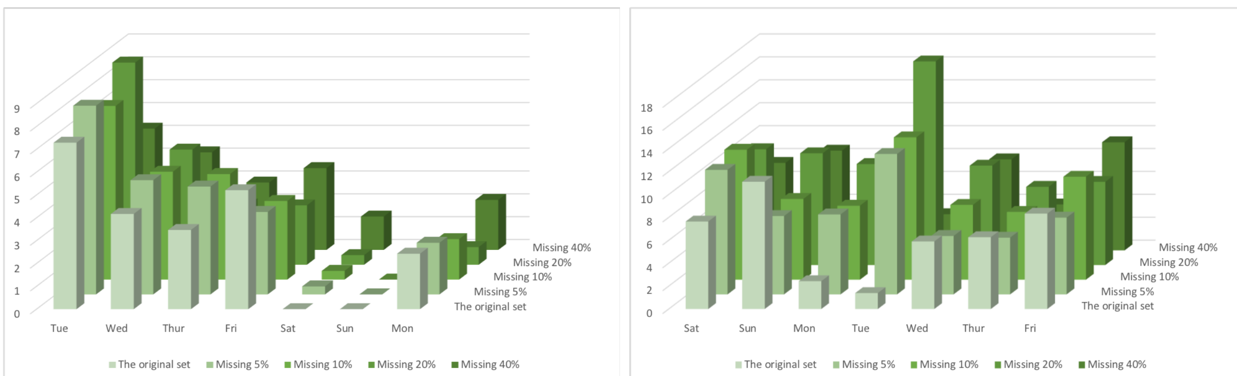
3) Missing data in the operational phase for the Dry Cleaners: Now the original data is used in the training phase to obtain the map, but the test days contain missing data. From Fig 6 we can see that missing data in the operational phase has much higher impact than in the training phase. As the



(a) Day of the week.

(b) Hour of the day.

Fig. 4: Distribution of electrical current readings for the butchers.



(a) The dry cleaners

(b) The butchers

Fig. 5: The number of alerts when the training phase contains missing data.

TABLE II: The optimal parameters for the original and missing data for the butchers

Dataset	Radius	Embedding dim.	Time delay
The original set	19.71	14	1
Missing 5%	14.03 (7.22)	10.9 (3.67)	1(0)
Missing 10%	10.94 (8.1)	8.0 (4.1)	1(0)
Missing 20%	8.81 (7.98)	6.8 (3.79	1(0)
Missing 40%	17.7 (1.44)	11.1 (0.7)	1(0)

training phase is longer, variety between the days allows for a created map to be robust. On the other hand, each day in the testing phase is impacted, causing number of alerts grows to $20-30\%$ in the cases of 20 or 40% missing data, and the total number of alerts during the testing week grows substantially (as expected less for 5 and 10% and more for 20 and 40%) .

4) Missing data in the operational phase for the Butcher: When the data missing is applied in the operational phase this has a similar effect as in the training phase, looking at the Fig 6, one can see that again most of the alert percentages are smaller than 10%, (only 4 values are slightly higher). As the butchers' data has more structure and no closing hours, it seems more robust to missing data in testing phase than the dry-cleaner's data.

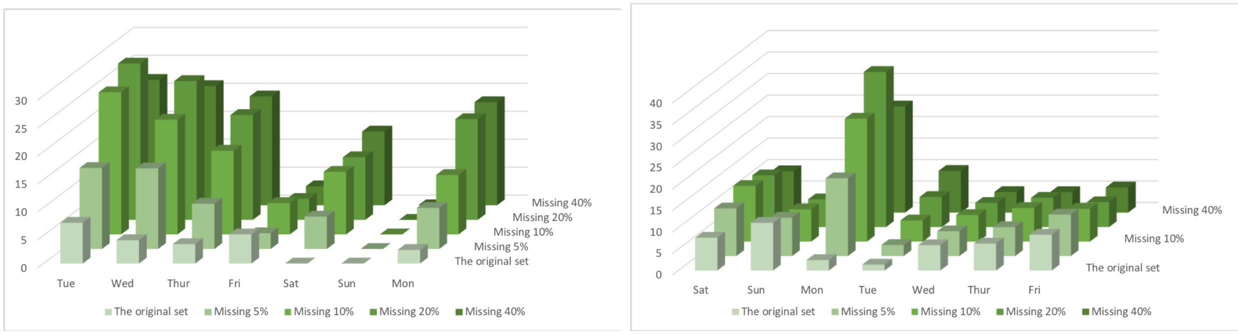
B. Adding noise

1) Noisy data in the training phase for the Dry Cleaners: From Fig 7 and Table III, we conclude that adding Gaussian noise to the dry-cleaners' training phase does change the optimal parameters, but doesn't change the map drastically. Fig 8, left panel shows that the number of alerts is inside the 10%.

TABLE III: The optimal parameters for the original and noisy data for the dry-cleaner

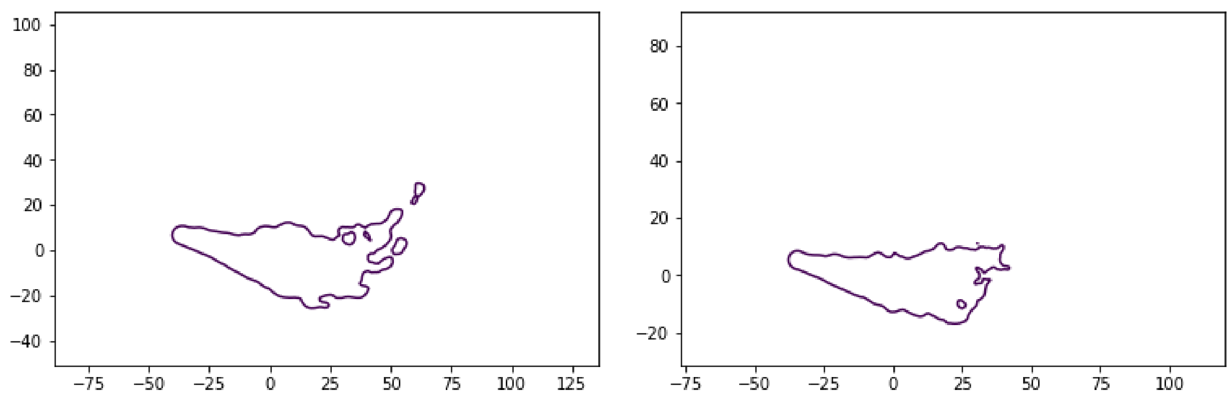
Dataset	Radius	Embedding dim.	Time delay
The original set	216.23	10	1
Noisy%	140.06 (73.56)	5.9(0.54)	1(0)

2) Noisy data in the training phase for the Butcher: Adding noise in the training phase to the butchers dataset produces



(a) The dry cleaners (b) The butchers

Fig. 6: The number of alerts when the operational phase contains missing data.



(a) The original data (b) Noisy data

Fig. 7: The maps obtained from the training phase for the dry-cleaners.

different map (Fig 9) with different optimal parameters (Table IV), which results in more alerts, see the Fig 8, right panel.

3) Noisy data in the operational phase for the Dry Cleaners: When the noise is added to the operational phase, the number of alerts increases (Fig 8, left panel). Note that Sunday is preserved, as the business is closed, but Saturday alerts increase inside the allowed envelope of 10%, as the most other days, only in the first two testing days, the percentage of alerts goes to 20% resp 13%.

TABLE IV: The optimal parameters for the original and noisy data for the butchers

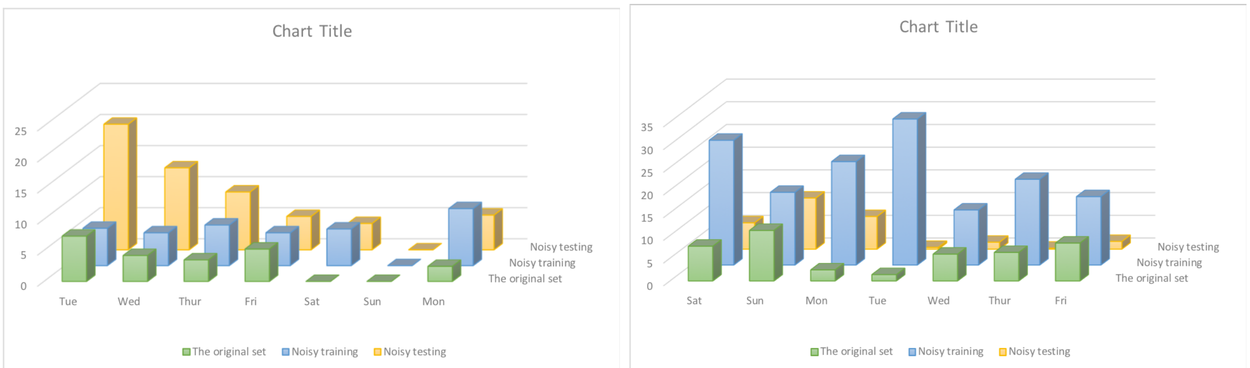
Dataset	Radius	Embedding dim.	Time delay
The original set	19.71	14	1
Noisy%	12.11 (2.73)	6.5 (0.92)	1(0)

4) Noisy data in the operational phase for the Butcher: On the other hand, the system seems quite robust to the added noise in the testing period. From the Fig 8, right panel, we see that the noise in the operational phase reduces the number of alerts each day.

C. Summary of missing and noisy data impact

We ran simulations to check the robustness of the RQA methods when training or testing input contained missing or noisy data. We used dry-cleaners and butchers' data in 5 min resolution, with the boundary set on 90%. The optimisation of parameters depends on input dataset which allows changes in optimal parameters while keeping similar shapes of maps. Our results show that in most scenarios, the method is robust on small changes in input, resulting in similar number of alerts.

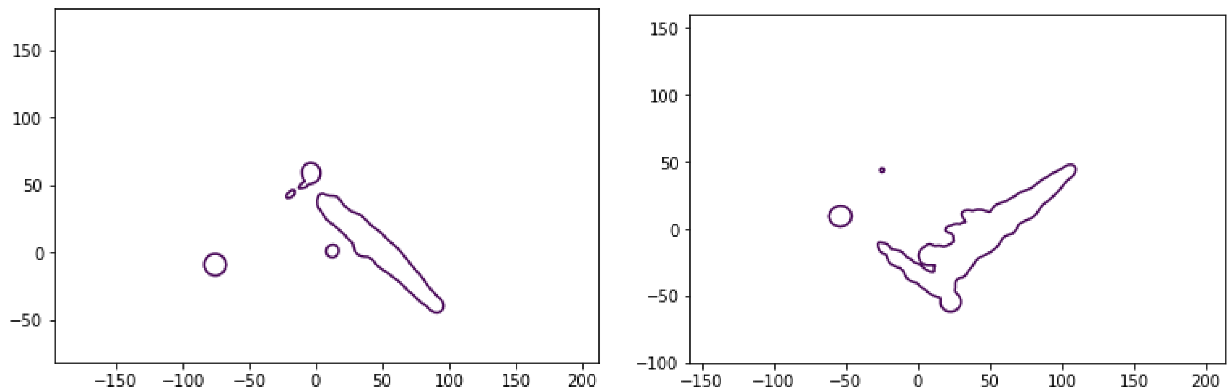
For the dry-cleaners and butchers the method was very robust to missing data in training period and could cope with up to 40% of missing data without significantly changing the maps. The same was true for butcher for missing testing data, while for dry-cleaners, the percentage of alerts moved to $20-30\%$ (from expected 10%) in the same setting. An activity threshold is used for the dry-cleaner data (as appliances are normally either switched on or off, we ignore periods of total rest, e.g. when the business is closed, in order to obtain 'tighter' maps). For that reason, it might be that the missing data impacts test days by relaxing this structure. Similar, but much smaller effect is observed when noise is added to the dry-cleaner in operational phase (noise is added only on the data that is greater than the threshold). The modified operational data ((both for noise and missing) are



(a) The dry cleaners

(b) The butchers

Fig. 8: The number of alerts when the training or operational phase contains noisy data.



(a) The original data

(b) Noisy data

Fig. 9: The maps obtained from the training phase for the butchers.

more scattered on the 2d plane (hence being more often outside the map) for the dry-cleaners.

On the other hand, butchers data copes well with missing data both in training and testing. In addition, it does so with added noise in the testing phase, but added noise in the training created more alerts over some days, between $20 - 30\%$. As the butcher's data is less volatile and more structured than the dry-cleaner's, by adding noise in the training period, the map changes, resulting projected points of the original operational data to be placed outside the map more often. In summary, our tests show that the method is relatively robust to missing and noisy data added in both training and operational phase. While there is a slight variability between robustness in different phases for two different kinds of data (butchers, more structured, no threshold, and dry-cleaner, more volatile, with an activity threshold) the largest change in the percentage of alerts (around 30%) that we observed was moderate.

IV. Office SME

In this section, we test robustness with smart meter data at half-hourly resolution of power of a small office SME. The data covers the period from 12th of May 2018 to 2nd of August 2018. In contrast to the previous sections, the analysis is now performed using the "Power" readings, which are collected via a smart meter.

A. Descriptive Analysis

In Figure 10, the whole time-series for the aforementioned period of available data is given. Regular spikes and a transition to lower consumption at about the sixth week can be observed.

In Figure 11, the daily average of the data is plotted. The transition to a state of lower consumption is now clear at the end of week 5 and the beginning of week 6.

We further investigate, the distribution of readings for each day of the week and each hour of the day, in Figure 12. The peaks shown in Figure 10 correspond to high usage between 11pm and 1am and also between 5am and 6am. This is due to overnight storage heaters. Regarding the day of the week, all days have similar distributions with Sunday having a slightly lower median.

B. Time Delay and Embedding Dimension

In this section we assess the robustness of the algorithms determining the time delay and embedding dimension [4]. We estimate the two parameters using i) longer periods of data,

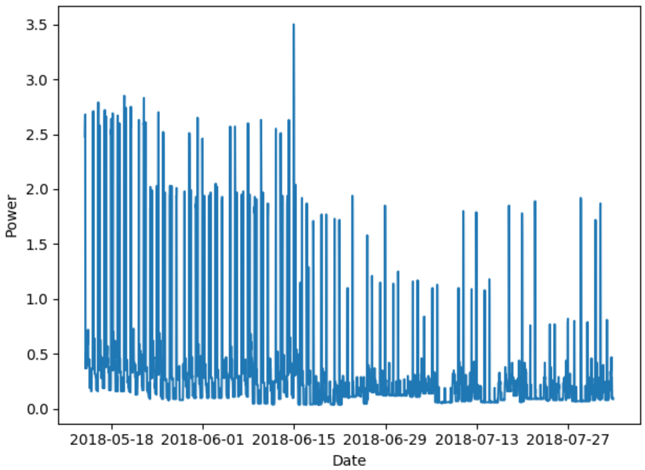
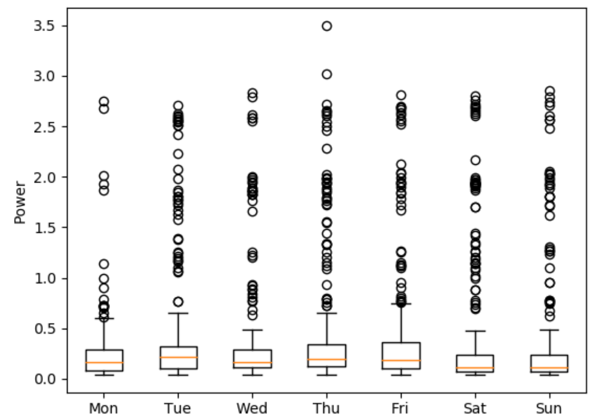


Fig. 10: Timeseries of power data for the office SME.



(a) Day of the week distribution of readings

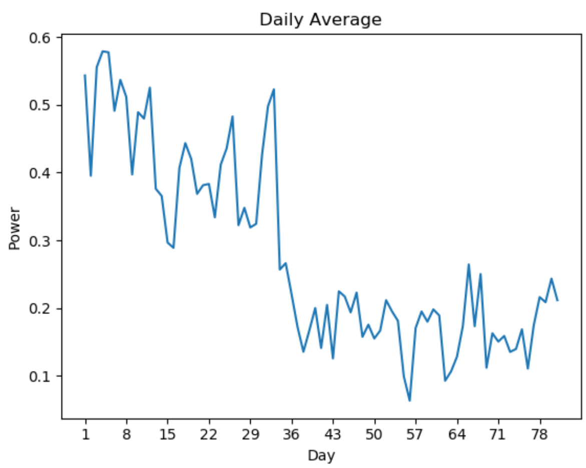
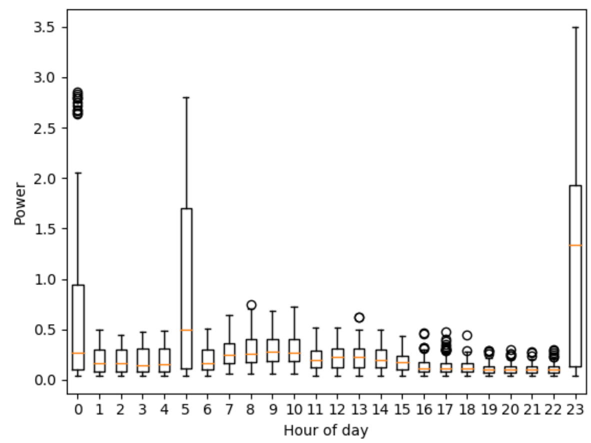


Fig. 11: Daily average



(b) Hour of the day distribution of readings

Fig. 12: Distribution of readings.

varying from one week to eleven weeks, see Figure 13a, and ii) using three weeks' data rolling over the available period, i.e. nine three-week periods, see Figure 13b.

Figure 13 shows that the embedding dimension is 16 with very few fluctuations. Also, the time delay is constantly $\tau = 1$, because there are no strong autocorrelations in the timeseries and the first minimum occurs below a threshold value. The drop to 1 for the embedding dimension happens when the optimisation algorithm does not converge.

In Figure 13a when the data in the training period includes the first transition week (week 6) the algorithm does not converge, because the sixth week is treated as noise. When more weeks of lower consumption are included in the training data, the new lower consumption data is treated as genuine signal, not as noise, and the embedding dimension recovers to the value of 16.

Also, in Figure 13b, the algorithm does not converge when the training period includes the transition point, weeks 5 and 6, for the same reasons we mentioned before.

In conclusion, the two algorithms seem to be robust over different periods of training.

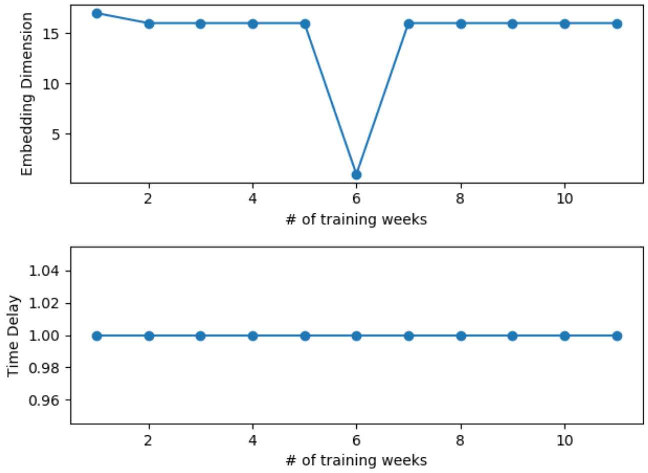
C. Radius

Finding the optimal radius is a time-consuming process (see [4]). For this reason we focus only on two training periods; a four and a five week period starting from the 12th of May 2018.

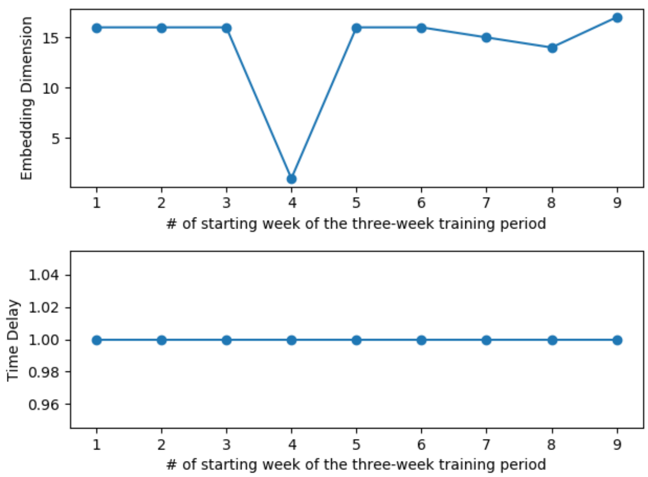
The radius optimisation depends on the optimal time delay and embedding dimension. We use the optimal values found in the previous section, $\tau = 1$ and $D = 16$, see Figure 13a.

We observe that the optimal radius occurs at low values, which also correspond to low values of the REC (2.5% and 3.5% respectively)[1]. The behaviour of the clustering measure is similar in both cases, it peaks at low values of REC and then

[1]REC is one of the RQA variables and equals to the percentage of recurrent points in a recurrence plot.

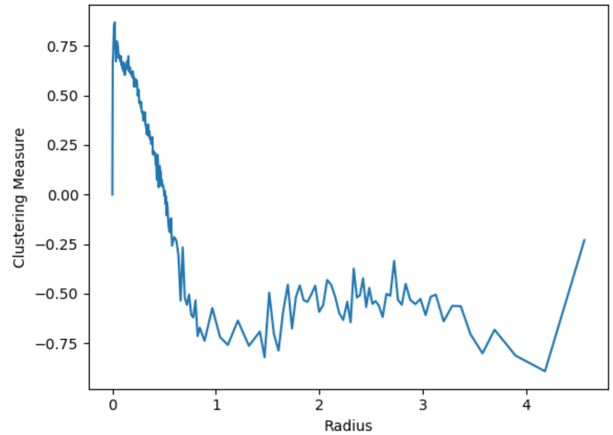


(a) Increasing number of weeks

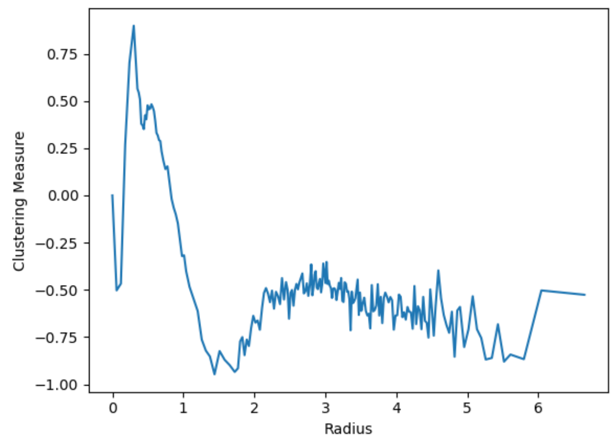


(b) Rolling period of two-weeks

Fig. 13: Optimal time delay and embedding dimension over different periods of data.



(a) 4 weeks



(b) 5 weeks

Fig. 14: Radius optimisation for two training periods of different size.

it fluctuates around a lower value at higher values of the REC. This is an indication that the radius optimisation is robust with the addition of more data.

D. Map

We use the four and five week period to train two models and produce the maps. We assess these maps on the week after the end of the training period, see Figure 15. We observe that the maps hardly change.

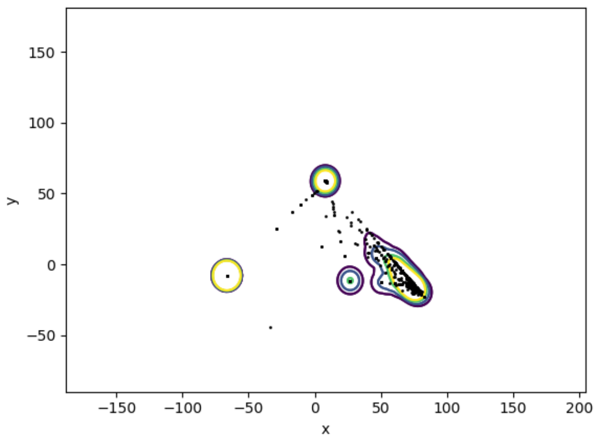
In Table V, we show the percentage of regular observations (i.e. inside the map) during training for different levels of sensitivity (expected ratio of regular observations) for the two training periods. The percentages between the two training periods do not differ, which reflects the similarity of the two maps as stated in the previous section.

Comparing to the sensitivity level, we observe that the percentage of regular readings starts deviating from the sensitivity level at values below 0.95. This is because of the low resolution data, which results in concentrated areas of projected points and a tight probability density.

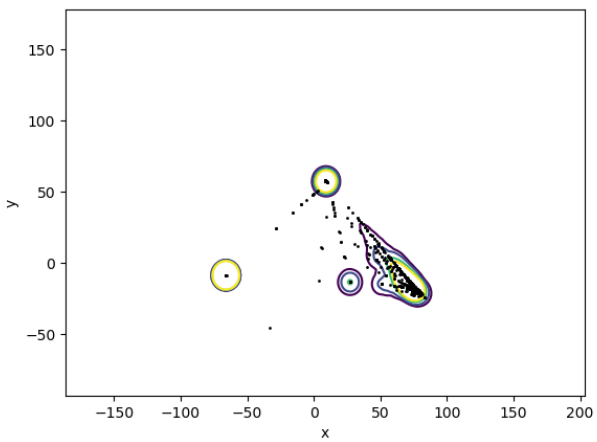
Sensitivity	% Inside (4 weeks)	% Inside (5 weeks)
0.95	0.9665	0.9679
0.9	0.9535	0.9478
0.8	0.9063	0.9018
0.75	0.8545	0.8417

TABLE V: Percentage of observations inside the map (i.e. regular observations), estimated during the training phase.

Singularly, in Table VI, we show the percentage of regular observations (i.e. inside the map) during the test week for different levels of sensitivity for the two models trained on four and five weeks respectively. We observe that the percentage of the regular observations drops significantly for all sensitivity levels for the five-week model. This is explained as follows. During week 6, there was a switch in demand behaviour to lower average values, see Figure 11 and the discussion in



(a) Contour at different sensitivity levels and projected data trained on 4 weeks



(b) Contour at different sensitivity levels and projected data trained on 5 weeks

Fig. 15: Maps for two different periods of data.

Sensitivity	% Inside (4 weeks)	% Inside (5 weeks)
0.95	0.9665	0.7685
0.9	0.9535	0.7212
0.8	0.9063	0.6367
0.75	0.8545	0.5526

TABLE VI: Percentage of observations inside the map, estimated during the test week - the week after the training period.

Section IV-A. The model has been trained on five weeks before this behavioural change and is being applied on the new behaviour. Hence, the model successfully finds more irregular data-points.

In conclusion, we demonstrated that the maps are robust under training periods with variable length. In addition, the maps identify the expected number of regular readings when the test data has the same patterns to the training data. More importantly, when the demand behaviour changes, the number of irregular points increases drastically, which can alarm the end-user for this change of behaviour.

V. Conclusions

While there are many methods in the literature for energy management of households and large systems, small and medium companies are currently to a certain degree neglected. SME energy management presents a challenge, because of the wide range of different, sometimes customised appliances; no data available (usually they have automatic metre readers instead of smart meters); very low sampling rates (automatic readings once per month); unwillingness of the owners to actively manage their energy demand, etc. For this reason, RQA based method that uses Principle Component Analysis is helpful. It produces 2-d maps - easily visualising normal behaviour and it is robust enough to be used across different businesses.

Several experiments with missing data, noise and different data resolution were run, in order to assess the robustness of the methods developed in [4]. We found that in the majority of the cases, the method is robust to the aforementioned tests. The number of alerts remains within the sensitivity levels. We demonstrated that the output does not change drastically with noise, missing data or duration of the training period. When optimal parameters vary with changes in the input, the PCA and maps co-adjust so that the final percentage of alerts is within the expected levels. We also showed, that when there is a change in behaviour, e.g. a transition to a lower demand, this can be identified by the proposed method.

In future work, we plan to test the method on similar and different businesses to the ones used. We also hope to look at seasonal trends (if they exist) and how they influence maps. Finally, we aim to apply these techniques for map creation to individual appliances and their combinations in order to understand the disaggregation of demand of SMEs. This will require more data to become available.

Acknowledgment

The work was supported by AND Technology Research through NSDEMIC (UK Government BEIS).

Appendix

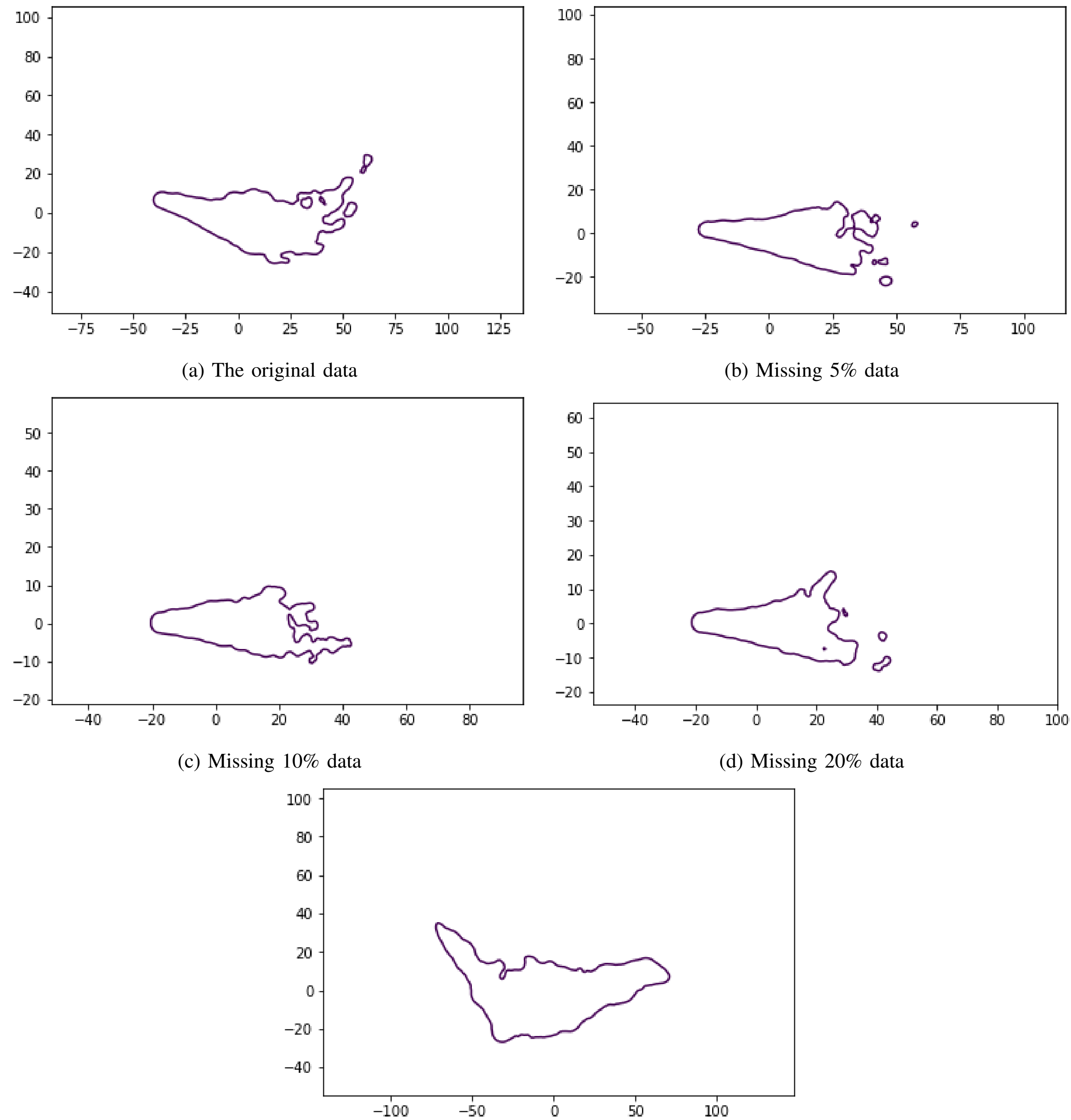
References

[1] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, “Recurrence plots of dynamical systems,” *EPL (Europhysics Letters)*, vol. 4, no. 9, p. 973, 1987.

[2] N. Marwan, M. Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” *Physics Reports*, vol. 438, no. 5, pp. 237–329, 2007.

[3] L. Hattam and D. Vukadinović Greetham, “Energy disaggregation for SMEs using Recurrence Quantification Analysis,” in *Proceedings of the Ninth International Conference on Future Energy Systems*, ser. e-Energy ’18. New York, NY, USA: ACM, 2018, pp. 610–617. [Online]. Available: http://doi.acm.org/10.1145/3208903.3210280

[4] G. Giasemidis and D. Vukadinović Greetham, “Optimising parameters inRecurrence Quantification Analysis of Smart Energy Systems,” in *The 9th International Conference on Information, Intelligence, Systems and Applications (IISA 2018)*, ser. IASC, 2nd Workshop on Intelligent Applications in Smart Cities, 2018.



(a) The original data

(b) Missing 5% data

(c) Missing 10% data

(d) Missing 20% data

(e) Missing 40% data

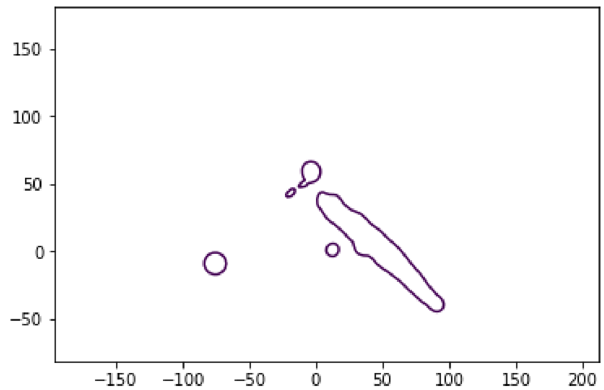
Fig. 16: The maps obtained from the training phase for the dry-cleaner.

[5] J. Zbilut and C. Webber, “Embeddings and delays as derived from quantification of recurrence plots,” *Physics Letters A*, vol. 171, no. 3, pp. 199–203, 1992.

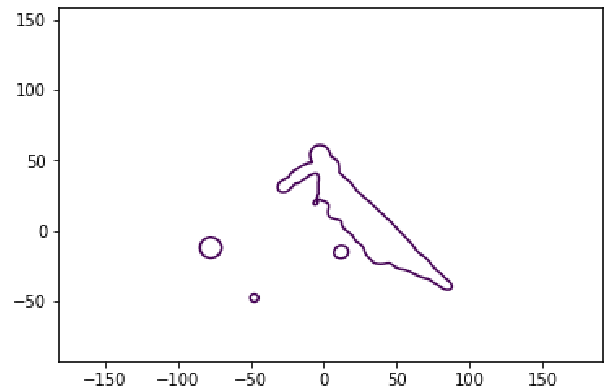
[6] A. Fabretti and M. Ausloos, “Recurrence plot and recurrence quantification analysis techniques for detecting a critical regime,” *International Journal of Modern Physics C*, vol. 16, no. 05, pp. 671–706, 2005.

[7] N. Marwan, S. Schinkel, and J. Kurths, “Recurrence plots 25 years later-gaining confidence in dynamical transitions,” *EPL (Europhysics Letters)*, vol. 101, no. 2, p. 20007, 2013.

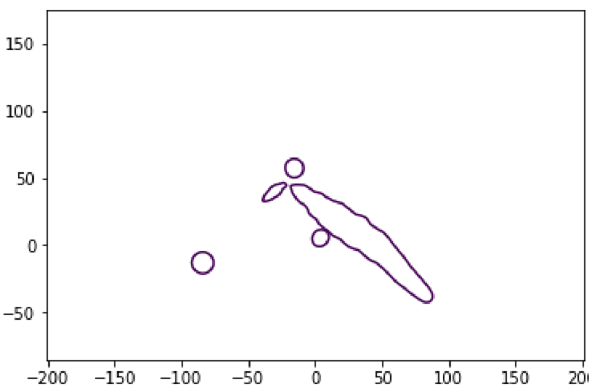
[8] N. Marwan, “How to avoid potential pitfalls in recurrence plot based data analysis,” *International Journal of Bifurcation and Chaos*, vol. 21, no. 04, pp. 1003–1017, 2011. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/S0218127411029008



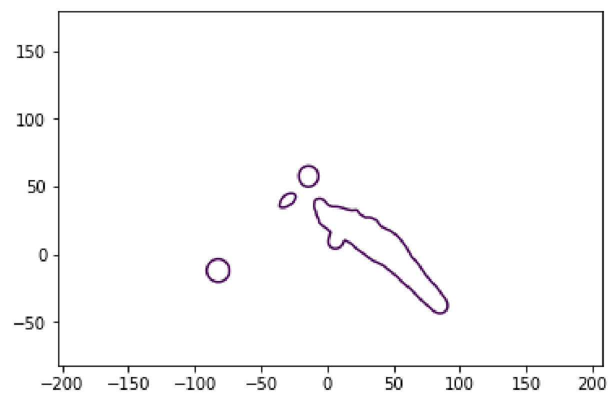
(a) The original data



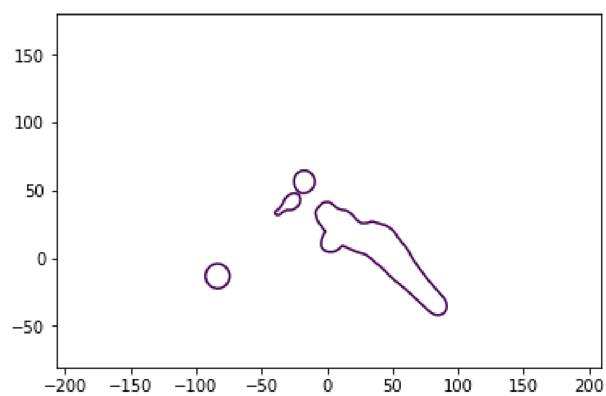
(b) Missing 5% data



(c) Missing 10% data



(d) Missing 20% data



(e) Missing 40% data

Fig. 17: The maps obtained from the training phase for the butchers.