COMSTA: 6116 Model 3Gsc pp. 1–12 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

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

Q1 Robert G. Aykroyd[a,*], Stuart Barber[a], Luke R. Miller[b]

[a] *University of Leeds, Leeds, UK*

[b] *University of Oxford, Oxford, UK*

ARTICLE INFO

*Article history:*
Received 12 September 2014
Received in revised form 21 July 2015
Accepted 22 July 2015
Available online xxxx

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

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 industrial processes.



## 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.

☆ Software to perform the data analysis example is available from www.maths.leeds.ac.uk/~stuart/research.

* Correspondence to: Department of Statistics, University of Leeds, LS2 9JT, UK.
*E-mail address:* R.G.Aykroyd@leeds.ac.uk (R.G. Aykroyd).

http://dx.doi.org/10.1016/j.csda.2015.07.009
0167-9473/© 2015 Published by Elsevier B.V.

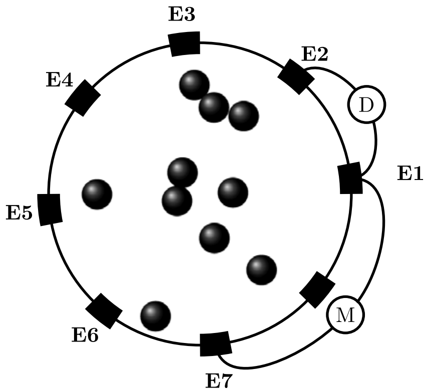


**Fig. 1.** Diagram of data collection protocol showing drive and measurement circuits connecting electrode on the pipe boundary with bubbles passing through the pipe.

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 Section 3. We propose the use of wavelets in relating time-series measurements to the response variable in Section 4. In particular, functions 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 Section 5 and real data in Section 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). Fig. 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, $\phi$, 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, 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 $\tau^2$ is added to yield the simulated data set.

Fig. 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