Cointegration and Nonstationarity in the Context of Multiresolution Analysis

K. Worden\textsuperscript{1}, E.J. Cross\textsuperscript{1} and A. Kyprianou\textsuperscript{2}
\textsuperscript{1}Dynamics Research Group, Department of Mechanical Engineering, Mappin Street, Sheffield S1 3JD, UK.
\textsuperscript{2}Department of Mechanical and Manufacturing Engineering, University of Cyprus, 75 Kallipoleos, Cyprus.
E-mail: e.j.cross@sheffield.ac.uk

Abstract. Cointegration has established itself as a powerful means of projecting out long-term trends from time-series data in the context of econometrics. Recent work by the current authors has further established that cointegration can be applied profitably in the context of structural health monitoring (SHM), where it is desirable to project out the effects of environmental and operational variations from data in order that they do not generate false positives in diagnostic tests. The concept of cointegration is partly built on a clear understanding of the ideas of stationarity and nonstationarity for time-series. Nonstationarity in this context is ‘traditionally’ established through the use of statistical tests, e.g. the hypothesis test based on the augmented Dickey-Fuller statistic. However, it is important to understand the distinction in this case between ‘trend’ stationarity and stationarity of the AR models typically fitted as part of the analysis process. The current paper will discuss this distinction in the context of SHM data and will extend the discussion by the introduction of multi-resolution (discrete wavelet) analysis as a means of characterising the time-scales on which nonstationarity manifests itself. The discussion will be based on synthetic data and also on experimental data for the guided-wave SHM of a composite plate.

1. Introduction
One of the main problems in the field of Structural Health Monitoring (SHM) is that of confounding influences arising in data-based approaches. If one adopts a data-based approach to SHM, one is immediately faced with the fact that data from damaged structures are very scarce. The high cost of many of the types of structure that concern SHM, like aeroplanes and bridges, means that it is inconceivable that one could acquire data from damaged structures by programmed experimentation. The complexity of structures also means that high-fidelity modelling is not often a solution. Fortunately, for the most important problem, that of damage detection, novelty detection methods are available which do not require samples of data from the damaged structure; only data from the normal condition is needed [1]. Sadly, another problem immediately arises. Novelty detection methods rely on detecting changes from an established normal condition in order to infer damage; if the structure changes for a benign reason like a change in the operational or environmental conditions, a false alarm may occur. (This matter has been the object of a great deal of research in the recent past and a good survey can be found in [2].) It is sometimes the case that the effects of operational and environmental changes manifest themselves as long term trends in measured time series. In recent work by the first two
authors here, the technique of cointegration from the econometrics literature has been adapted to SHM purposes as an effective means of purging time series data of long term trends [3]. The analysis is based largely around an algorithm called the Johansen procedure which is able to combine two or more time series which share common trends and form a residual signal free of those trends [4, 5]. The work has led to some insight into how issues of nonstationarity impact on SHM feature selection and these ideas are carried forward in the current paper.

In general, in the context of SHM, the changes in a signal induced by environmental variations around the structure of interest happen on a longer time scale than those which characterise the dynamical behaviour of the structure. However, one would usually expect the effects of damage to manifest themselves in the dynamical behaviour. In an ideal situation, one might hope to remove the influence of the environment by applying a high-pass filter to the data. Unfortunately, things are not often quite so simple, and nonstationarity from benign causes may appear on different time scales. A good example is provided by the SHM of bridges. The behaviour of a bridge will almost always be dependent on ambient temperature, and temperature variations can occur on widely disparate time scales. There will be seasonal variation with a period of a year and superimposed on this will be a daily variation as a result of the day/night cycle. The behaviour of the bridge will also change as a result of operational variations; traffic loading being the major concern. In this case, there are daily variations as a result of rush hour etc., but there will also be a weekly pattern of change as the working week moves into the weekend. All these time scales are still long compared to the natural periods of the structure, but one can see that a simple division of the signal into low and high-frequency components may be less informative than one might desire.

Fortunately there exist more refined means of distinguishing the behaviour of a structure on different time scales: one of the most prevalent being multi-resolution analysis [6]. For the moment, it suffices to say that this results from the application of a sequence of filtering operations. Starting from the original signal, one applies a high-pass filter with a given cut-off and designates this the first level of detail, at the same time a low-pass filter with the same cut-off is applied and the result is designated the approximation. This process is repeated on the approximation signal to yield a second level of detail and a new approximation. The process can be repeated until the low-frequency approximation is reduced to monotonic or even constant behavior (Figure 1).

The whole process is best described in the context of wavelet analysis and the sequential division into detail and approximation is the motivation for one of the fastest algorithms for estimating wavelet coefficients [7]. Before explaining in more detail how wavelet analysis can

![Figure 1. A schematic for multi-resolution analysis.](image-url)
perhaps shed light on nonstationarity and cointegration of signals, it will be useful to distinguish two types of nonstationarity.

2. Nonstationary time series

In the field of econometrics where the technique of cointegration emerged, two models are widely used to capture and explain nonstationarity in the time series of various economic indicators. In their very simplest forms these models look like [8]:

- **Deterministic time trend**

  \[ y_t = \alpha t + \beta + \psi(L) \epsilon_t \]  

- **Unit Root**

  \[ y_t = y_{t-1} + \beta + \psi(L \epsilon_t \]  

where \( t \) is the current time instant, \( y_t \) is the variable of the event whose dynamics are studied, \( \alpha \) and \( \beta \) are constants, \( L \) describes the lag operator (i.e. \( y_{t-1} = L y_t \)) and \( \psi(L) \) is a ratio of polynomials in \( L \), \( \epsilon_t \) is the time series of innovations modeled by an independent sequence of Gaussian random variables.

The essential difference between equations (1) and (2) is the extra factor \((1 - L)\) that is contributed by the lag \( y_{t-1} \) in (2). The root that this factor contributes to the characteristic polynomial of the model is unity and hence the time series modelled by (2) is called a unit root process [8]. The presence of this unit root implies that the model is neutrally stable. This kind of stability is the cause of the qualitative differences that are directly observable between time series modeled by (1) and (2) which will now be illustrated.

Figure 2 shows time series generated by two different models,

\[ y_t = y_{t-1} + 0.5 + \epsilon_t \]  

and,

\[ y_t = 0.5 t + \epsilon_t \]  

i.e. models of a unit root process and a process with a deterministic linear trend, respectively.

A detailed look at the dynamics of the processes shows that both models actually incorporate the same linear trend, namely \( y_t = 0.5 t \). The ‘noise’ process here \( \epsilon_t \) is drawn from the zero mean Gaussian distribution \( N(0, 0.5) \).

The first conclusion one draws from Figure 2 is that the process containing the unit root drifts away from the backbone linear trend and persist in staying adrift. This is as if, due to the presence of an external disturbance, the equilibrium position of the process changes for significant time periods to a new position, until a new disturbance again changes this position to a new one and carries on doing this as long as external disturbances capable of changing the new positions occur. These shifts in the equilibrium are a major concern to economicists. In the context of SHM, it may be that deterministic trends will more often be the norm. Both of the types of nonstationarity shown here are manifesting themselves on the longest time scales and in this ‘ideal’ case, one might imagine that a single high/low frequency division of the data will separate the trends from the noise dynamics. In fact, this turns out to be the case, Figures 3 and 4 show that a division of the signals into levels of detail and an approximation, suffices to ‘trap’ the trend in the approximation level. (Each of the Figures in question shows the first three levels of detail \((D_1, D_2 \text{ and } D_3)\) and the remaining approximation \((A_3)\) at this level of decomposition.)

It is possible at this point to speculate on a deeper connection between cointegration analysis and multi-resolution analysis. Cointegration is based on the idea, roughly stated, that one can sometimes combine a number of signals sharing common long-term trends into a signal which
Figure 2. Time series generated by (3) (black line) and (4) (blue line), $\sigma=5.0$

Figure 3. Multi-resolution analysis for signal with deterministic trend.

is purged of these trends. By this process - implemented in [3] for example by the Johansen procedure - one constructs a signal which is more stationary than the originals. One constructs combination of signals, which has the stationarity characteristics of differenced versions of the original signals and it is clear how differencing in general will remove trends - for example, a first difference would be sufficient to remove the linear deterministic trend illustrated earlier. The derivative operator - in spectral terms - is equivalent to a type of high-pass filter, so one can see how the sequence of detail/approximation divisions in the multi-resolution process will result in a sequence of signal components steadily changing their stationarity properties. There is no room in the current paper to develop these ideas in mathematical detail and this will be
pursued in future work.

As discussed in the introduction, the nonstationarity in a signal may not show up in the longest time scales and it will generally be the case that it will show itself in more than one level of approximation or detail in the multi-resolution analysis. In order to illustrate this and to show how the multi-resolution analysis can potentially be exploited further, an experimental study involving guided-wave SHM will be discussed. However, before discussing the data, it will be useful to provide a little more detail on how the multi-resolution analysis can be accomplished in practice.

3. Wavelet Analysis

The wavelet transform is a linear transformation that decomposes a given function $x(t)$ into a superposition of elementary functions $\psi_{a,b}(t)$ derived from an analysing or mother wavelet $\psi(t)$ by scaling and translation i.e.,

$$\psi_{a,b}(t) = \psi^*\left(\frac{t - b}{a}\right)$$  \hspace{1cm} (5)

where $*$ denotes complex conjugation, $b$ is a translation parameter indicating the time locality and $a$ ($a > 0$) is a dilation or scale parameter. Because of the incorporation of the translation parameter, the wavelet transform - unlike the Fourier transform - is ideally suited to the analysis of nonstationary signals [6]. Given the basis of elementary functions in equation (5), the continuous wavelet transform (CWT) is defined as,

$$W^x_\psi(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t - b}{a}\right) dt$$  \hspace{1cm} (6)

The continuous wavelet transform has many useful properties which have been very widely studied (a good reference is [9]); however, if the data of interest is sampled and thus confined to a discrete set, the discrete wavelet transform (DWT) can be used with reference to a finite set of translation and dilation parameters. The dyadic transform results from setting the dilations and translations to be $a_j = 2^j$ and $b_{j,k} = k/2^j$. Within the dyadic framework, the orthogonal wavelet transform (OWT) can be defined. A function $\psi(t)$ is called an orthogonal wavelet if the family,
ψ_{m,k}(t) = 2^{m/2}ψ(2^m t - k), \quad (m, k \in \mathbb{Z}) \quad (7)

forms an orthonormal basis of $L^2(\mathbb{R})$, which is to say that,

$$
<\psi_{m,k}, \psi_{n,l}> = \delta_{mn}\delta_{kl} \quad (8)
$$

for all allowed integers $m, n, k, l$, where $<$, $>$ is the usual inner product defined by,

$$
<h, g> = \int_{-\infty}^{\infty} h^*(t)g(t)dt \quad (9)
$$

and $\delta_{mn}$ is the Kronecker symbol which equals unity if $m = n$, and to zero otherwise.

The orthogonal wavelet transform can now be defined by,

$$
x_{m}^{k} = \int_{-\infty}^{\infty} x(t)\psi_{m,k}(t)dt \quad (10)
$$

The transform is simply a linear combination of the basis functions. The scale decomposition leads to a partitioning in the time-domain that is finer at the higher scales. Any decomposed function can be represented as a sum of $m$ wavelet levels,

$$
x_{m}(t) = \sum_{k} x_{m}^{k} \psi_{m,k}(t) \quad (11)
$$

These levels represents the time behaviour of the signal within different scale bands and gives their contribution to the total signal energy; higher levels refer to higher scales. This decomposition clearly corresponds to the division into levels of detail and approximation discussed earlier. From now on, this paper will adopt a terminology based on the wavelet levels rather than further speaking of approximation and detail. Each level is determined by a number of wavelets which can be fitted into the interval of interest. If Daubechies’ wavelets are used (as they will here), it can be shown that level $m$ is composed of $2^m$ basis functions [7]. (In fact, in this paper, due to a different labelling of the levels, level $m$ will be composed of $2^{(m-1)}$ wavelets.) Because the higher levels (with this labelling) have the higher frequency content, on the basis of arguments offered earlier, one might expect the ‘degree of stationarity’ of the levels to increase with level number.

Similarly to the continuous wavelet transform, the inverse transform or synthesis formula for the orthogonal decomposition can be obtained, in fact,

$$
x(t) = \sum_{m,k} x_{m}^{k} \psi_{m,k}(t) \quad (12)
$$

which simply represents a sum over the levels.

The OWT was computed throughout this work using the C-code provided by [10] and made use of the Daubechies family of wavelets $\psi_{r}(t)$ [11]. The Daubechies’ wavelets cannot be given in a closed form and must be determined recursively from the so-called scaling functions which are solutions of the functional equation [12],

$$
\varphi(t) = \sqrt{2} \sum_{k=0}^{2^{r}-1} c_k \varphi(2t - k) \quad (13)
$$

where,

$$
\int_{-\infty}^{\infty} \varphi(t)dt = 1 \quad (14)
$$
4. The Experimental Data

The results of the analysis methods introduced here will be based on a benchmark study, originating from the Brite-Euram project DAMASCOS, which attempted to use Lamb-waves to detect damage in a composite plate subjected to cyclic temperature variations. The test set up will be briefly described below.

Data collected by the DAMASCOS consortium originated from a series of tests which measured the propagation of Lamb-waves through a composite plate. The plate material was a carbon fibre reinforced plastic (CFRP) laminate with a $0^\circ/90^\circ$ lay-up. Two identical piezoceramic discs bonded to the midpoint of the edges of the plate were used to transmit and receive fundamental symmetric and antisymmetric Lamb-waves. In the particular test this paper concerns, the instrumented composite plate was placed in an environmental chamber whilst Lamb-waves were recorded every minute. The test composed of three different parts. In the first part, the temperature was held constant at 25$^\circ$C. In the second part the temperature was cycled between 10 and 30$^\circ$C every nine hours. In the final part of the testing, the temperature continued to be cycled, however, damage was introduced by drilling a 10mm hole in the plate between the two sensors. More details of the test can be found in [13].

Manson [13] used the DAMASCOS benchmark data to investigate the feasibility of discovering features for damage detection from the response spectrum that are insensitive to environmental variations. Following the work of Manson, the features chosen for analysis are 50 spectral lines from the frequency spectrum (numbers 46-95) which occur around the peak. Figure 5 illustrates the variation of amplitude of the 50 spectral lines under investigation for the duration of the testing.

The first part of the test corresponds to the sample points between 0 and 1078; during this period the temperature was held constant. Between points 1079 and 2159, the plate was cycled in temperature three times as described above, the end-points of the temperature cycles are indicated by the vertical dashed lines in Figure 5. The damage to plate was introduced at point 2207 and this is indicated by a solid vertical line in the figure. A final temperature cycle for

Figure 5. Features (spectral lines 46-95) over the whole period of the Lamb-wave environmental test.
the plate with damage was applied up to point 2519 in the figure (again a dashed line shows the bounds). Beyond that point, the damaged plate was subjected to different, less systematic temperature variations for various reasons.

In a previous analysis [3], the authors used cointegration to create a feature that was insensitive to temperature-induced variation but still sensitive to damage. To simplify matters and improve the numerical conditioning of the problem, only the 20 features corresponding to the first 20 spectral features were used. The details of the procedure can be found in [3], but the basic principle was simply to create a cointegrated residual using the Johansen procedure on training data encompassing the temperature variation, but not the damage. This residual could then be monitored for anomaly detection; the results were excellent with the damage detected unambiguously. Figure 6 shows a 'control chart' with 3σ bounds for normal behaviour on the cointegrated residual; the damage is detected promptly. The vertical dashed lines in the figure now show the bounds of the training set used, while the solid vertical line still indicates the onset of damage.

5. Results

The availability of the multi-resolution analysis (wavelet level decomposition) allows two interesting possibilities for looking at the experimental data. The first is the possibility of checking the idea that the stationarity of the level signals will increase steadily with level number. In order to verify such an idea, one clearly needs a measure of stationarity of some sort. There are a number of different statistical tests for stationarity [8], but the one adopted here will be the augmented Dickey-Fuller (ADF) statistic. The computation of the ADF statistic is a little complicated (see [3], for a tutorial treatment in the context of SHM problems), but the essential idea is that the ADF statistic becomes increasingly negative as signals become more stationary.

Before applying the ADF statistic, the wavelet level decomposition must be carried-out. This was accomplished here using the orthogonal wavelet transform as discussed earlier. Because the OWT requires the number of samples to be a power of two, the feature data were truncated to 4096 (2^12) points by removing the first portion of the data where the temperature was held...
constant. Because of the details of the OWT implementation, this led to 13 levels; however, the first two levels are essentially all copies of the mother wavelet and do not contain significant information from the signals. The Daubechies $N = 4$ wavelet was used for the analysis; this was considered a good choice because, although it is less smooth than the higher-order wavelets, its shape is quite well matched to the ‘triangular wave’ temperature ramping cycles. The level decomposition thus yielded a $13 \times 20$ matrix of level time series. The ADF statistic was computed for all of these signals and the results were averaged across the features to give a mean ADF statistic per level. This quantity is plotted in Figure 7.

![Figure 7](image.png)

**Figure 7.** Mean ADF statistic across the feature set as a function of wavelet level.

The figure gives excellent support for the proposal that the multi-resolution decomposition process yields a sequence of signals steadily increasing in stationarity. In terms of the computation of the ADF statistic, the first 3 or 4 level values are probably not to be trusted as the signals actually reflect the nature of the mother wavelet rather than the the original signal.

The second interesting possibility made available by the multi-resolution analysis is that of computing a cointegrated residual for damage detection on a level-by-level basis. Essentially one applies the Johansen procedure to the set of signals at a given level $i$ across the feature set. This procedure generated a large body of interesting results and a detailed discussion will need to be postponed to a later publication for reasons of space. However, it will be possible to consider one or two interesting facets of the work. First, one might wish to consider the ‘most stationary’ of the levels for analysis, as this might be optimal for SHM purposes in some sense. The cointegrated residual for level 12 (levels are labelled here from 0 to 12) is given in Figure 8 (the region of training data is shown by vertical dashed lines, the onset of damage is indicated by the dotted line; the horizontal lines are the $3\sigma$ confidence intervals as before).

There is an extremely clear indication of damage indicated by a spike in the residual. Further, a control chart based on variance would show a clear change of regime with the onset of damage. (The later spike is not a concern as it is in the later, uncontrolled, part of the test which does not concern us here.) Having considered what happens at the most stationary level, one might wish to consider what happens at arguably the ‘least stationary’ level i.e. that corresponding to the time scale of the temperature variations. Although the choice is not trivial the most likely
candidate for least stationary level is level 5 (Figure 9).

One can see that the 'period' of the mother wavelet for level 5 matches well the period of the temperature cycling. Now, a subtlety arose in the application of the Johansen procedure at level 5 (and actually many of the other levels). A little thought will show the reader that this level can only accommodate 16 basis functions and therefore the 20 features will actually be linearly dependent. This was evident from the fact that the Johansen procedure became very ill-conditioned and gave combination coefficients 4 or 5 orders of magnitude greater than
the data themselves. Regularisation was needed, and the simplest way to accomplish this was to add noise to the feature data. A short trial-and-error process showed that a noise rms of approximately 0.25% of the level variance gave good results (in future, the noise level will be determined by cross-validation). After regularisation, the cointegrated residual shown in Figure 10 was obtained.

This figure shows that even the most nonstationary signal component can be purged of the trends by the Johanse procedure. In fact, Figure 10 is in a sense inappropriate as the noise corrupted (regularised) signals have been used. The noise is only needed for regularisation and it is the clean signals which should be used to compute the cointegrated residual as in Figure 11.

The results are excellent. Although there are some excursions outside the confidence limits before the onset of damage, they are small and could perhaps be eliminated by optimisation of the regularisation parameter as discussed earlier. More significantly, the residual causes an early alarm for the true damage; this is because of the limited number of time points where a basis function can be placed at level 5. This issue is not addressed here as it is part of a wider issue concerning on-line implementation of the approach.

6. Conclusions
The material of this paper has rather an exploratory nature, and as such, long conclusions are probably not appropriate. The idea has been to explore some issues relating to ideas of nonstationarity as applied in the context of SHM. The main objective has been to see if a multi-resolution analysis can lead to insight into the use of cointegration methods for the removal of trends from SHM data; such trends being a common problem when novelty detection methods are applied. The results indicate that the multi-resolution approach has some rather interesting uses, not least by allowing the removal of trends at independent time scales. This is interesting because the structural changes from environmental and operational variations will often show themselves on different time scales. At a deeper level, the research has given support to the idea that cointegration and multi-resolution analysis are fundamentally connected; this idea will be
pursued in further work.

Acknowledgements
The authors would like to express their gratitude to Dr Graeme Manson and Professor Wieslaw Staszewski of the University of Sheffield for providing the DAMASCOS data and the wavelet code respectively; they would also like to thank the UK Engineering and Physical Sciences Research Council (EPSRC) for funding this research.

References
[1] Worden K, Manson G and Surace C 2007 In Proceedings of DAMAS 2007 (Torino) p 3.
[2] Sohn H 2007 Philosophical Transactions of the Royal Society A: Physical and Engineering Sciences 356 539.
[3] Cross E J, Worden K and Chen Q 2011 Cointegration; A novel approach for the removal of environmental trends in structural health monitoring data Preprint: submitted to Proceedings of the Royal Society - Series A.
[4] Johansen S 1995 Likelihood-Based Inference in Cointegrated Vector Autoregressive Models (Oxford University Press).
[5] Juselius K 2006 The Cointegrated VAR Model: Methodology and Applications (Oxford University Press).
[6] Kaiser G 1994 A Friendly Guide to Wavelets (Birkhauser).
[7] Mallat S 1998 A Wavelet Tour of Signal Processing (Academic Press).
[8] Hamilton J D 1994 Time Series Analysis (Princeton University Press).
[9] Chui C K 1992 An Introduction to Wavelets (Academic Press).
[10] Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1986 Numerical Recipes - The Art of Scientific Computing (Cambridge University Press).
[11] Daubechies I 1992 Ten Lectures on Wavelets (SIAM monographs).
[12] Meyer Y 1993 Wavelets: Algorithms and Applications (SIAM monographs).
[13] Manson G 2002 Proceedings of the 3rd International Conference on Identification in Engineering Systems (Swansea).