
White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/88481/

Version: Accepted Version

Article:
Aykroyd, RG, Barber, S and Miller, LR (2016) Classification of multiple time signals using localized frequency characteristics applied to industrial process monitoring. Computational Statistics and Data Analysis, 94. pp. 351-362. ISSN: 0167-9473

https://doi.org/10.1016/j.csda.2015.07.009

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Classification of multiple time signals using localized frequency characteristics applied to industrial process monitoring

Robert G. Aykroyd[a], Stuart Barber[a] and Luke R. Miller[b]

[a] *University of Leeds, Leeds, UK.*
[b] *University of Oxford, Oxford, UK.*

Abstract

A general framework for regression modeling using localized frequency characteristics of explanatory variables is proposed. This novel framework can be used in any application where the aim is to model an evolving process sequentially based on multiple time series data. Furthermore, this framework allows time series to be transformed and combined to simultaneously boost important characteristics and reduce noise. A wavelet transform is used to isolate key frequency structure and perform data reduction. The method is highly adaptive, since wavelets are effective at extracting localized information from noisy data. This adaptivity allows rapid identification of changes in the evolving process. Finally, a regression model uses functions of the wavelet coefficients to classify the evolving process into one of a set of states which can then be used for automatic monitoring of the system. As motivation and illustration, industrial process monitoring using electrical tomography measurements is considered. This technique provides useful data without intruding into the industrial process. Statistics derived from the wavelet transform of the tomographic data can be enormously helpful in monitoring and controlling the process. The predictive power of the proposed approach is explored using real and simulated tomographic data. In both cases, the resulting models successfully classify different flow regimes and hence provide the basis for reliable online monitoring and control of

Address for correspondence: Robert G. Aykroyd, Department of Statistics, University of Leeds, LS2 9JT, UK. E-mail: R.G.Aykroyd@leeds.ac.uk. Software to perform the data analysis example is available from `www.maths.leeds.ac.uk/~stuart/research`.

industrial processes.

Keywords: Electrical tomography, Logistic regression, Process control, Remote sensing, Wavelets

---

1. Introduction

High-frequency data are routinely collected in a wide range of monitoring and forecasting applications such as financial trading, meteorology, environmental science, industrial process engineering and internet marketing. Data sets consist of multiple time series which accumulate rapidly and must be analyzed in real-time. This may mean that wide-ranging analysis is impractical and that the focus must be on answering well-defined questions. The aim is to summarize the incoming data-stream without losing essential information.

An exemplary application is the monitoring of industrial processes, where measurements taken while the process is evolving must be converted into parameters which can be used to monitor and control the process. Electrical tomography is a widely used technique which aims to investigate the interior of a region using voltages taken outside the region. This provides indirect information about the internal conductivity distribution, which reflects the state of the process. Such techniques are widely used in geophysical, industrial and medical investigations. The predominant method of analysis estimates the conductivity at points forming a fine grid — see for example Aykroyd (2015), Lionheart (2004) and Watzenig and Fox (2009). This leads to an over-parameterized regression type problem, known as an ill-posed inverse problem. Stable solution then requires substantial regularization, which can mask features of interest. Although image reconstruction is useful for process visualization, for automatic control an image is unnecessary (Stitt and James, 2003; Hoyle, 2004). Such reconstruction may be time-consuming and the image will still require post-processing to obtain control parameters. Hence direct control parameter estimation, rather than process visualization, is the more appropriate output of a data analysis in many real situations.

Clearly, there is a need in many other applications, as well as industrial monitoring, for methods which are simple, fast and can operate largely unsupervised. Wavelets are an ideal tool for our purpose since their multiscale nature enables the efficient description of both transient and long-term signals; this will be illustrated in §3. We propose the use of wavelets in relating time-series measurements to the response variable in §4. In particular, func-

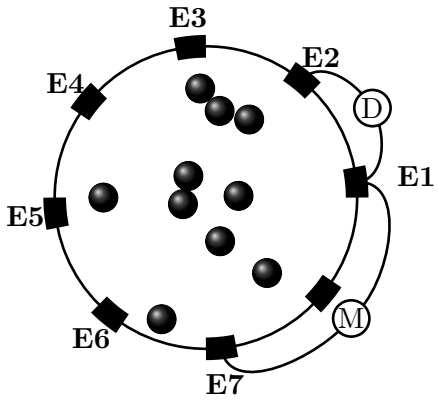

Figure 1: Diagram of data collection protocol showing drive and measurement circuits connecting electrode on the pipe boundary with bubbles passing through the pipe.

tions of wavelet coefficients, which emphasize key frequency information, are proposed and used as explanatory variables in a predictive regression model. Efficacy of the proposed method is demonstrated on simulated electrical tomography data in §5 and real data in §6. We initially illustrate our methods by application to simulated data sets which we describe in the next section.

2. Description of simulated data

To motivate the data simulation, consider the flow of a gas upwards through a liquid in a pipe. The gas fraction and bubble size are determined by the inlet size and the input pressure. To control process efficiency it is important to monitor the flow regime, and to adjust the input parameters when necessary. In our simulation, bubble flow (many small bubbles) and churn flow (few large bubbles) will be considered. The spatial distribution of the bubbles then defines the conductivity distribution which determines the measurements. To create conductivity distributions, bubbles enter the plane of the electrodes at random with frequency and diameter determined by the flow regime. The diameter of the bubbles also determines the length of time the bubble remains in the plane of the electrodes.

The data collection scheme is motivated by the widely used reference protocol for an eight-electrode electrical tomography system; for details, see West et al. (2005). Figure 1 shows a cross section through the pipe with the eight electrodes on the boundary. To start the process a drive circuit passes a current between the fixed reference electrode (E1) and a second electrode

(E2). For each current pattern an induced potential field is created within the pipe which depends on the electrodes in the drive circuit, and upon the conductivity distribution within the pipe. Then a measurement circuit is created connecting the reference electrode and each of the other electrodes in turn and the voltage is recorded. In the diagram electrode E7 is part of the measurement circuit. With the reference electrode fixed, seven other electrodes can be part of each of the drive and measurement circuits, leading to a total of 49 measurements. Further, the process is allowed to evolve for n time points.

For each spatial conductivity distribution, c, the value of the potential field, ϕ, within the pipe is found by solving a system of Maxwell's equations, $\nabla \cdot (c \nabla \phi) = 0$, with certain boundary conditions on the electrodes and on the insulating pipe between electrodes, see West et al. (2005). This is a system of second-order partial differential equations the solution of which requires substantial numerical effort. Here the finite element method is implemented using the EIDORS library (Polydorides and Lionheart, 2002) in MATLAB. Once the potential field has been calculated, the voltages are then given by the difference in the potential at the relevant locations. Once noise-free voltages are obtained, uncorrelated Gaussian noise with mean zero and variance τ^2 is added to yield the simulated data set.

Figure 2(a) shows a typical trace from a single sensor pair, with no noise in (a) and increasing levels of noise corruption in (b)–(d). In each, the first half of the trace corresponds to bubble flow and the second half to churn flow. In the noise-free and low-noise cases, (a) and (b), the change of flow regime can easily be seen. The signals from each of the flow regimes exhibit extremely different behaviour, which can be explained in terms of the frequency of the change in the measurements. The bubble flow, with its many small pockets of gas in liquid, produces more rapidly varying measurements than the churn flow. Even in (c), when the noise has $\tau = 0.05$, it is just possible to distinguish between the two flow regimes. By $\tau = 0.1$ however it is almost impossible to make any reliable judgement by eye.

The left and right columns of Figure 2 show standard regularized least-squares reconstructions for the same noise levels as the central column using total-variation regularization during bubble flow and churn flow respectively. For a review of similar approaches see Lionheart (2004). For noise-free data the reconstructions are clearly distinguishable. For the churn flow a dark (gaseous) region in the top-left is clear. For the bubble flow there are no well-defined bubbles, but it is clear that gaseous areas occur across the pipe. For

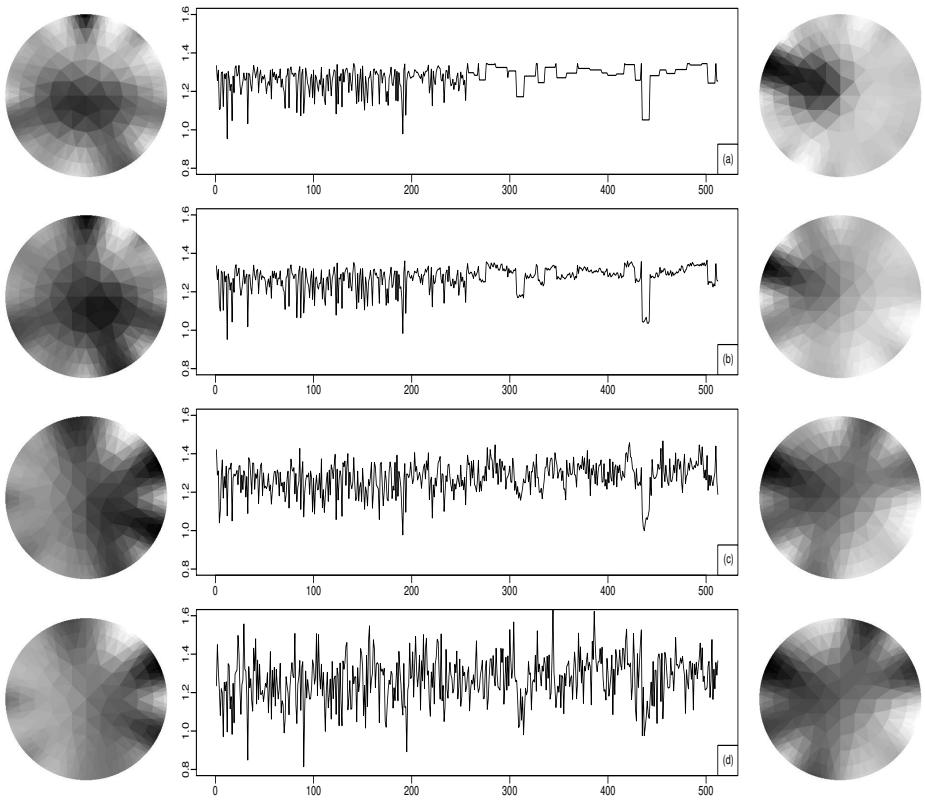

Figure 2: Example simulated signals with increasing levels of noise and corresponding image reconstructions.

low noise the picture is similar, though there is more random variation across the pipe. For the moderate and high noise cases it is virtually impossible to see any clear difference and any image-based classification will be unreliable.

3. Wavelet transforms of tomographic data

We propose using wavelets, a type of basis function, to analyze and classify our multiple time-series data. A brief introduction to wavelets and their relevant properties is given here; for more details see, for example, Nason (2008). Calculations were performed in the language R (R Development Core Team, 2014) using the WaveThresh package (Nason, 2010).

For data such as tomography sensor traces, where the flow state changes over time and different flow states have different frequency characteristics, a

standard spectral analysis via Fourier expansion is clearly unsuitable. To overcome this problem, wavelets can be used which are able to simultaneously represent a signal in the time and frequency domains. However, wavelets only represent a data set of length n at $\lfloor \log_2(n) \rfloor$ resolution levels or scales. Wavelets at a fine scale (high frequency) resolution level are highly localized, representing brief transient effects, while those at a coarser scale measure lower frequency activity. Each resolution level represents activity at approximately twice the frequency of the previous level.

We use the non-decimated discrete wavelet transform (Nason and Silverman, 1995; Coifman and Donoho, 1995). The NDWT has complete time localization at each scale, resulting in an over-complete basis. We denote a wavelet function by $\psi(t)$, and create scaled and shifted copies via

$$\psi_{jk}(t) = 2^{j/2}\psi(2^j(t-k)),$$

where $j = 1, 2, \ldots, J-1$ and $k = 0, 1, \ldots, n-1$ represent the resolution level and location respectively. The NDWT maps a discrete signal vector $y = (y_i : i = 1, 2, \ldots, n)$ to a collection of wavelet coefficients $\{d_{jk}\}$ at levels $j = 1, 2, \ldots, J-1$ and locations $k = 0, 1, \ldots, n-1$ defined by

$$d_{jk} =< y, \psi_{jk} > .$$

Figure 3 shows wavelet transforms of selected noise-free tomographic signals (column 1). The recorded signals are examples of bubble flow (a,c) and churn flow (b,d). The top two rows show data from a measurement circuit comprising adjacent electrodes; the bottom two rows correspond to measurements between electrodes at opposite sides of the pipe. The second column shows the non-decimated wavelet transform for each recorded signal. Here, the numbers on the vertical axes indicate the resolution levels from zero (coarsest) to 7 (finest). Columns 3 and 4 show the "activity measures" which we shall define later in §4. Although the mean values in (a) and (b) are similar, as are those in (c) and (d), the frequency patterns are very different, and this is reflected in the corresponding wavelet coefficients and activity measures. Finer resolution levels (5–7) show a lot of activity for bubble flow and considerably less for churn flow. At the coarsest levels, there is slightly less activity for bubble flow and more for churn flow. This indicates a difference in the frequencies at which the different signals are active, which is detectable using the wavelet transform.

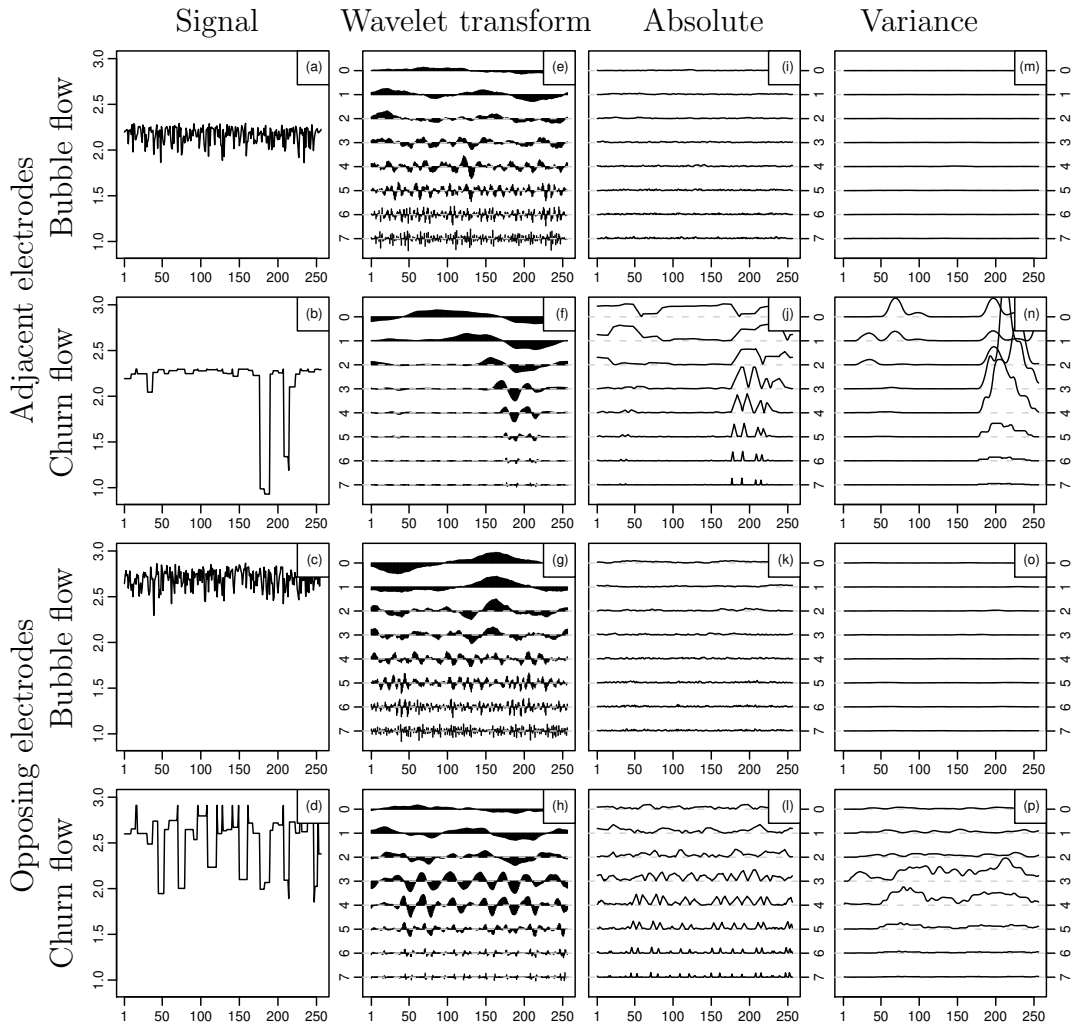

Figure 3: Example data, showing four of the 49 time series, two each from bubble (a, c) and churn (b, d).

4. Methodology

Suppose we have a response and p explanatory variables connected by some unknown relationship and training data which can be used to investigate this relationship. The fitted model is then used to predict the unknown response in test data where there is only knowledge of the explanatory variables.

Here two types of feature enhancing and data reducing transformations will be proposed. Let d_{ijt} be the non-decimated wavelet coefficient of explanatory variable i at level j and at time t. In our application it is important to have a measure which is invariant to relabelling of the explanatory variables. This can be achieved by combining the explanatory variables at each time point. It is important, however, to retain the distinct information at each resolution level hence we consider two activity measures which aggregate over explanatory variables using functions γ_1 and γ_2. The first retains separate values for distinct resolutions and time points:

$$Z_{\cdot jt} = \gamma_2 \left(\sum_{i=1}^{p} \gamma_1(d_{ijt}) \right). \tag{1}$$

As an example, and defining the case to be used later, let $\gamma_1(\cdot) = |\cdot|$ and γ_2 be the identity. This defines an absolute activity measure where the magnitudes of the wavelet coefficients are summed over explanatory variables:

$$Z^{(\mathrm{abs})}_{\cdot jt} = \sum_{i=1}^{p} |d_{ijt}|. \tag{2}$$

In the second measure, values are combined within a backwards facing time window of length w (truncated at the start of the data trace):

$$Z_{\cdot jt} = \gamma_2 \left(\sum_{i=1}^{p} \gamma_1 \left(d_{ij(\max\{1,t-w+1\})}, \ldots, d_{ijt} \right) \right). \tag{3}$$

As an example, and again defining the case to be used later, let $\gamma_1(\cdot) = \mathrm{var}(\cdot)$ and γ_2 be the identity. This defines the variance activity measure where variances of the wavelet coefficients over a time window are calculated and summed over the explanatory variables:

$$Z^{(\mathrm{var})}_{\cdot jt} = \sum_{i=1}^{p} \mathrm{var} \left(d_{ij(\max\{1,t-w+1\})}, \ldots, d_{ijt} \right). \tag{4}$$

Figure 3 shows typical signals during bubble and churn flow. Also shown are the corresponding wavelet decompositions and derived activity measures in columns 3 and 4. When comparing corresponding bubble and churn signals, there are clear differences, although it is important not to be drawn to transient patterns.

Consider classifying flow into one of two states using a logistic regression model to predict the response y_t from the activity measures $z_{.jt}$. A step-wise variable selection procedure is used to identify important resolution level variables. Two optimization criteria will be considered: (i) minimizing AIC, and (ii) maximizing the correct classification rate. In both cases the model fitting procedure uses a training data set, but for final model assessment the correct classification rate is calculated when the fitted model is used to predict on an independent test data set.

Let V represent the set of included variables and V' the remaining variables, let $AIC(V)$ be the value of the AIC for the model containing variables V, and let V_{-v} and V_{+v} represent the set of variables with v removed or added. The variable selection procedures uses the following algorithm.

---

1. Initialize with current model V empty and all variables in V'.
2. Forward: Consider adding a variable to current model V:
   - Let $v' = \arg\min_{u' \in V'} AIC(V_{+u'})$, and move v' to V if $AIC(V_{+v'}) < AIC(V)$
3. Backward: Consider removing a variable from current model V:
   - Let $v = \arg\min_{u \in V} AIC(V_{-u})$, and move v to V' if $AIC(V_{-v}) < AIC(V)$
4. Swap: Consider swapping variables between V and V':
   - Let $(v, v') = \text{args} \min_{u \in V, u' \in V'} AIC(V_{-u+u'})$, and move v to V' and v' to V if $AIC(V_{-v+v'}) < AIC(V)$
5. Repeat steps 2, 3 and 4 until there is no change in V and V'.
6. Output V and calculate the correct classification rate using the test data set.

A second approach maximizes the correct classification rate, but this time the fitting algorithm proceeds until all variables are included and returns the best model for each possible number of variables. The correct classification rate is then calculated for each of these models using the test data set.

5. Simulation results

We now discuss the results of applying our proposed method to simulated tomographic data. Our interest is not only in the method's performance on the specific tomography application, but also in how one would select the best approach for other data sets. Thus, we discuss a range of different variations of our method and what criteria one might use to choose a preferred approach. The Haar wavelet was used throughout; we have found other wavelets to give similar or worse results.

Our proposed methods were trained on 100 sets of simulated data, in each case testing the resulting model's performance on a separate independent test data set. Two training regimes were considered. One ("pure") calculates activity measures separately for 256 observations of bubble flow and 256 observations of churn flow. The other ("mixed") calculates activity measures for a combined data set of 256 observations of bubble flow followed by 256 observations of churn flow. In the mixed case, wavelet coefficients near the transition from bubble to churn will include information from both flow types, while in the pure case the coefficients will all represent uncorrupted information on the flow types. Some care must be taken since the longer mixed data set results in one extra, more coarse, resolution level in the wavelet decomposition. The test data sets consisted of 168 observations of bubble flow, followed by 256 observations of churn flow, and finally another 88 observations of bubble flow, resulting in 512 observations in each case.

The most common statistical application of wavelets is denoising (Donoho et al., 1995). Hence a further option is whether to threshold the wavelet coefficients in an attempt to remove the noise. This may be helpful if the recorded data are corrupted by measurement error which partially masks the information contained in the signals.

Figure 4 shows a summary of the model fitting results. For each modeling approach, the correct classification rates on the test data from the 100 training/test pairs are summarized as a boxplot. The top row shows classification using wavelet absolute value, whereas the other rows are for the wavelet variance activity measure with selected window widths w. Columns represent low and high noise levels (similar results were seen for moderate noise). Within each panel, the left-most boxplot is classification using the raw data for comparison. The remaining boxplots can be divided into four groups defined by type of training data (mixed or pure), and according to whether wavelet thresholding was or was not used. Further, within each

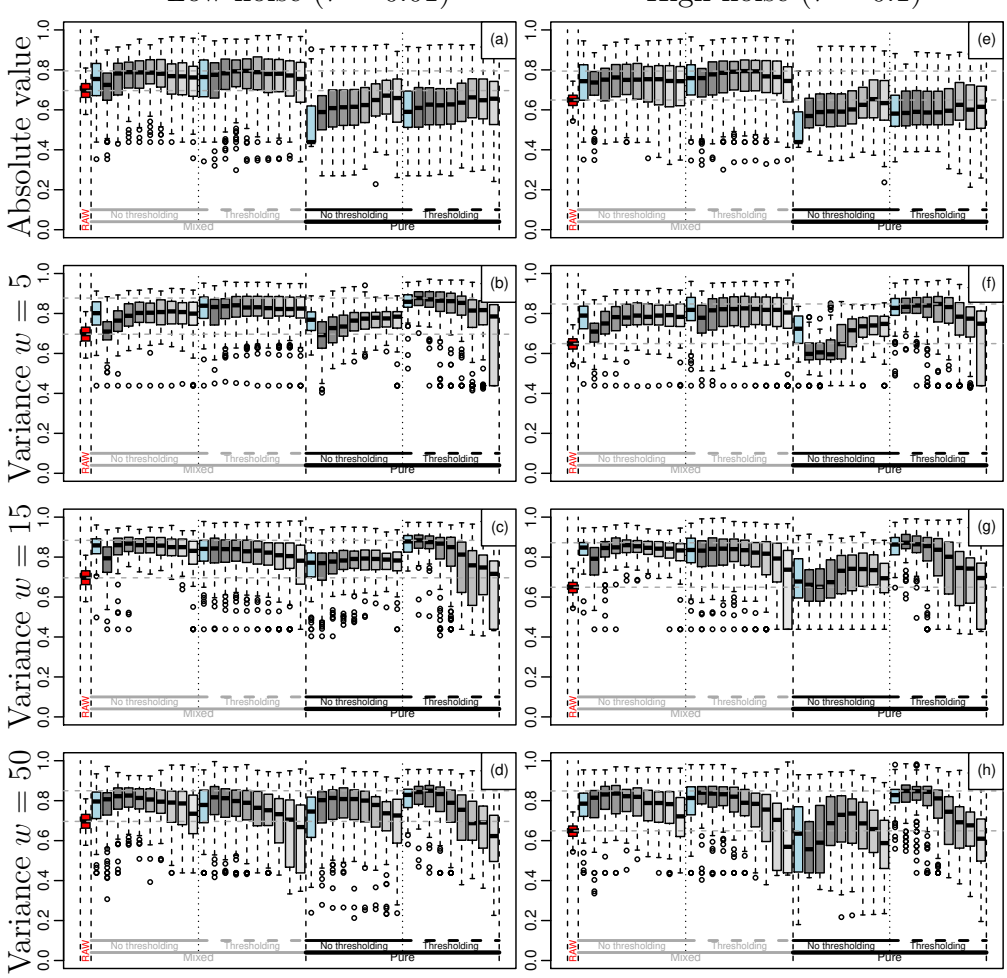

Figure 4: Box plots showing correct classification rates for the different classification methods. The top row has the wavelet absolute value method whereas the other rows use the wavelet variance method with selected window widths. Columns show low (left) and high (right) noise levels. Dashed lines indicate the median correct classification rate for the best method and the model using the raw data.

group of 9, or 10, boxplots the first shows the results after model selection using AIC, then the remainder show results using the best 1, 2,... variables as determined by correct classification rate on the training data set. In each panel, two horizontal dashed lines show the median correct classification rate from (a) the best method and (b) using the raw data.

The first row of Figure 4 shows that, in general, when using the absolute value activity measure, thresholded mixed training data is the preferred approach. However, in our application, wavelet variance generally works better.

The remaining three rows in Figure 4 correspond to the use of the wavelet variance activity measure using window widths of 5, 15 and 50 time points. Overall the patterns are reasonably consistent across window width and noise levels. In general, the classification rates are better when using a window of 15, and the method without thresholding and trained on pure data performs the worst. The other three approaches perform to a similar level, with the approach trained on pure data and using thresholding on the best one or two variables performing the best.

Table 1: Top 10 classification approaches for high noise level.

	Measure	Training	Thresholded	Levels	Rate
1	Var	pure	Y	1	0.877
2	Var	pure	Y	2	0.874
3	Var	pure	Y	3	0.874
4	Var	pure	Y	AIC	0.864
5	Var	mix	N	2	0.860
6	Var	mix	N	3	0.860
7	Var	mix	Y	2	0.857
8	Var	mix	N	4	0.853
9	Var	mix	Y	1	0.852
10	Var	mix	N	5	0.850
⋮	⋮	⋮	⋮	⋮	⋮
24	Abs	mix	Y	6	0.794
⋮	⋮				⋮
57	RAW				0.649
⋮	⋮				⋮

When the approaches are ranked by correct classification rate, Table 1,

the best approaches all use the variance activity measure with thresholding and trained on pure data. The first uses only the best single variable, then two variables and then three variables. Notably, the approach selecting variables using AIC is the fourth best. Most of the rest of the approaches shown use mixed training data with the wavelet variance activity measure, but some use thresholding and some do not, and there are a wide variety of number of variables included. It is only at rank 24 when the best approach using the wavelet absolute value activity measure appears.

In summary, the preferred approach for the simulated tomographic data is to use:

- wavelet thresholding,
- variance activity measure with a window width of 15,
- the single best resolution level, and
- trained on pure data.

For other applications, ideally one should conduct a simulation study to identify the best approach. If, however, it is not possible to perform such an investigation to choose the best resolution levels, then using AIC variable selection performs well.

We now study the preferred approach more closely, presenting more detailed results using the same 100 pairs of training/testing data. Figure 5 shows results from approaches using the wavelet variance activity measure with a window width of 15 for both training approaches and with and without thresholding. The diagram shows the joint relative frequency that each wavelet resolution level appears as best or second-best in terms of correct classification rate. In each graph, the vertical axis identifies the best, and the horizontal axis identifies the second best. The left hand column, marked "M", gives the marginal distribution of the best level.

When using mixed data it is clear that the useful information is contained in levels 4 and 5. Without thresholding, once level 4 or 5 is included, then the second most important is level 8. For mixed data with thresholding, levels 4 and 5 are still the most important but high levels are also sometimes included as the most important. This time there is no clear-cut second best with most levels being represented.

For the pure data without thresholding, the pattern is very different. The best levels are, almost equally, any of 1 to 4, and with all other levels equally second best. The picture for pure data with thresholding is very similar to

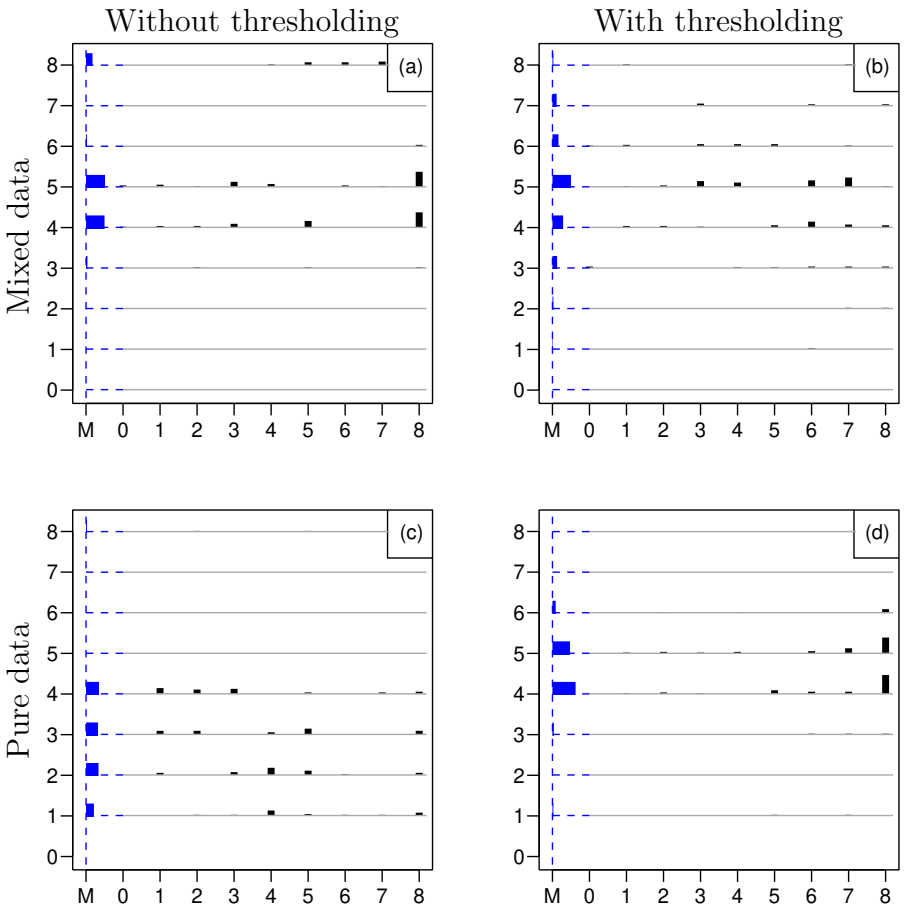

Figure 5: Distribution of best variables (vertical axis) and second best (horizontal axis) using the correct classification rate criterion for wavelet variance method.

the case of mixed data without thresholding. The best are levels 4 and 5 with the highest level appearing often as second best.

Figure 6 shows the performance of the four approaches, along with classification using the raw data, for varying noise levels. It can clearly be seen that for low noise levels the selected methods have similar correct classification rates, but as the noise increases the methods' relative performance of the approaches changes.

When the best model is chosen by AIC, (a), three of the approaches have very similar high correct classification rates. These are mixed data with and without thresholding, and pure data with thresholding. For low and

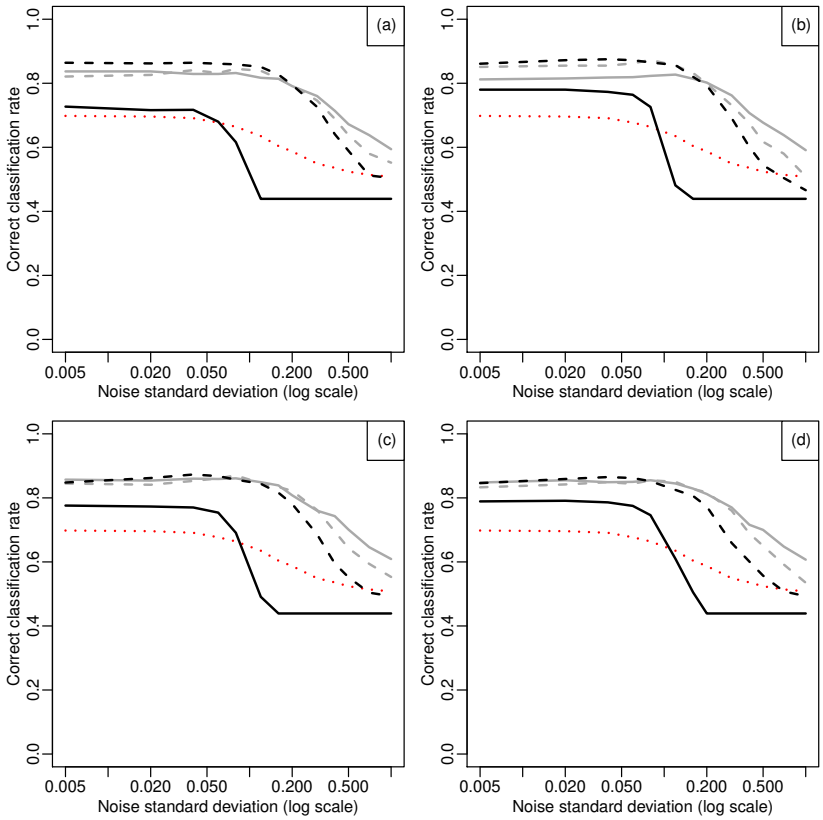

Figure 6: Correct classification rate, using wavelet variance method, for different noise levels: (a) Variables selected by AIC, and (b)–(d) best 1–3 variables respectively, selected by best correct classification rate. Explanatory variables are raw data (dotted line), activity measures trained on mixed data (grey) and trained on pure data (black). Solid lines denote no thresholding and dashed lines thresholding.

moderate noise levels the best is pure data with thresholding, but for high noise levels this becomes the worst of the three. The performance using pure data without thresholding is always lower than the others.

When the correct classification rate using the best 1–3 variables are considered, in panels (b)–(d), the general pattern is similar. In all cases using pure data without thresholding is the worst approach. Using only the best single variable thresholding is better, and with pure data in particular, but as the number of variables included increases the approaches using mixed data improve and eventually become the better.

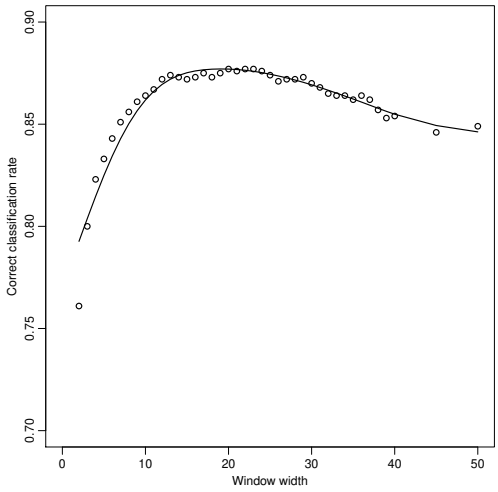

Figure 7: Correct classification rate over 100 test-training pairs for wavelet variance using different window widths.

Figure 7 shows how the correct classification rate varies with window width. The approach uses wavelet thresholding and training on pure data. The approach performs badly if the window width is very small, then rapidly improves until about 10, then is stable until about 20 after which it gradually worsens. From this we see that the approach is not too sensitive to the exact choice of window width though in the range 15 to 20 appears best.

Clearly, if a very narrow window is used then the resulting flow prediction is less reliable due to random variation. As the width increases, the effect of variability reduces and prediction becomes more reliable. However, as the width further increases the prediction is slower to detect a change in flow regime and hence the classification becomes less reliable.

Figure 8 shows how the correct classification rate and the probability of churn changes with time using the preferred approach. As expected the correct classification rate, panel (a), is generally high. Initially, the correct classification rate is almost 99%, but at the first switch in flow regime this drops to zero before more slowly returning to a high level. The rate again dips at the second change of regime. Notice that the method has a higher correct classification rate for bubble than for churn, that the drop in rate is much more rapid than the recovery, and also that when changing from bubble to churn the dip in rate is slightly more dramatic and the recovery

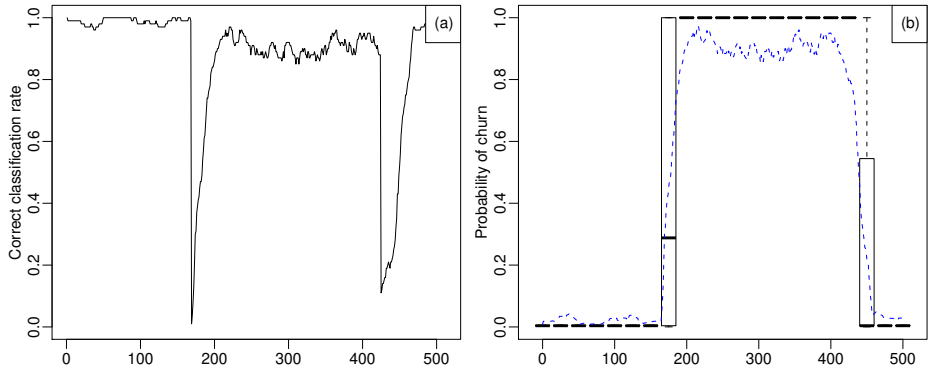

Figure 8: Using the best wavelet variance method, (a) shows the correct classification rate along the trace and (b) shows boxplots of estimated probability of churn.

more rapid than when changing from churn to bubble.

Panel (b) shows boxplots (drawn without outliers) of the estimated probability of churn for the 100 training/test data set pairs at different points along the trace, along with a dotted line showing the mean values. Over most of the trace the range of estimates is so narrow that the box and whiskers appear as a single line. The location of the mean, however, shows that we are dealing with a skew distribution caused by the (omitted) outliers. Only at the change of regime does this pattern differ. At the first change, from bubble to churn, the box covers the entire range of possible values. For the second change, from churn to bubble the box is not so wide, but is still more than half the allowable range with the whiskers covering the full range.

6. Detecting flow change in real data

We now consider detecting the change in flow characteristics on data gathered in a laboratory experiment. Rice was allowed to flow down a vertical tube surrounded by eight sensors, leading to 49 data traces. In this case, the presence or absence of flow can be detected easily using the mean level of the recorded voltages. To validate our method, the mean level was first removed using a wavelet smoother and the resulting residuals analyzed for local frequency characteristics which could diagnose the presence or absence of flow. Figure 9 shows a typical trace (dots), the fitted local mean level, and the residuals remaining after removal of the local mean. The local mean was used to classify observations as two types, “flow” and “no flow”, with an

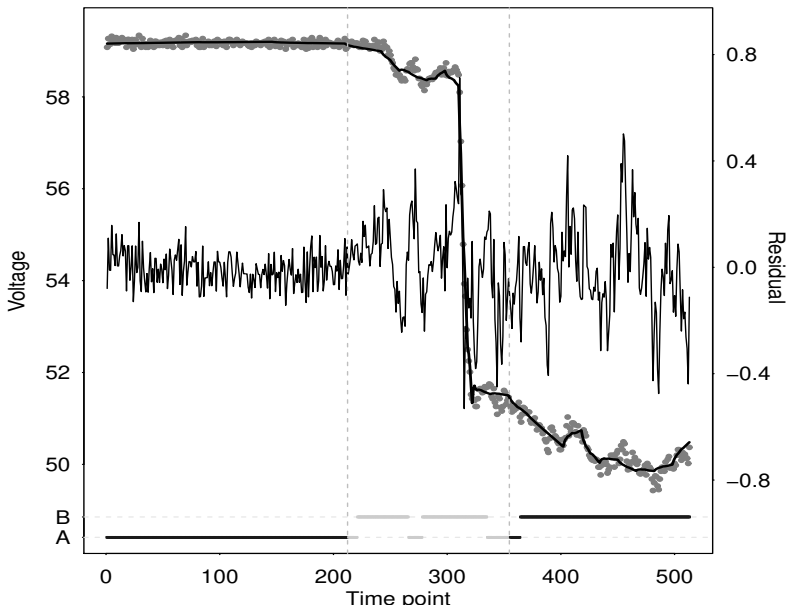

Figure 9: Example trace from rice flow data. The plot shows raw data as dots with a smoothed local mean curve. The vertical lines dividing the trace into "flow", "intermediate" and "no flow" states based on the local mean. The trace across the centre shows the residuals after removal of the local mean, with bars at the bottom of the figure show classification to "flow" (A) or "no flow" (B) using our wavelet-based approach.

indeterminate intermediate stage; these segments are separated in the figure with vertical dashed lines. The intermediate values were not used in training. In our application, this intermediate stage is not a clearly distinguished flow type in its own right, and hence there is no interest in classifying observations to an intermediate state and our response variable is binary. Of course, other applications may have more than two states of interest and hence require different classification methods.

The data were analyzed using the best classification approach found in our simulation study (wavelet variance, pure training data with thresholding, using the single best resolution level). Three replicate experiments were conducted, giving six possible training/test combinations. For each pair of training and test data sets, correct classification rates were computed based only on the observations classified by mean level as flow or no flow. Across the six train/test pairings, the correct classification rate ranged from 97%–100% with mean 99%. One reason for this excellent performance is that the

duration of the intermediate stage of around 150 observations is longer than the wavelet variance window (15 observations). Therefore, once the "no flow" observations commence there is no carry-over effect from the "flow" data.

Other applications may have no intermediate stage. Then carry-over effects following the abrupt change from type A to type B may lead to the first few type B observations being classified as type A. For our data, removing intermediate-stage measurements in the test data, results in correct classification rates of 88.3% to 97.3%, with mean 92.4%.

Figure 9 shows an example of the actual classification, including during the intermediate phase. It is clear that the vast majority of "flow" or "no flow" observations are correctly classified with very little uncertainty. In the intermediate stage, we see blocks of observations being classified to each of "flow" and "no flow". This may occur if the remaining rice is flowing irregularly rather than as a smooth reduction in the flow rate.

7. Discussion

We have proposed a general framework which uses localized frequency variables derived from wavelet coefficients to construct regression models which use transient signal features as predictors. As an explicit example we used this approach to classify the flow regime from electrical tomography data, but many other possible uses can be considered.

Each resolution level corresponds to a frequency band and by considering which levels are most useful in a model we may learn about the frequencies where the processes differ. Using our preferred model (pure data, thresholding, single best resolution level) on the simulated data, we see from Figure 5 that the preferred level is 4 or 5, compared to a finest resolution at level 8. Hence, we deduce that the difference between bubble and churn flow is most striking at a frequency corresponding to some 16–32 observations. (Although we note that the estimation of the instantaneous wavelet variance uses a window of coefficients, thus incorporating data from a wider window.)

It is worth noting that there can be "leakage" between different levels of the NDWT. When the key underlying frequency is somewhere between two levels, there may be difficulty in choosing a single "best" level. This may account for the fact that levels 4 and 5 are preferred almost equally often in our simulation. In such cases, it is possible that averaging activity measures across different resolution levels may be helpful.

We have discussed the use of wavelet variances since the key information in our application lies in the amount of "activity" at given frequencies. Depending on the application, the important information in the signal may be in the mean value of the explanatory variables, in their frequency component, or in other characteristics of the data. Many of these can also be encoded by suitable statistics derived from the NDWT. For example, scaling coefficients are computed as part of the wavelet transform. These are effectively local weighted averages and could be used (possibly via activity measures) to encode localized mean level data as the time series progresses.

Other authors have also considered the use of wavelet coefficients as predictors in other settings; for example Ramsey and Lamport (1998) constructed regression models relating the DWT of both covariate and response time series in economic data and Küçük and Ağiralioğlu (2006) construct similar models in streamflow modeling. Further, wavelet packets (a generalisation of wavelets which results in a much richer set of potential basis functions) have been used in modeling transient underwater signals (Learned and Willsky, 1995), sales data (Michis, 2009), wastewater filtering (Lee et al., 2009), sleep state modeling (Nason et al., 2001) and wind speed prediction (Hunt and Nason, 2001; Nason and Sapatinas, 2002). The greater flexibility of wavelet packets comes at a price; the number of potential covariates given n observations on a single predictor variable is $\log_2(n)$ when using the non-decimated wavelet transform, but $2n-2$ when using the non-decimated wavelet packet transform. Some form of dimension reduction is required to construct a regression model.

Similar wavelet variances to the ones we use have been considered by Maharaj and Alonso (2007, 2014). These authors use wavelet variances and correlations evaluated on disjoint blocks in classifying seismic, EEG, and ECG data. An alternative statistic which can be computed from the NDWT is the evolutionary wavelet spectrum (EWS) based on the locally stationary wavelet process (LSW) model due to Nason et al. (2000); Fryzlewicz and Ombao (2009) and Krzemieniewska et al. (2014) use selected (j, k) pairs from the EWS in classifying seismic and acoustic time series. In these applications, the interest was in allocating an entire time series to one of a number of classes, in contrast to our goal of segmenting a time series as it evolves.

Sanderson et al. (2010) developed a multivariate LSW process model, which Cho and Fryzlewicz (2015) have used in segmentation of multiple locally stationary time series, while Park et al. (2014) have developed estimators for the dependence structure between the multivariate time series.

This multivariate LSW model could be used to extend our approach if the structure of the multiple explanatory time series is of interest.

The obvious use of our models is online monitoring of an evolving process. Wavelets, being computationally efficient and effective at modeling transient features, are well suited to this purpose. Moreover, we note that as new data become available only the most "recent" wavelet coefficients would need recomputing (the exact number being different at the varying resolution levels). Further work would be required to develop efficient algorithms which update only the necessary wavelet coefficients. On the other hand, if the purpose is retrospective analysis of an entire data set, there would be no need to restrict our wavelet coefficients to backward-looking time windows.

Many modern applications collect multiple time series data which need to be analyzed sequentially and in real time. The wavelet-based classification approach described here concentrates characteristic features into a small number of variables which are then used in the classification steps. Although we have only used local frequency characteristics, one might also wish to include mean level information. This could easily be done within our framework by including additional activity measures based on the scaling function coefficients. Our framework produces a method which is fast and allows rapid identification of changes in the monitored process. Although here it is applied to electrical tomography data, the same framework can be used to produce similar classification methods in a wide range of applications.

Acknowledgements: We thank Krzysztof Grudzien and Andrzej Romanowski for collecting the data analyzed in §6 as part of UK Engineering and Physical Sciences Research Council grant GR/R22148/01. Also we are grateful to the anonymous referees for their role in making this a better paper.

References

Aykroyd, R. G., 2015. Statistical image reconstruction. In: Wang, M. (Ed.), Industrial Tomography: Systems and Applications. Woodhead Publishing Ltd, pp. 401–428.

Cho, H., Fryzlewicz, P., 2015. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. J. R. Statist. Soc. B 77 (2), 475–507.

Coifman, R. R., Donoho, D. L., 1995. Translation-invariant de-noising. In: Antoniadis, A., Oppenheim, G. (Eds.), Wavelets and Statistics. Vol. 103 of Lecture Notes in Statistics. Springer-Verlag, New York, pp. 125–150.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., Picard, D., 1995. Wavelet shrinkage: Asymptopia? (with discussion). J. R. Statist. Soc. B 57, 301–369.

Fryzlewicz, P., Ombao, H., 2009. Consistent classification of nonstationary time series using stochastic wavelet representations. Journal of the American Statistical Association 104, 299–312.

Hoyle, B., 2004. Industrial process tomography progress. In: Proc: 3rd International Symposium on Process Tomography in Poland. Łodz, p. 9.

Hunt, K., Nason, G. P., 2001. Wind speed modelling and short-term prediction using wavelets. Wind Engineering 25, 55–61.

Krzemieniewska, K., Eckley, I., Fearnhead, P., 2014. Classification of nonstationary time series. Stat 3, 144–157.

Küçük, M., Ağiralioğlu, N., 2006. Wavelet regression technique for streamflow prediction. Journal of Applied Statistics 33, 943–960.

Learned, R. E., Willsky, A. S., 1995. A wavelet packet approach to transient signal classification. Applied and Computational Harmonic Analysis 2, 265–278.

Lee, H. W., Lee, M. W., Park, J. M., 2009. Multi-scale extension of PLS algorithm for advanced on-line process monitoring. Chemometrics and Intelligent Laboratory Systems 98, 201–212.

Lionheart, W. R. B., 2004. EIT reconstruction algorithms: pitfalls, challenges and recent developments. Physiological Measurements 25, 125–1142.

Maharaj, E., Alonso, A., 2007. Discrimination of locally stationary time series using wavelets. Computational Statistics and Data Analysis 52, 879–895.

Maharaj, E., Alonso, A., 2014. Discriminant analysis of multivariate time series: Application to diagnosis based on ECG signals. Computational Statistics and Data Analysis 70, 67–87.

Michis, A. A., 2009. Forecasting brand sales with wavelet decompositions of related causal series. International Journal of Business Forecasting and Marketing Intelligence 1, 95–110.

Nason, G. P., 2008. Wavelet Methods in Statistics with R. Springer, New York.

Nason, G. P., 2010. wavethresh: Wavelets statistics and transforms. R package version 4.5.
URL `http://CRAN.R-project.org/package=wavethresh`

Nason, G. P., Sapatinas, T., 2002. Wavelet packet transfer function modelling of nonstationary time series. Statistics and Computing 12, 45–56.

Nason, G. P., Sapatinas, T., Sawczenko, A., 2001. Wavelet packet modelling of infant sleep state using heart rate data. Sankhya B 63, 199–217.

Nason, G. P., Silverman, B. W., 1995. The stationary wavelet transform and some statistical applications. In: Antoniadis, A., Oppenheim, G. (Eds.), Wavelets and Statistics. Vol. 103 of Lecture Notes in Statistics. Springer-Verlag, New York, pp. 281–300.

Nason, G. P., von Sachs, R., Kroisandt, G., 2000. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. J. R. Statist. Soc. B 62, 271–292.

Park, T., Eckley, I. A., Ombao, H. C., 2014. Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. IEEE Transactions on Signal Processing 62 (20), 5240.

Polydorides, N., Lionheart, W., 2002. A MATLAB toolkit for three-dimensional electrical impedance tomography: a contribution to the electrical impedance and diffuse optical reconstruction software project. Measurement Science and Technology 13, 1871–1883.

R Development Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.R-project.org`

Ramsey, J. B., Lamport, C., 1998. The decomposition of economic relationships by time scale using wavelets: expenditure and income. Studies in Nonlinear Dynamics and Econometrics 3, 2342.

Sanderson, J., Fryzlewicz, P., Jones, M., 2010. Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. Biometrika 97 (2), 435–446.

Stitt, E. H., James, K., 2003. Process tomography and particle tracking: Research tool and commercial diagnostic tool. In: Proc. 3rd World Congress on Industrial Process Tomography. Banff, pp. 2–10.

Watzenig, D., Fox, C., 2009. A review of statistical modelling and inference for electrical capacitance tomography. Measurement Science and Technology 20, 1–22.

West, R. M., Meng, S., Aykroyd, R. G., Williams, R. A., 2005. Spatial-temporal modeling for electrical impedance imaging of a mixing process. Review of Scientific Instruments 76 (7), 073703.