University of Bath

SAX-Quantile Based Multiresolution Approach for Finding Heatwave Events in Summer Temperature Time Series

Manuel Herrera [a], Aida A. Ferreira [b],
David A. Coley [a] and Ronaldo R. B. de Aquino [c]

[a] *EDEn - ACE, University of Bath*
Bath, UK
E-mail: amhf20, d.a.coley @bath.ac.uk
[b] *Federal Institute of Pernambuco*
Recife, Brazil
E-mail: aidaferreira@recife.ifpe.edu.br
[c] *Federal University of Pernambuco*
Recife, Brazil
E-mail: rrba@ufpe.br

Time series pattern discovery is of great importance in a large variety of environmental and engineering applications, from supporting predictive models to helping to understand hidden underlying processes. This work develops a multiresolution time series method for extracting patterns in weather records, particular temperature data. The topic is important, as, given a warming climate, morbidity and mortality are expected to rise as heatwave frequency and intensity increase. By analysing summer temperature quantiles at different levels of coarseness, it was found that compounding models can contain a complete description of severe weather events. This new multiresolution quantile approach is developed as an extension of the symbolic aggregate approximation of the temperature time series in which quantiles are computed at every stretch of the piecewise partition. The process is iterated at different scales of the partition, and it was found to be a very useful approach for finding patterns related to both heatwave periods and intensities. The method is successfully tested using real weather records from Brazil (Recife) and the UK (London), and it was found that in both locations heatwave intensity and frequency are increasing at a substantial rate. In addition, it was found that the rate of increase in intensity of the heatwaves is far outstripping the rate of increase in mean summer temperature: by a factor of 2 in Recife and a factor of 6 in London. The approach will be of use to those looking at the impact of future climates on civil engineering, water resources, energy use, agriculture and health care, or those looking for sustained extreme events in any time series.

Introduction

Multiresolution time series analysis [1] decomposes the data into components at different frequencies. The approach is often used to allow ordinary predictive models using smaller time frames to deal with higher frequency components, but automatically larger time steps for the lower frequencies. Multiresolution time series analysis also helps noise removal processes and data compression [2].

This paper proposes a time series decomposition in various time-steps, thereby generating new data over different scales from a common source. As the proposal focuses on summer temperature time series, the desire is to better explain the extremes of the historical series in comparison to only using mean conditioned models. The work proposes a multiresolution quantile (MRQ) approach, computing the quantiles for each level of resolution in the time series.

In order to represent the warmer temperatures, the quantiles under analysis need to be focused on both upper and lower quantiles. On one side, quantile regressions at upper values gather how weather interacts with the near maximum temperatures during the day. On the other side, analysing lower quantiles provides insights about when these values are relatively high during the night and the potential factors affecting these circumstances. This is key, as it is known that the lack of a diurnal cycle was one of the reasons for the high mortality during the Paris heatwave of 2003 [3,4]. Short distances between upper and lower quantiles are of key importance in establishing criteria regarding the existence of heatwave events, conditioned to steadily higher minimum temperatures [5]. The analyses of previous heatwaves are of particular

interest for those looking to create future weather time series when looking at ventilation and overheating issues in the built environment [6,7,8], drought periods in agriculture [9], the management of water resources [10,11] and energy systems [12,13].

Various methods for the multiresolution analysis of time series are discussed in the literature. The discrete wavelet transform (DWT) is a natural way to reach such a decomposition and is useful in indexing time series [14]. However, different alternatives have been proposed for those looking for improved efficacy. For example, those based on fuzzy inductive reasoning [15] decompose the time series into a trend series and another complementary series describing the deviation from this trend. Alternatives based on symbolic aggregate approximation (SAX) have also been successfully developed [16,17]. The main advantage of SAX based methods is that they provide a suitable avenue for pattern recognition models associated with time series [18].

There are a few works related to multiresolution time series with respect to climate or weather time series, including solar irradiance and climate reconstructions [19] and a hybrid approach to removing noise from a climate time series [20]. However, the work introduced in this paper is closer to the hierarchy-of-clusters approach proposed for use in multiresolution image analysis [21]; being a variation of 1d-SAX [22] but extending the more common SAX methodology to deal with time series quantiles. MRQ starts with the typical piecewise aggregate approximation (PAA) to divide the series into segments of equal length [23]. After that, the quantiles for each segment of data are saved for further analysis instead of only the average values, and SAX run on these quantities. Finally the differences between upper and lower quantiles at each resolution level is computed by a lower-bounding distance measure [24]. Possible heatwave events could then be detected based on the persistence of minimum distances at various time-based resolution levels.

Indexing and mining time series for multiresolution purposes

Indexing time series is traditionally used as a way to efficiently store a large temporal database [25]. However, its use has been expanded to allow the extraction of patterns in time series. While common indexing and mining models are based on measures formed from the average, the method proposed here is quantile based: this is to allow for better treatment of the higher temperatures.

Multiresolution based on average values

Wavelets are mathematical functions that represent data or other functions in terms of the averages and differences to a prototype function [26]. An interesting wavelet property w.r.t multiresolution approaches is that the first coefficients forming a wavelet expression contain an overall approximation of the time series under study, while additional coefficients take into account data characteristics in greater detail [27]. This endows the wavelets with the properties required for a suitable methodology for investigating time series at various resolutions. Haar's Discrete wavelet transform [14] is a widely used case in which the prototype function follows an orthonormal system, at discrete times, for the space of square-integrable functions on the unit interval [0, 1]. Haar's DWT is the simplest case of wavelets but can be developed to efficiently catch up on any change on the time series average. Furthermore, different DWTs can be used as bases and combined in order to produce a better approximation of the original time series, with more bases giving a higher resolution (see Figure 1).
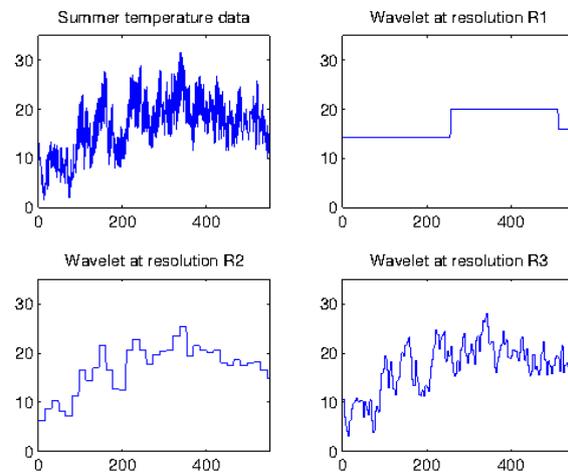

Fig. 1. Example of a combination of various wavelet resolution levels for a time series.

Haar's DWT's are ultimately made up by step functions. Approaches based on piecewise aggregate approximation (PAA), which divides the time series dataset into equally spaced segments, provide similar results to Haar's DWT in terms of time series decomposition. Consequently, DWT results and PAA based outcomes are equivalents to computing distances between time series [23]. The efficient results found by using PAA based methods, together with their capabil-

ity to be straightforwardly part of further data mining processes, represent a key advantage for the work presented here.

Multiresolution based on quantiles

The symbolic aggregate approximation of time series (SAX) [24] represents the time series as a sequence of symbols. It was primarily developed to reduce the dimensionality of a numerical series into a short chain of characters. However, it has been found useful for data mining tasks such as: indexing [28], clustering [29,30], and classification [31]. SAX is based on three steps:

1. Divide a time series into segments of length L
2. Compute the average of the time series on each segment
3. Represent the average values as a symbol from an alphabet of size N

The time series division is based on a previous PAA phase. SAX is based on the assumption that time series values follow a Gaussian distribution for each of the segments into which PAA divided the series. The conversion of the average values into a symbol makes use of ($N1$) breakpoints that divide the area under the Gaussian distribution into N equi-probable areas and then the average value per segment is quantized according to the areas of this distribution [24]. As a result a "word" is composed containing as many letters as segments in the PAA.

This alphabetic approach is then useful in further analyses using methods such as hashing [32], variations of Markov models [33,34], and suffix tree approaches [35]. In addition, it automatically has associated a sliding windows approach in which every timeframe is encoded by a letter. Figure 2 represents the output of a single use of SAX process for a temperatures time series from London (April to September, 1989).

The SAX variation known as 1d-SAX [22] extends the usual alphabetic symbols to a system able to contain information about the average and the trend of the series within a segment. A natural extension of 1d-SAX takes median values on each interval, and provides similar results to linear regression. The new MRQ method proposed generalizes this approach by computing upper and lower quantiles at every segment of the PAA partition. The PAA partition becomes a bi-level PAA-quantile based partition (PAA-Q), and we have termed the modified SAX approach SAX-
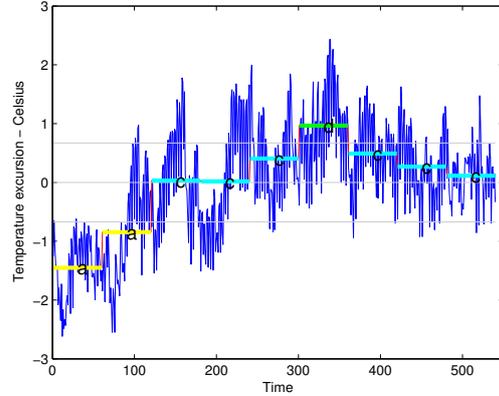

Fig. 2. Example of the SAX conversion process for a time series with length 549, w = 9 and resolution 4 (a, b, c, d). Temperature variations from the long term baseline.

Quantile, or SAX-Q for short. PAA-Q is defined in a similar way to the single PAA for a time series of length n. $Q_u = q_{u1}, \cdots, q_{uw}$ and $Q_l = q_{l1}, \cdots, q_{lw}$; corresponding to upper and lower quantiles respectively. Then, the time series dimension is reduced from n to $2 \times w$ dimensions, and the data is divided into w equal sized frames. The vectors, Q_u and Q_l become a suitable data representation taking into account the extremes of the time series instead of the average values.

This multiresolution approach facilitates working at the different coarseness of PAA-Q used in SAX-Q to allow for posterior analysis [36]. MRQ focuses specifically on quantile information for discovering patterns, in our case of steadily high summer temperatures at upper and lower quantiles i.e. heatwaves. Having enough variety in the coarseness levels of the PAA-Q partition, and computing upper and lower quantiles at each resolution level, we believe allows heatwave events to be identified successfully.

First, with larger PAA-Q segments, the series is filtered to allow the rapid identification of patterns associated with high temperatures even at lower quantiles. This is easily achieved by splitting the series using a binary SAX codification of values lower and higher than a certain threshold. For the set of segments of interest (those encoded as high temperatures) we approach a finer PAA-Q resolution. The overall MRQ process is:

1. Set PAA to coarsest scale with segment: k_1 days in length

   (a) Compute upper and lower quantiles: α_u and α_l levels, respectively

(b) Approach SAX-Q coding as '1' those segments encoded with higher alphabet values for both lower and upper quantiles; '0' otherwise

2. Target the set of PAA-Q encoded as '1'

(a) Define 'runs' as sub-series segment of consecutive PAA-Q stretches coded as '1'

(b) Set PAA-Q for the runs to a finer scale: k_2 days length (with $k_2 < k_1$)

(c) Go to 1. (a)

Patterns extraction from summer temperatures

Two meteorological databases in Brazil and UK have been analysed. Both of them contain 50 years of data taken at 8 hour intervals during the period 1961 to 2010.

Brazil summer database

The Brazil weather database was collected at Curado weather station in Recife (Pernambuco State, Brazil). For each one of the 50 years analysed, the months of October to March (of the following year) were selected to represent the summer period. The average temperature for this period is 27.59^{o}C with an associated standard error of 1.71. The choice of the number of segments to approach the first PAA basis partition takes into account a suitable length of every segment and the error related to the approximation (see Table 1). Thus, for a segment length of size 30 (10 days) the average root mean square error over (RMSE) the 50 years is only $1.62/1.58 = 1.02$ times higher than the average RMSE provided by working with a length of size 9 (3 days) and 1.06 times higher than working with segments of size 3 (1 day). This is better than working with a coarser partition based on segments of size 180 (60 days) which have an average RMSE 1.09 and 1.15 higher than partitions of 3 and 1 days respectively.

Table 1
Average RMSE in the PAA approximation of the Recife 1961 2010 8 hourly temperature time series at different segment sizes.

Segment length	540	180	60	30	9	3
Avg. RMSE	1.72	1.73	1.68	1.62	1.58	1.50

Figure 3 shows the performance of SAX-Q for a given PAA-Q partition and one selected summer. The summer of 1989-90 is chosen for consistency with the choice of using the summer of 1989 for the UK anal-
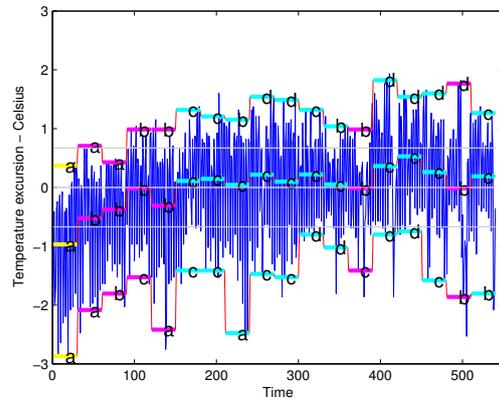

Fig. 3. SAX-Q conversion process for Recife's temperature time series with length 540, w = 18, resolution 4 (a, b, c, d), and quantiles 0.05 and 0.95. Temperature variations in summer 1989-90 are given w.r.t. the baseline period of 1961-2010.

ysis. Upper and lower quantiles, together with the median (quantile 50 %) are shown in Figure 3.

After obtaining the output of the MRQ process for Recife's summer temperatures, it is possible to identify the warmest temperatures with the beginning and the end of the summer period. Figure 4 shows the histogram of the normalised probabilities regarding the frequencies of 2-level SAX-Q in the observed time series. This shows that at the beginning of the summer there is often a brief period of high temperatures, and from the end of December and until February also high temperatures.
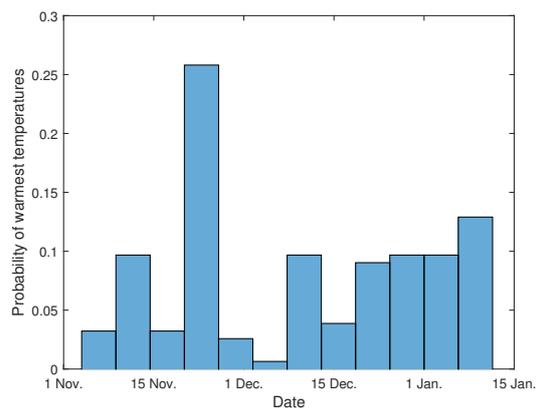

Fig. 4. Probability of warmest temperatures during Recife summers (1961-2010).

Figure 5a shows how the frequency of periods of high temperature increases in the latter years of the

study period. This Figure uses a colour scale in which light red represents no heatwaves that year, red is 1 period of high temperatures, and dark red is 2 or more periods of high temperatures. It can be seen how the ratio of years with heatwaves to years without is 10/13 in the first half of data. However, post year 1985, the situation is completely different, with two or more heatwaves in some years and the ratio of years with heatwaves to years without increasing to 14/7.

Figure 5b complements the information given in 5a by showing how the mean temperature during each heatwave changes over the 50 years of the study. It is clear that there is an increasing trend in the mean, so not only are the number of heatwaves increasing, their magnitude are also increasing, with the mean heatwave temperature being 1.5^oC higher in 2005 than at the start of the data. Applying linear regression suggests an increase in heatwave magnitude of 0.4^oC per decade; this should be compared to the overall temperature trend during the summer period which is less than 0.2^oC per decade.

UK summer database

The UK weather data was collected at the Heathrow weather station (London, UK) [37]. For each one of the 50 years analysed, the months of April to September were selected to represent the summer period. The average temperature for this period is 15.78^oC with an associated standard error of 4.71. The temperature is clearly below Recife's values. However, the proposed MRQ process automatically adapts itself to locally set conditions for screening heatwaves. The choice of the number of segments to approach the first PAA basis partition takes into account a suitable length for every segment and the error related to the approximation (see Table 2). Thus, for a segment length of size 61 (20 days) the average RMSE over the 50 years is $3.57/2.96 = 1.20$ times higher than the average RMSE provided by working with a length of size 9 (3 days) and 1.33 times higher than working with segments of size 3 (1 day). This is better than working with a coarser partition based of segments of size 183 (60 days) which have an average RMSE 1.36 and 1.52 higher than partitions of 3 and 1 days respectively.
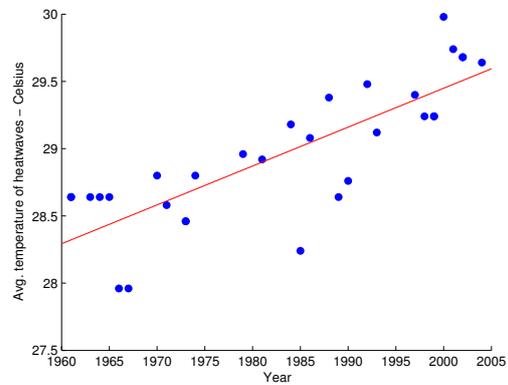
Table 2

Average RMSE in PAA approximation of time series at different segment sizes. London 1961 - 2010.

Segment length	549	183	61	9	3
Avg. RMSE	4.89	4.05	3.57	2.96	2.68
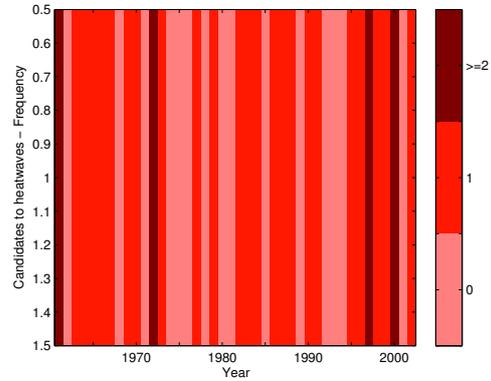
(a) Frequency of warmer temperatures

(b) Trend of average temperatures associated with heatwave events

Fig. 5. Change in frequency and intensity of heatwave candidates. Recife 1961 - 2005.

Figure 6 shows the performance of SAX-Q for a given PAA-Q partition and the summer of 1989. The year 1989 was chosen as it represents a moderately warm year and is used as a standard warm summer by the Chartered Institution of Building Services Engineers (CIBSE, UK) [38].

After obtaining the output of the MRQ process for London it is found that the warmest temperatures typically occur during the last week of July and the first days of August. Figure 7 shows the histogram of the normalised probabilities of 2-level SAX-Q in the observed time series.

Figure 8a shows how the heatwave frequency changes over the study period. In the first 25 years the ratio of years with heatwaves to years without is 18/7. However, post year 25 (1986), the ratio of years with heatwaves to years without increases to 21/4. This increase
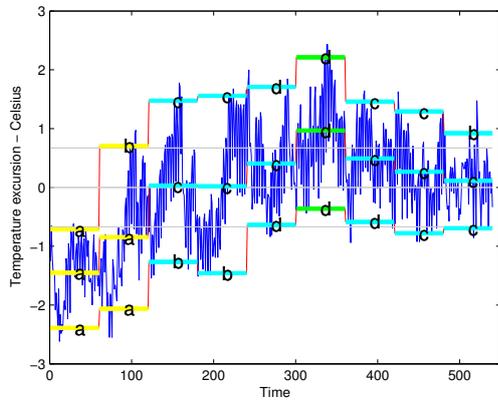
Fig. 6. SAX-Q conversion process for London's temperatures time series with length 549, w = 9, resolution 4 (a, b, c, d), and quantiles 0.05 and 0.95. Temperature variations in summer 1989 w.r.t the baseline period 1961-2010.
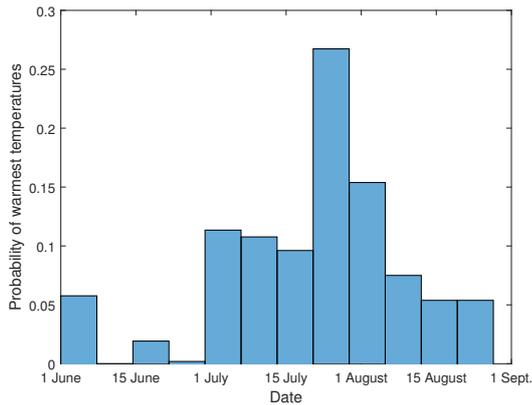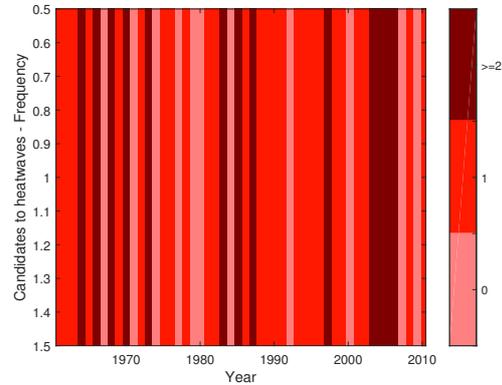

Fig. 7. Probability of warmest temperatures during London summers (1961-2010).

(a) Frequency of warmer temperatures

(b) Trend of average temperatures associated with heatwave events

Fig. 8. Change in frequency and intensity of heatwave candidates. London 1961 - 2010.
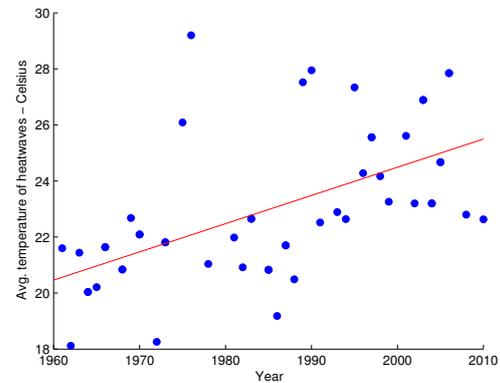
in heatwave frequency is even higher than that observed for the previous case-study of Recife.

Figure 8b shows how the mean temperature associated with each heatwave has increased with the mean heatwave temperature being about 4^oC higher towards the end of the series that near the beginning. Applying linear regression suggests an increase in heatwave magnitude of near 1^oC per decade, this should be compared to the overall temperature trend during the summer period which is less than 0.15^oC per decade.

It is interesting to note the difference in the values of the candidate heatwave temperatures identified by the process for Recife and London. While the candidate temperatures for heatwaves in Recife have a minimum value over 27^oC, in London 18^oC at night is enough to indicate the temperature is potentially part of a heatwave a 9^oC difference between the two cities. However, the average temperature of the selected dataset in Recife is around 29^oC, compared with 24^oC in London, only a 5^oC difference.

The proposed method is blind to the temperatures themselves and suggests a possible heatwave based only on the variability and the size of the variation, thus it is useful in that it is universal. The corollary however is that if given an unsuitable time series it might identify heatwaves where the maximum temperature reached is still below any common definition of a heatwave. In a practicable application there would hence need to be the use of an heuristic for rejecting unsuitable time series or locations, or one applied after identifying the heatwave; for example that the peak temperature must be at least 28^oC. However, this constant would depend greatly on the location: 28^oC is a high

temperature in London for example, but not in Saudi Arabia.

Conclusions

The paper presents a new method based on a fundamental modification of the SAX approach for time series, called SAX-Q, which focuses on the symbolic approximation of the quantiles instead of using the mean. Its iterative use leads to an automatic detection process for periods of interest w.r.t time series extremes.

Multi-resolution analysis appears to be an ideal tool to uncover relationships at different temporal scales in weather data that are normally hidden by background noise. A posterior analysis of a multi-resolution SAX-Q has been carried out the summer temperature record in Brazil and the UK and heatwaves identified in these two, very different, climates; as the methodology automatically adapts itself to local conditions. Fundamental questions about when, how often and the intensity of high temperatures have been answered for both locations. Most importantly, in both locations heatwave intensity and frequency are increasing at a substantial rate. In addition, it was found that the rate of increase in intensity of the heatwaves is far outstripping the rate of increase in mean summer temperature: by a factor of 2 in Recife and a factor of 6 in London.

An interesting extension to this work would come from the relaxation of the PAA-Q partition proposed, by instead using a dynamical basis for segments or change points for the initial SAX-Q partition of the time series. This might provide better insights about the length of the hottest periods lying in the data.

Acknowledgements

References

[1] Qiang Wang, Vasileios Megalooikonomou, and Christos Faloutsos. Time series analysis with multiple resolutions. *Information Systems*, 35(1):56–74, 2010.

[2] Jeffrey D Scargle. Wavelet and other multi-resolution methods for time series analysis. In *Statistical Challenges in Modern Astronomy II*, pages 333–347. Springer, 1997.

[3] Marine Laaidi, Abdelkrim Zeghnoun, Bénédicte Dousset, Philippe Bretin, Stéphanie Vandentorren, Emmanuel Giraudet, and Pascal Beaudeau. The impact of heat islands on mortality in Paris during the August 2003 heat wave. *Environmental Health Perspectives*, 120(2):254, 2012.

[4] Laura Schaeffer, Perrine de Crouy-Chanel, Vérène Wagner, Julien Desplat, and Mathilde Pascal. How to estimate exposure when studying the temperature-mortality relationship? a case study of the paris area. *International Journal of Biometeorology*, 60(1):73–83, 2016.

[5] Manuel Herrera, Matthew Eames, Alfonso Ramallo-Gonzalez, Chunde Liu, and David Coley. Quantile regression ensemble for summer temperatures time series and its impact on built environment studies. *Proceedings of International Environmental Modelling and Software Society (IEMSS)*, 2016.

[6] David Coley, Tristan Kershaw, and Matt Eames. A comparison of structural and behavioural adaptations to future proofing buildings against higher temperatures. *Building and Environment*, 55:159–166, 2012.

[7] Tristan Kershaw, M Eames, and David Coley. Comparison of multi-year and reference year building simulations. *Building Services Engineering Research and Technology*, 31(4):357–369, 2010.

[8] M Eames, T Kershaw, and David Coley. A comparison of future weather created from morphed observed weather and created by a weather generator. *Building and Environment*, 56:252–264, 2012.

[9] RL Wilby, H Orr, G Watts, RW Battarbee, PM Berry, R Chadd, SJ Dugdale, MJ Dunbar, JA Elliott, C Extence, et al. Evidence needed to manage freshwater ecosystems in a changing climate: turning adaptation principles into practice. *Science of the Total Environment*, 408(19):4150–4164, 2010.

[10] Birgitte Christierson, Jean-Philippe Vidal, and Steven D Wade. Using UKCP09 probabilistic climate information for UK water resource planning. *Journal of Hydrology*, 424:48–67, 2012.

[11] Bruno M Brentan, Edevar Luvizotto Jr, Manuel Herrera, Joaquín Izquierdo, and Rafael Pérez-García. Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics*, 309(1):532–541, 2017.

[12] Jane O Ebinger. *Climate impacts on energy systems: key issues for energy sector adaptation*. World Bank Publications, 2011.

[13] Ronaldo R. B. de Aquino, Hugo T. V. Gouveia, Milde M. S. Lira, Aida A. Ferreira, Otoni Nobrega Neto, and Manoel A Carvalho. *Engineering Applications of Neural Networks: EANN 2012, London, UK*, chapter Models Based on Neural Networks and Neuro-Fuzzy Systems for Wind Power Prediction Using Wavelet Transform as Data Preprocessing Method, pages 272–281. Springer Berlin Heidelberg, 2012.

[14] Kin-Pong Chan and Ada Wai-Chee Fu. Efficient time series matching by wavelets. In *Data Engineering, 1999. Proceedings., 15th International Conference on*, pages 126–133. IEEE, 1999.

[15] François E Cellier and Àngela Nebot. Multi-resolution time-series prediction using fuzzy inductive reasoning. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 1621–1624. IEEE, 2004.

[16] Nuno Castro and Paulo J Azevedo. Multiresolution motif discovery in time series. In *SDM*, pages 665–676. SIAM, 2010.

[17] Vasileios Megalooikonomou, Qiang Wang, Guo Li, and Christos Faloutsos. A multiresolution symbolic representation of time series. In *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, pages 668–679. IEEE, 2005.

[18] Anand Rajaraman, Jeffrey D Ullman, Jeffrey David Ullman, and Jeffrey David Ullman. *Mining of massive datasets*, volume 1. Cambridge University Press Cambridge, 2012.

[19] Hee-Seok Oh, Caspar M Ammann, Philippe Naveau, Doug Nychka, and Bette L Otto-Bliesner. Multi-resolution time series analysis applied to solar irradiance and climate reconstructions. *Journal of Atmospheric and Solar-Terrestrial Physics*, 65(2):191–201, 2003.

[20] Taesam Lee and TBMJ Ouarda. An emd and pca hybrid approach for separating noise from signal, and signal in climate change detection. *International Journal of Climatology*, 32(4):624–634, 2012.

[21] Jessica Lin, Michail Vlachos, Eamonn Keogh, and Dimitrios Gunopulos. Multiresolution clustering of time series and application to images. In *Multimedia Data Mining and Knowledge Discovery*, pages 58–79. Springer, 2007.

[22] Simon Malinowski, Thomas Guyet, René Quiniou, and Romain Tavenard. 1d-SAX: A novel symbolic representation for time series. In *Advances in Intelligent Data Analysis XII*, pages 273–284. Springer, 2013.

[23] Eamonn Keogh, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, 3(3):263–286, 2001.

[24] Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, 15(2):107–144, 2007.

[25] Eamonn Keogh. Indexing and mining time series data. In *Encyclopedia of GIS*, pages 493–497. Springer, 2008.

[26] Amid Bakhtazad, Ahmet Palazoglu, and José A Romagnoli. Process data de-noising using wavelet transform. *Intelligent Data Analysis*, 3(4):267–285, 1999.

[27] Piotr Wojdyłło. Wavelets and mallat's multiresolution analysis. *Fundamenta Informaticae*, 34(4):469–474, 1998.

[28] Durga Toshniwal. Feature extraction from time series data. *Journal of Computational Methods in Sciences and Engineering*, 9(1, 2S1):99–110, 2009.

[29] Saeed Aghabozorgi and Teh Ying Wah. Clustering of large time series datasets. *Intelligent Data Analysis*, 18(5):793–817, 2014.

[30] Francisco Martínez-Álvarez. Clustering preprocessing to improve time series forecasting. *AI Communications*, 24(1):97–98, 2011.

[31] Jidong Yuan, Zhihai Wang, Meng Han, and Yange Sun. A lazy associative classifier for time series. *Intelligent Data Analysis*, 19(5):983–1002, 2015.

[32] Xiaoyue Wang, Abdullah Mueen, Hui Ding, Goce Trajcevski, Peter Scheuermann, and Eamonn Keogh. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2):275–309, 2013.

[33] Jessica Lin and Yuan Li. Finding structural similarity in time series data using bag-of-patterns representation. In *Scientific and Statistical Database Management*, pages 461–477. Springer, 2009.

[34] Carmelo Cassisi, Michele Prestifilippo, Andrea Cannata, Placido Montalto, Domenico Patanè, and Eugenio Privitera. Probabilistic reasoning over seismic time series: Volcano monitoring by hidden markov models at mt. etna. *Pure and Applied Geophysics*, pages 1–22, 2016.

[35] Faraz Rasheed, Mohammed Alshalalfa, and Reda Alhajj. Efficient periodicity mining in time series databases using suffix trees. *Knowledge and Data Engineering, IEEE Transactions on*, 23(1):79–94, 2011.

[36] Abdullah Mueen and Nikan Chavoshi. Enumeration of time series motifs of all lengths. *Knowledge and Information Systems*, 45(1):105–132, 2015.

[37] UK Meteorological Office. Met Office Integrated Data Archive System (Midas) land and marine surface stations data (1853-current). *NCAS British Atmospheric Data Centre*, (accessed 01/10/14), 2012.

[38] CIBSE. Design Summer Years for London – CIBSE TM49, 2014.