On the Relationship between Seasons of the Year and Disaggregation Performance

João Gois
ITI, LARSyS, ARDITI, Portugal
joao.gois@iti.larsys.pt

Christoph Klemenjak
University of Klagenfurt, Austria
klemenjak@ieee.org

Lucas Pereira
ITI, LARSyS, Técnico Lisboa, Portugal
lucas.pereira@tecnico.ulisboa.pt

ABSTRACT

This paper pursues the question of how seasons of the year affect disaggregation performance in Non-Intrusive Load Monitoring. To this end, we select the dishwasher, a common household appliance that may exhibit usage cycles depending on the user. We utilize an auto-correlation function to detect usage patterns of dishwashers in each season. Then, we examine the dissimilarity across each season with the help of the Keogh Lower Bound measure. Finally, we conduct a disaggregation study using the REFIT dataset and relate the outcome to the dissimilarity across seasons. Our findings indicate that in cases where energy consumption shows similarity throughout seasons, the performance of load disaggregation approaches can be positively affected.

CCS CONCEPTS

• **General and reference → Performance**; **Evaluation**; **Empirical studies**.

KEYWORDS

NILM, Seasonality, Auto-Correlation, Similarity, Performance

ACM Reference Format:
João Gois, Christoph Klemenjak, and Lucas Pereira. 2020. On the Relationship between Seasons of the Year and Disaggregation Performance. In *The 5th International Workshop on Non-Intrusive Load Monitoring (NILM'20), November 18, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3427771.3427856

1 INTRODUCTION

The continuous reduction of greenhouse gas (GHG) emissions is a growing need for achieving a more sustainable worldwide energy distribution. There have been several initiatives throughout the years to discuss the impacts resulting from improper energy usage. In the 2008 climate change act, one of the commitments of the UK government was to achieve 80% reduction in GHG emissions by 2050, compared with a 1990 baseline [9].

Energy efficiency of the domestic sector must improve to lower emissions and mitigate the risks of global climate change [12]. It is emphasized that electricity consumption of households depends on a complex series of interlinked and interacting socio-economic, dwelling, and appliance-related factors. Hence, there can be high consumption volatility among users. Furthermore, in [6], seasonal variation is a visible feature in house occupancy levels, e.g., lower consumption in evening hours for warmer months, as opposed to the remaining colder months.

Seasonality effects are widely studied in the context of load forecasting [11]. The results stress that without accounting for seasonality, the quality of predictions is seriously compromised. To the best of our knowledge, the effects of seasons on the performance of NILM algorithms have received little attention by the community. Instead, the most common method to evaluate NILM algorithms consists of selecting test sets with varying sizes that have not been used for training, independently of the season [15].

In this work, we study how season transitions affect the performance of NILM algorithms. To this end, we propose two auto-correlation based indicators to capture temporal correlations and assess the similarity of appliance consumption time series. The remainder of this paper is structured as follows: Section 2 presents background and related work. Section 3 provides insights on the appliance data used in our work followed by three case studies. Finally, we present conclusions, limitations, and an outlook on future work in Section 4.

2 BACKGROUND AND RELATED WORK

2.1 Correlation and Auto-Correlation

Statistical correlation summarizes the strength of the relationship between two random variables. Pearson's correlation coefficient [13] can be used to measure the linear association of two variables, which is a number in $[-1, 1]$, where values close to -1 indicate a negative correlation, while values in the opposite direction indicate a positive correlation. The zero value indicates no correlation at all.

In time series analysis, an important feature is the (potential) serial correlation. For a time series y_t, the correlation coefficient between two values is called Auto-Correlation Function (ACF) [3], given by:

$$\mathrm{Cor}(y_t, y_{t-k}) = \frac{Cov(y_t, y_{t-1})}{\sqrt{Var(y_t)Var(y_{t-k})}}, \quad k = 1, 2, \ldots \tag{1}$$

where k is the time gap considered between observations. In general, the correlation between values that are k time periods apart is called lag-k correlation. In a stationary time series, the moments of random variables are non-changing over time, thus the index t is dropped and the ACF is written as function of lag k

$$\rho(k) = \mathrm{Cor}(y_t, y_{t-k}), \quad k = 1, 2, \ldots \tag{2}$$

NILM'20, November 18, 2020, Virtual Event, Japan

ACM ISBN 978-1-4503-8191-8/20/11…$15.00
https://doi.org/10.1145/3427771.3427856

One of the approaches to estimate the ACF is by means of the Pearson's correlation coefficient

$$\tilde{\rho}(k) = \frac{\sum_{s=1}^{n-k}(x_{s+k} - \bar{x}_{(k)})(x_s - \bar{x}_{(1)})}{\sqrt{\sum_{s=k+1}^{n}(x_s - \bar{x}_{(k)})^2 \times \sum_{t=1}^{n-k}(x_t - \bar{x}_{(1)})^2}}, \quad k = 1, 2, \ldots \tag{3}$$

where $\bar{x}_{(1)} = \frac{1}{n-k}\sum_{i=1}^{n-k} x_i$ and $\bar{x}_{(k)} = \sum_{i=k+1}^{n} x_i$. This measure also gives insight on the amount of explained variability: the square of the Pearson correlation coefficient gives the percentage of variability of x_{t+k} that is explained by x_t. The drawback of this approach is that the number of data pairs decreases at higher lags, which leads to less precise estimations as k increases. Alternatively, a plug-in approach based on estimated auto-covariances can be used, and the estimate for the k-th auto-correlation coefficient terms turn out to be in

The estimation of ACF from an observed time series assumes that the underlying process is stationary. However, the formulae given previously can still be applied to non-stationary series. The ACF plots then usually exhibit some typical patterns. Therefore, the time series is decomposed into *trend*, *seasonality*, *cyclic* and *residual* components, which are taken into account while finding correlations on the data.

The trend is a long-term average tendency of the data to increase (upward) or decrease (downward) over time, either linear or non-linear. The ACF is generally large and positive for small lags with a slow decay as the lag increases. With a different nature, seasonal variations (seasonality) operate in a regular and periodic manner over a span of less than a year. It is present in periodically recorded data, e.g., daily, weekly, or monthly. The ACF plots exhibit more significant spikes at multiples of the seasonal frequency. Cyclic variations operate over more than one year and the ACF peaks around the average cycle length. The residual component (or remainder) is a factor that causes purely random or irregular variation in the variable under study. In this work, seasonality analysis is of special interest.

2.2 ACF Applications to Load Disaggregation

Literature in ACF applications is not vast, and existing attempts are mostly to gain additional insights in consumption data. For example, in [8] and [7], ACF is used to detect periodicity and load patterns in Hilbert transforms of pulse waveforms from energy consumption data. It is emphasized that the properties of the integral transform and the use of statistical tools, such as the ACF, provides valuable information to the end-use disaggregation: level of demand and duty cycles of appliances for a given household and specific season.

In [10], the temporal structure of power time series is analyzed via first-order auto-correlation to compare energy consumption in commercial and residential buildings. Among other results, it was highlighted that commercial buildings have more evident seasonal patterns than residential ones, which is related to people's habits and time-scheduled equipment.

3 MATERIALS AND METHODS

3.1 Appliance Consumption Data

We focus on the analysis of dishwasher consumption since it may exhibit seasonal ACF patterns (e.g. daily or weekly usage) depending on the user . For example, in [20], it is concluded that fully-loaded modern dishwashers use less electricity, water, and detergent than even the most efficient hand-washers. Hence, it is expected that more conscious householders will do full loads, resulting in seasonal ACF patterns. Such a seasonal pattern contrasts with the fridge, which exhibits a recurrent cycling pattern over time.

In the course of this work, we employ data of dishwasers provided by the REFIT dataset [16], which is composed of electrical consumption measurements, both aggregate and appliance levels, for 20 households in the UK. The original data are timestamped and sampled at 8 second intervals.

To capture all four seasons of a year, the period in which the experiments are performed is exactly one year - ranging from 2014-06-21 to 2015-06-20. During this period, we identified a dishwasher in 15 of 20 households in REFIT. Houses 4, 8, 12, 16 and 18, do not have a dishwasher. We divided the data according to the seasons in the UK as follows: i) Summer: from 2014-06-21 to 2014-09-21; (ii) Fall: from 2014-09-22 to 2014-12-20; (iii) Winter: from 2014-12-21 to 2015-03-19; (iv) Spring: from 2015-03-20 to 2015-06-20.

3.2 Seasonality Identification with ACF

In this section, the ACF is used to unveil seasonality patterns in consumption data of dishwashers. The data was down-sampled to one sample per minute. At each activation, the dishwasher is used for a considerable time span (in view of minutes). Therefore, this granularity is sufficient to preserve usage patterns albeit working with down-sampled versions of the original data.

A two-step approach was followed for seasonality identification: (1) ACF calculation, and (2) identification of the most significant peaks. Concerning the first, ACFs are calculated for each season using the `plot_acf` function from Python's package `statsmodels`[1]. In order to capture household routines, which normally follow a weekly basis, a different ACF is calculated for each consecutive two weeks of data. The second step is based on the concept of topographic prominence in a signal, which measures the vertical distance between the peak and its lowest contour line. The inspiration for using this technique comes from real-world detection of mountains [14]. The implementation in Python can be seen in detail in [19].

3.2.1 Results and Discussion. The resulting ACFs and prominent peaks were then analyzed for each household to find which ones exhibit seasonal usage of the dishwasher. For illustration purposes, Figure 1 shows the ACF plot for household 10 during the first two weeks of summer. The prominent peaks occur approximately every 1440 lags, which indicates the dishwasher is used once a day (i.e., every 1440 minutes).

In contrast, for household 17, the ACF does not exhibit any seasonal pattern as can be observed in Figure 2, where several peaks represent irregular uses of the dishwasher.

The results obtained for the dishwashers in each household are summarized in Table 1. It should be noted that some households exhibit seasonal ACF pattern only for one season, while household 5 keeps that pattern for all seasons. On the other hand, the seasonal ACF pattern was not noticed for households 1, 3, 11, 13, 14, 19, and

[1]Python statsmodels, https://www.statsmodels.org/

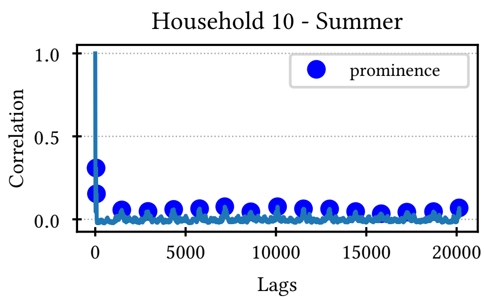

Figure 1: Autocorrelation function of the dishwasher in household 10 for the first two weeks in summer.

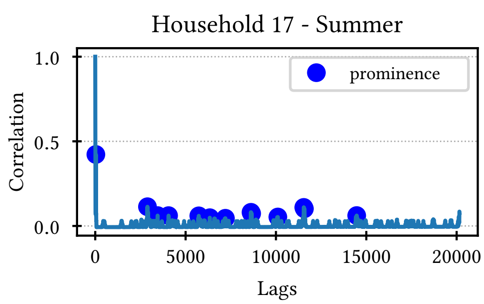

Figure 2: Autocorrelation function of the dishwasher in household 17 for the first two weeks in summer.

20. Also, only household 5 has a seasonal ACF in spring, which may occur because most households have missing values for this season. Ultimately, for about 85% of the cases the ACF pattern is not seasonal, which means that in general the householders in this study do not use the dishwasher according to a pattern, as initially assumed.

Table 1: Seasonality results for the first two weeks of each season. Seasonal patterns are identified with "X".

Season	Households														
	1	2	3	5	6	7	9	10	11	13	14	15	17	19	20
Summer				X		X	X	X				X			
Fall		X		X	X			X					X		
Winter		X		X				X				X			
Spring				X											

3.3 Seasons Comparison with ACF

In this section, we compare the dishwashers' ACF in the transitions between consecutive seasons. To this end, the similarity between ACFs is assessed using the Keogh Lower Bound (LBK) [18], which is a lower-bounding measure for Dynamic Time Warping (DTW). LBK minimum bounds can be set up to check a lesser number of alignments between the series and quickly determine candidates for the set of nearest neighbours. Therefore it enables to reduce the computational time and quadratic complexity inherent in DTW [5]. LBK function is defined as

$$LBK(Q,C) = \sum_{i=1}^{n} (c_i - U_i)^2 I(c_i > U_i) + (c_i - L_i)^2 I(c_i < L_i) \quad (4)$$

where $U_i = max(q_{i-r} : q_{i+r})$ and $L_i = min(q_{i-r} : q_{i+r})$ are, respectively, upper and lower bounds for time series Q for reach r and $I(.)$ the indicator function.

Figure 3 exhibits the dishwasher similarity scores for the summer-fall, fall-winter, and winter-spring season transitions, obtained by comparing the last two weeks of a season and the first two weeks of the next one.

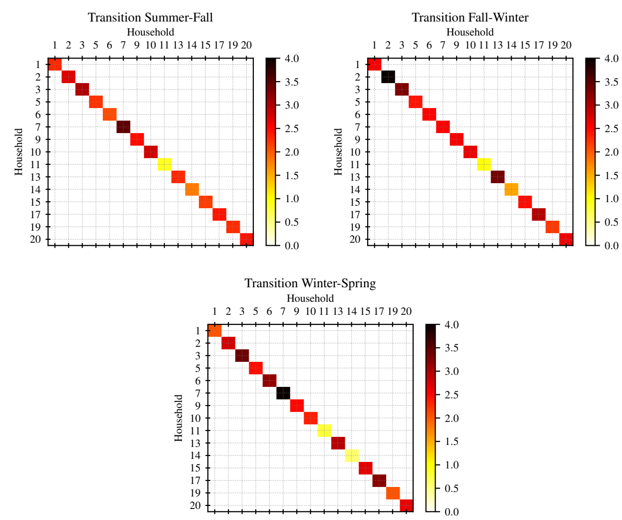

Figure 3: LBK scores of ACF comparison across seasons.

3.3.1 Results and Discussion. For this analysis, the households were split into two groups according to the level of similarity among the dishwashers. More precisely, dishwashers with a LBK score greater than 2.5 are considered to have higher dissimilarity. The obtained results are summarized in Table 2.

Table 2 seems to be the opposite of the previous table, with 60 out of the 75 cases reporting similarity across seasons. This leads to the conclusion that the consumption does not seem to change that much across seasons, even though there are no specific seasonal ACF patterns. Also, although the ACF pattern is seasonal for household 2 in both fall and winter, the ACFs are not similar across those seasons (figure 3).

Table 2: LBK summarized results for each transition. "X" represents the households with higher dissimilarity among the dishwashers.

Transition	Households														
	1	2	3	5	6	7	9	10	11	13	14	15	17	19	20
Summer-Fall		X	X			X		X							X
Fall-Winter	X	X	X							X					
Winter-Spring			X		X	X				X			X		X

3.4 Seasons and Disaggregation Performance

In this section, we present the results of a disaggregation experiment of the dishwasher from the aggregate data. This was performed to assess if the similarity of dishwashers across season transitions affects the algorithm performance. Since the original data was not necessarily sampled every 8 seconds, it was decided to down-sample to 10 seconds for the disaggregation experiment.

3.4.1 Training and Test Data. For this experiment, two test sets were considered from the aggregate data, each with two consecutive weeks of data (T1 and T2). More precisely, T1 and T2 contain, respectively, aggregate data from the last 2 weeks before season change and the first 2 weeks of the new season. For each test set, the same training set is used, each of which contains only aggregate data from the first season (the whole season excluding the last 2 weeks). Table 3 summarizes the train and test sets for each season transition.

Table 3: Train and test sets used in the disaggregation of dishwashers.

Train	Test 1 (T1)	Test 2 (T2)
2014/06/21-2014/09/07	2014/09/08-2014/09/21	2014/09/22-2014/10/05
2014/09/22-2014/12/06	2014/12/07-2014/12/20	2014/12/21-2015/01/04
2014/12/21-2015/03/05	2015/03/06-2015/03/19	2015/03/20-2015/04/03

3.4.2 Disaggregation Algorithm and Performance Metric. The disaggregation algorithm used in this experiment is the Factorial Hidden Markov Model (FHMM), which is available in NILMTK [2]. The selected performance metric is the Mean Absolute Error normalized by standard deviation (NMAE) [17].

3.4.3 Results and Discussion. To assess if similarity across season transitions affects the algorithm performance, it is necessary to compare the performance between the households that exhibit similar ACFs and the ones that do not as presented in the previous sub-section.

Looking at the results for both T1 and T2 (Figure 4 and Figure 5, respectively), for transitions fall-winter (fw) and winter-spring (ws), the performances in the presence of a seasonal pattern (highlighted with _s in the figure) appear to be better.

In contrast, in the transition summer-fall (sf), the distributions are not conclusive. Yet, for this transition the performance is comparable, independently of the similarity in the ACFs. This is especially visible in T2 where the median value is the same. Ultimately, these results suggest that the disaggregation performance is positively affected in households with similar ACFs in the transitions across seasons. Furthermore, the fact that the same behavior does not occur in the transition from summer to fall indicates that other factors are influencing the performance. In this particular case, the performance is equally good independently of the similarity between ACFs, which may indicate that at some levels of complexity the disaggregation will not vary significantly. Still, in future work, it would be interesting to explore this specific transition in more detail.

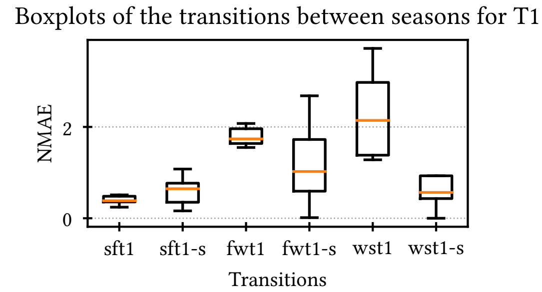

Figure 4: Comparison of performance between households with similar ACF in the transitions between seasons for T1.

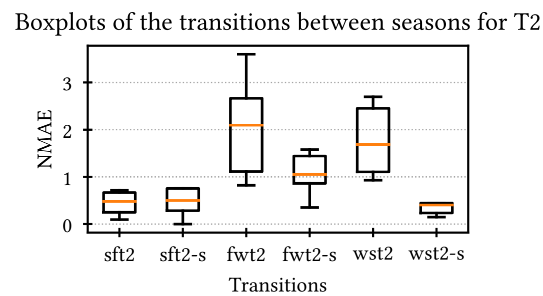

Figure 5: Comparison of performance between households with similar ACF in the transitions between seasons for T2.

4 CONCLUSIONS AND FUTURE WORK

In this work, we show how the ACF can be used to identify seasonality and compare energy consumption patterns across seasons for the dishwasher appliance in the REFIT dataset. Also, we provide a disaggregation experiment to assess to what extent similar consumption among dishwashers affects NILM performance. One limitation of this work is that the amount of data used is short to enable fair comparisons. In fact, we would need at least two years of data to assess if the results hold across seasons. A second limitation is we have only considered one NILM algorithm. Nonetheless, the results show that seasonality in dishwashers is not as common as initially expected. On the other hand, the consumption patterns seem to be similar independently of the seasons, which can potentially improve the performance of the NILM algorithms.

It is important to remark that these results are limited to one country, thus future work should explore data from different regions. Likewise, it is of interest to consider different NILM algorithms and other appliances, such as the washing machine. Finally, study and compare other measures to assess seasonality and similarity of time series, e.g. Partial Auto Correlation (PACF) [4] for the former and Global Alignment Kernels (GAK) [1] for the latter.

ACKNOWLEDGMENTS

João Gois received funding from project M1420-01-0247-FEDER-000018. Lucas Pereira received funding from FCT under grants CEECIND/01179/2017, and UIDB/50009/2020.

REFERENCES

[1] Mourtadha Badiane, Martin O'Reilly, and Pádraig Cunningham. 2018. Kernel Methods for Time Series Classification and Regression. In *Proceedings for the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2018)*. CEUR-WS, Dublin, Ireland, 12. http://ceur-ws.org/Vol-2259/

[2] Nipun Batra, Rithwik Kukunuri, Ayush Pandey, Raktim Malakar, Rajat Kumar, Odysseas Krystalakos, Mingjun Zhong, Paulo Meira, and Oliver Parson. 2019. Towards reproducible state-of-the-art energy disaggregation. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. ACM, New York, NY, USA, 193–202.

[3] G.E.P Box, G.M. Jenkins, and G.C. Reinsel. 1994. *Time Series Analysis: Forecasting & Control* (3rd ed.). Prentice Hall.

[4] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung. 2015. *Time Series Analysis: Forecasting and Control*. John Wiley and Sons Inc.

[5] A Dalal and Jamie Olson. 2007. Feature selection for prediction of user-perceived streaming media quality. In *Proceedings of International Symposium on Performance Evaluation of Computer and Telecommunication Systems SPECTS*. San Diego, CA, USA, 285–294.

[6] Amir Fazeli and Mark Gillott. 2013. Analysing the effects of seasonal variation on occupancy in an electricity demand model. *International Journal of Low-Carbon Technologies* 8, 4 (2013), 282–288.

[7] Antonio Gabaldon, Roque Molina, Alejandro Marin-Parra, Sergio Valero-Verdu, and Carlos Álvarez. 2017. Residential end-uses disaggregation and demand response evaluation using integral transforms. *Journal of Modern Power Systems and Clean Energy* 5, 1 (2017), 91–104.

[8] Antonio Gabaldón, Mario Ortiz-García, Roque Molina, and Sergio Valero-Verdú. 2014. Disaggregation of the electric loads of small customers through the application of the Hilbert transform. *Energy Efficiency* 7, 4 (2014), 711–728.

[9] HM Government. 2008. Climate change act 2008. [Accessed 9 February 2020].

[10] Simon Henriet, Umut Simsekli, Gaël Richard, and Benoit Fuentes. 2018. Energy Disaggregation for Commercial Buildings: A Statistical Analysis. In *», International Workshop on Non-Intrusive Load Monitoring (NILM2018)*. Austin, TX, USA, 1–4.

[11] Maria Jacob, Cláudia Neves, and Danica Vukadinović Greetham. 2020. Short term load forecasting. In *Forecasting and Assessing Risk of Individual Electricity Peaks*. Springer, 15–37.

[12] Rory V Jones, Alba Fuertes, and Kevin J Lomas. 2015. The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renewable and Sustainable Energy Reviews* 43 (2015), 901–917.

[13] Wilhelm Kirch. 2008. Pearson's correlation coefficient. *Encyclopedia of Public Health; Springer: Dordrecht, The Netherlands* (2008), 1090–2013.

[14] Andrew Kirmse and Jonathan de Ferranti. 2017. Calculating the prominence and isolation of every mountain in the world. *Progress in Physical Geography* 41, 6 (2017), 788–802.

[15] Christoph Klemenjak, Stephen Makonin, and Wilfried Elmenreich. 2020. Towards Comparability in Non-Intrusive Load Monitoring: On Data and Performance Evaluation. In *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, Washington DC, DC, USA, 5.

[16] David Murray, Lina Stankovic, and Vladimir Stankovic. 2017. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Scientific Data* 4 (1 2017), 160122. https://doi.org/10/f9k7k9

[17] Lucas Pereira and Nuno Nunes. 2018. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, 6 (Nov. 2018), e1265. https://doi.org/10/gfb8gr

[18] Toni M Rath and R Manmatha. 2002. Lower-bounding of dynamic time warping distances for multivariate time series. *University of Massachusetts Amherst Technical Report MM* 40 (2002), 1–4.

[19] Python scipy. [n.d.]. Topographic Prominence. https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.peak_prominences.html. Accessed: 2020-07-24.

[20] R Stamminger, R Badura, G Broil, S Dörr, and A Elschenbroich. 2003. A European comparison of washing dishes by hand. In *University of Bonn, Proceedings of the International Conference on Energy Efficiency in Domestic Appliances and Lighting (EEDAL), Turin, Italy.*