# On Engle-Granger Cointegration Using Treed Gaussian Processes

KEITH WORDEN and ELIZABETH J. CROSS

## ABSTRACT

Over the last decade or so, cointegration has emerged as arguably the state-of-the-art in terms of removing Environmental and Operational Variations (EOVs) from structural health monitoring (SHM) data. When data channels share common trends which can be removed by linear projection, the Johansen procedure, a maximum-likelihood approach developed within the field of econometrics, is provably optimal. Unfortunately, SHM problems can present where the trends occupy a nonlinear submanifold of the feature space, and in this case, linear cointegration/projection fails. It is still possible to make progress in this case by moving to the older Engle-Granger approach to cointegration, where one linearly regresses one of the feature space variables on the others; nonlinear cointegration is then 'simply' the application of an appropriate nonlinear regressor. Over the years, a number of nonlinear regression algorithms have been applied, motivated by machine learning or evolutionary computation; each with pros and cons. The aim of the current paper is to demonstrate an approach based on Treed Gaussian Processes (TGP); the advantage being that the algorithm allows switching between cointegration models in different parts of the feature space. Examination of the switching points can provide insight into the physical processes driving the nonlinearity. The approach is demonstrated here on the well-known Z24 Bridge data set, where the ambient temperature drives EOVs which cannot be removed by linear methods.

## INTRODUCTION

The past three decades have seen considerable progress in data-based Structural Health Monitoring (SHM) [1], notably in terms of its recent transition into industrial usage. Along the way, a number of significant challenges have been overcome, and significant challenges remain. One of the most significant problems has been that of *confounding influences*, or *Environmental and Operational Variations* (EOVs) [2]. A dominant issue in SHM has always been that of acquiring damage-state data for high-value structures; the usual situation being that only data from the undamaged structure

Keith Worden, Elizabeth J. Cross, Dynamics Research Group (DRG), Department of Mechanical Engi- neering, The University of Sheffield, Sheffield, S10 2TN, U.K.. Email: k.worden@sheffield.ac.uk

will be available. However, the lack of damage-state data has not precluded the development of effective damage detection algorithms. Such algorithms are usually based on the idea of *novelty* or *change* detection; in the language of machine learning they are designated *unsupervised learning* algorithms. Novelty detection works by establishing a statistical model of the normal condition of the structure in terms of feature data from the undamaged state; the simple premise is that any subsequent test data which deviate significantly from the model, thus signal change, from which damage is inferred. Unfortunately, because such algorithms detect *changes*, they can be confused by any benign changes in the environment or operation of the structure; the result can be an undesirable false alarm. For example, if one is concerned with detecting damage by monitoring the natural frequencies of a bridge, variations in temperature or mass loading (traffic) can affect the frequencies as much (if not more), as damage.

The solution to the problem of confounding influences is to find a systematic means of removing them from feature data; various effective means have been proposed over the years, usually in terms of *subtraction* or *projection* schemes [3]. Arguably the most effective of these schemes has proved to be *cointegration*. Cointegration was developed in the econometrics literature in the context of identifying long-term trends in economic or financial date [4]. The basic premise is that, if a number of variables share common long-term trends, this *nonstationary* behaviour can be removed by forming an appropriate linear combination of the variables, for which the trends cancel. The implication for SHM should be clear: if a set of feature variables share a dependence on long-term confounding influences, a cointegrated feature vector can be formed, purged of those influences. A series of papers starting with [5,6], showed the effectiveness of cointegration for SHM purposes. The theory in these papers was based on the Johansen procedure [7], which is able to form a maximum-likelihood estimate of the necessary linear combinations. As powerful and elegant as the Johansen procedure proves to be, its dependence on linear combinations is a significant restriction. In essence, the procedure constructs a linear transformation of feature data in which the transformed variables are stationary (purged of trends). In the situation that the shared long-term trends in the feature data are nonlinearly related, the Johansen procedure breaks down. A general solution to the nonlinear problem is simply stated, one seeks a nonlinear transformation on the feature variables which purges the trends. Unfortunately, exactly like the situation for modal analysis in structural dynamics, such nonlinear transformations cannot be constructed in general.

Despite the lack of a nonlinear Johansen procedure, all is not lost; in fact, all one needs to do is take a step backwards to an earlier take on cointegration. The first works on cointegration adopted a different approach – the Engle-Granger (EG) method [4]. Although the EG approach entails a number of careful steps, the essence of the method is to linearly regress one of the variables of interest on the others; if the regression is effective, the model residual (noise) will be purged of the long-term trends as required[1].

[1]A great deal of care is usually needed in any cointegration strategy, apart from the actual removal stage. For example, one would usually establish that the variables of interest are all integrated to the same order; one would also carry out stationarity tests on each variable [4]. These stages are omitted in the discussion in this paper because they have already been considered in the literature of the case study here – the Z24 Bridge. In addition, one can argue on the basis of engineering insight that the integration order for civil infrastructure data is not an issue [8].

The nonlinear generalisation of EG is therefore evident (at least in basic terms), one simply carries out a nonlinear regression of the variable of interest on the others. With this viewpoint, the problem is thus reduced to a choice of multivariate regression algorithm. This choice is influenced by a number of factors: explanatory power of the algorithm, computational cost etc. The algorithm presented here will be the *Treed Gaussian Process* (TGP) algorithm; while this is a universal approximator, it does come at a computational cost. However, an important aspect of the TGP algorithm is that it can fit *piecewise-smooth* models, and this makes it well-suited to the case study presented here, in which the long-term trends switch between two linear regimes. In this case, the TGP can discover important physics in the data, i.e. the points at which the trends switch. The problem considered here will be that of monitoring and damage detection for the Z24 Bridge [9]. In fact, the TGP model has already proved effective for the Z24 case in terms of fitting surrogate models [10].

The layout of the paper is as follows. The next section will give a brief summary of the TGP model, followed by a section introducing the problem of interest – monitoring the Z24 Bridge. The results of the analysis will then be presented and discussed, followed by conclusions.

## TREED GAUSSIAN PROCESSES (TGPS)

TGPs are part of a class of algorithms called *regression trees*. The idea of a regression tree is fairly simple to state (much of the theory and practice of such trees can be attributed to Breiman and co-workers, and a good reference is [11]. The idea is to partition the independent variable space into regions over which the response behaviour is smooth and fit low-order regression models over each region. If the partitioning is carried out by hand, the resulting problem is still amenable to linear least-squares methods. The idea, however, of a regression tree in general is that the partitions are determined from the data as part of the modelling problem; this renders the estimation problem highly nonlinear and alternatives to least-squares are needed. Breiman and co-workers established a greedy algorithm for fitting the trees that gave good (but suboptimal) solutions. If an effective partition of the data is found, linear regression models over each distinct region can give excellent results; however, in principle, any algorithm can be used once the data have been partitioned into sensible regions. Once the concept of Classification and Regression Trees (CART) was established, arguably the next major advance was the development of a Bayesian framework for the algorithm [12, 13]. The new algorithm was based on rigorous concepts of probability theory and proved an effective departure from the greedy algorithm. In Bayesian CART, a prior probability distribution was proposed over all possible tree structures as well as all possible coefficients; this was then refined by using the data to determine which tree was supported by the greatest evidence. The original formulation is too complicated to describe here without taking this paper a long way from its illustrative objectives. In the original Chipman formulation [12, 13], all the regression models within the tree were linear; this restriction was later removed by Gramacy, who replaced the linear models by more powerful Gaussian Process (GP) models; Gramacy's work also extended the Bayesian formulation of the problem significantly [14]. The Treed Gaussian Process (TGP) partitions the variable space in much the same way as a Bayesian CART and then fits GP regressors over each independent

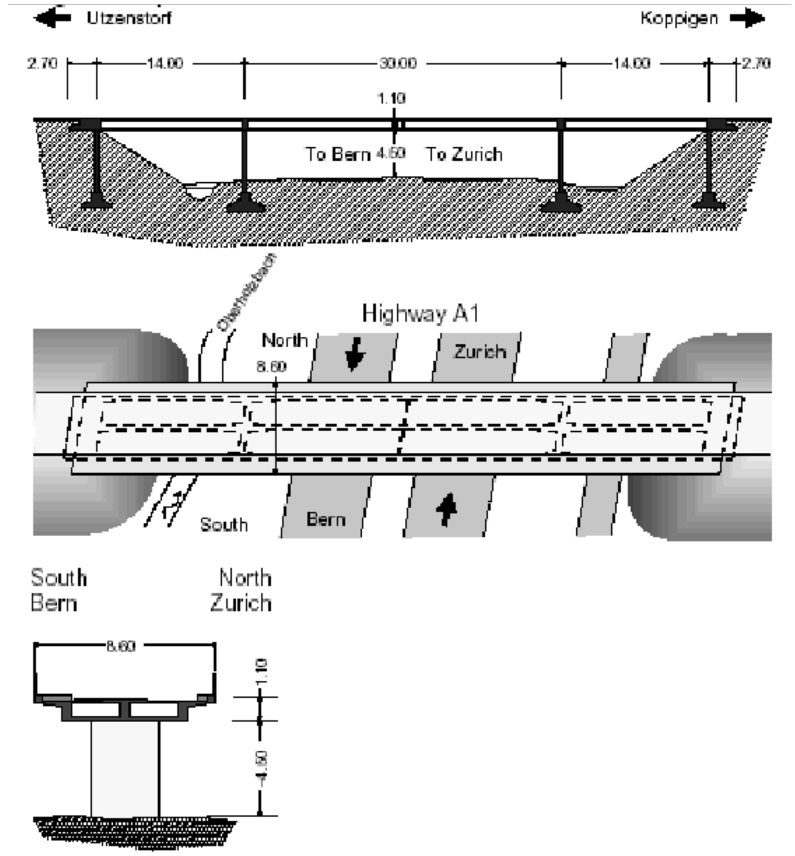


Figure 1. Schematic of the Z24 Bridge.

region. The software used for modelling in this work is the TGP package written by Gramacy in the language R [15]. Although the full capability of the TGP is needed in general switching models, it will be shown that a treed linear model also proves effective here.

## THE Z24 BRIDGE

The Z-24 Bridge was a pre-stressed concrete highway bridge in Switzerland (1), subjected to a comprehensive monitoring campaign under the 'SIMCES project' [16], prior to its demolishment in the late 1990s. The bridge has since become a landmark benchmark study in SHM and the subject of a great many papers. The monitoring campaign, which spanned a whole year, tracked modal parameters and included extensive measurement of the environmental factors affecting the structure, such as air temperature, soil temperature, humidity etc. The Z24 monitoring exercise was an important study in the history of SHM developments because towards the end of the monitoring campaign researchers were able to introduce a number of realistic damage scenarios to the structure. In order, these scenarios were [17]:

- Pier settlement.
- Tilt of foundation followed by settlement removal.
- Concrete spalling.
- Landslide.

- Concrete hinge failure.
- Anchor head failure.
- Tendons rupture.

The SHM features of interest here are the natural frequencies of the bridge, which were tracked over the period of a year and additionally over the period where the bridge was damaged according to the various scenarios. Apart from the almost-unique opportunity provided by the test to extract real damage-state data, the fact that the monitoring extended over an entire year meant that a complete picture of annual variations in the confounding influences were obtained. Modal properties of the bridge were extracted from acceleration data [9]. Figure 2 shows a time history of each of the four natural frequencies between 0-12 Hz of the bridge. The solid vertical line marks the start of the period where the different levels of damage (starting with pier lowering) were introduced. Gaps where the monitoring system failed have been removed. On inspection of Figure 2, the natural frequencies of the bridge are by no means stationary. There are some large fluctuations in the first half of the time history before the introduction of any damage. These fluctuations occurred during periods of very cold temperatures and have been associated with an increase in stiffness caused by stiffening of the asphalt layer on the bridge deck. The natural frequency time histories are an excellent example of how damage-sensitive parameters can also be very sensitive to environmental variations, in this case temperature. For the purpose of the current paper, it is important to note that the effect of temperature on the bridge is *nonlinear* – in fact, piecewise-linear. Figure 3 plots how the first natural frequency changes with temperature); the bilinear form of the temperature dependence means that switching models should prove useful. For the purposes of this case study, the second natural frequency will be discussed, as the first frequency is rather insensitive to damage.

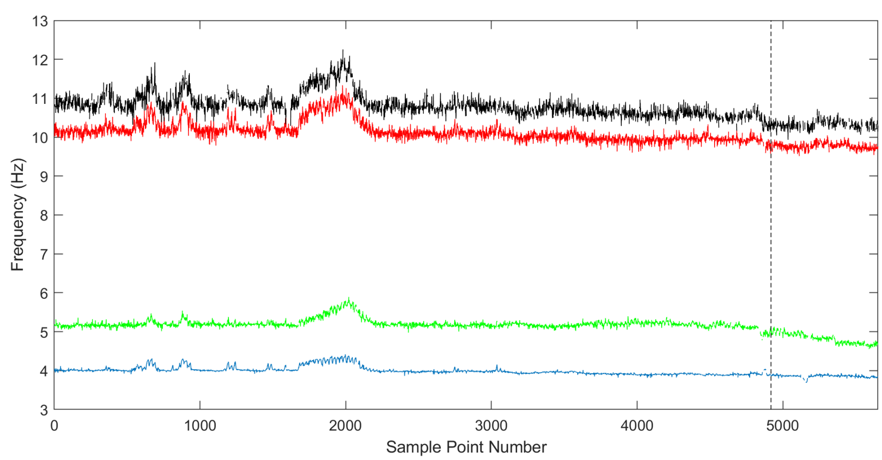


Figure 2. First four natural frequencies of the Z24 Bridge (increasing vertically). Note that the second natural frequency is the most sensitive to damage (right of the dotted line), but also varies significantly with temperature in the undamaged state (left of the dotted line).

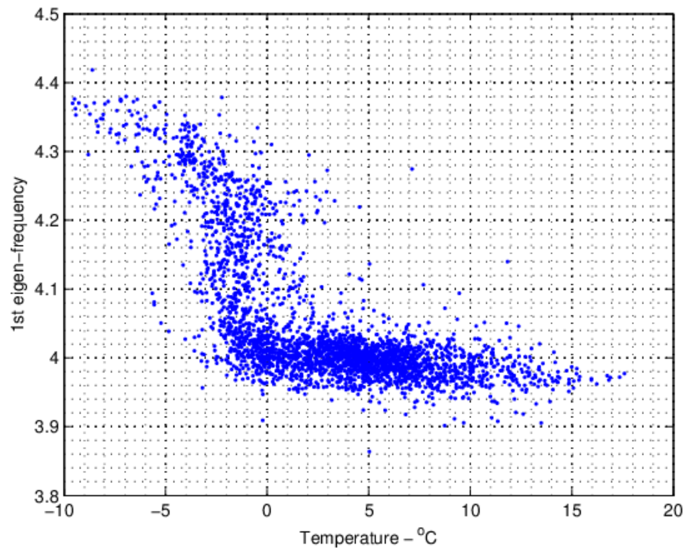


Figure 3. Detail of variation of second natural frequency with temperature.

## RESULTS

The results here are all based on selecting the second natural frequency $f_2$ data from the Z24 Bridge, and regressing them on the others $\{f_1, f_3, f_4\}$. In explicit terms, the model is thus,

$$f_{2i} = g(f_{1i}, f_{f3}, f_{4i}) + \epsilon \quad (1)$$

with the *cointegrated residual* $\epsilon$ thus,

$$\epsilon = f_{2i} - g(f_{1i}, f_{f3}, f_{4i}) \quad (2)$$

where $i$ is the sampling index of the data points.

If the process is succesful, the residual will be purged of the influence of temperature and will then only show novelty if damage is present. With this in mind, the idea will be to construct a control chart for $\epsilon$ by plotting it together with some confidence interval. Damage will then be signalled if the control variable departs the confidence interval. A significant advantage of the Bayesian approach followed here is that confidence intervals are produced naturally. As a baseline, the first results presented here will be for a treed linear model; this will make the overall model piecewise-linear (PWL). In fact, as the nature of the temperature variation is PWL, one would expect the treed-linear model to be both effective and parsimonious. In any machine learning problem, it is important to select appropriate training data; in this case, one requires data over which there is significant variation in temperature, but no effect of damage. When the algorithm is applied, the reults on the training data are shown in Figure 4.

The fit to the training data is excellent. A number of predictions are possible, the 'kriging mean' is shown here as it is arguably the simplest – representing the usual posterior mean of the Bayesian predictive distribution. As one can see, the predictions follow the measured data closely with the true values within the 99% confidence interval.

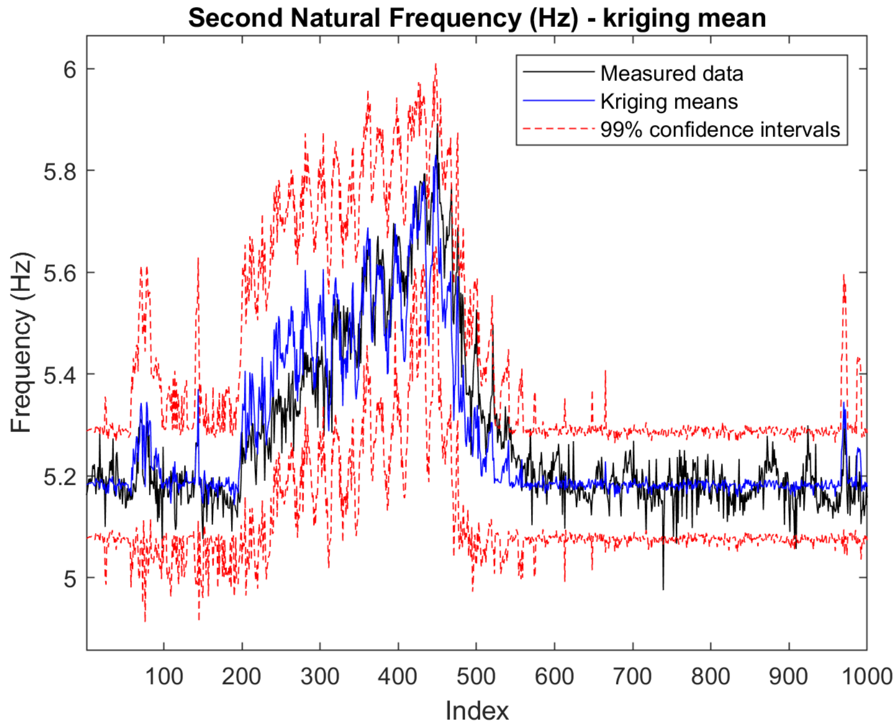


Figure 4. Cointegrated residual for a Bayesian treed linear model: training data.

Figure 5 shows the results over the test data, which in this case cover all the normal condition data immediately following the training data and the damage-state data.

The results are excellent; the residual stays within bounds throughout all the normal period and departs as required when the damage-states arise at about point 1400. The model is thus an effective novelty detector, purged of the confounding influence of temperature. As mentioned earlier, the model also provides physical insight in that it gives an estimate of the point at which the natural frequencies induce the switch in the model. This information is available in the tree which is estimated as the most probable (*maximum a postiori* (MAP)); the result here is shown in Figure 6.

The next model fitted was a full treed GP. In fact, Gramacy's algorithm has a useful point of sophistication. During the analysis, when a GP is fitted on a specific leaf of the tree, the algorithm can compare the Bayesian model evidence for a GP to that for a linear model; if the evidence supports a linear model, it is accepted [14]. The results on the training data are actually visually indistinguishable from those for the treed linear model, so they are not shown here. In contrast, the results on the testing data (Figure 7), are a little different.

In fact, the results are not quite as good as for the linear model; the confidence intervals expand a little more towards the end of the dataset and reduce the sensitivity to damage. The damage is detected later at a higher level of severity. This behaviour is not completely unexpected as the temperature dependence is known to be piecewise-linear. Occam's razor would appear to select the linear model. A slightly surprising outcome of fitting the TGP is that it provides a much simpler MAP tree (Figure 8), the model switches only once, on the first natural frequency ('x1' in the figure represents $f_1$). The

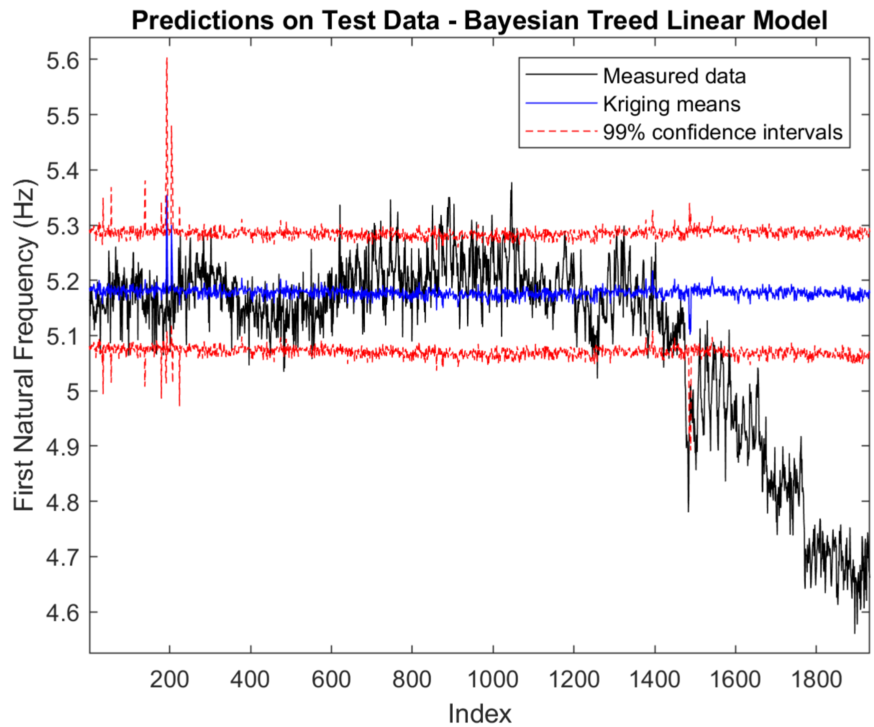


Figure 5. Cointegrated residual for a Bayesian treed linear model: test data.

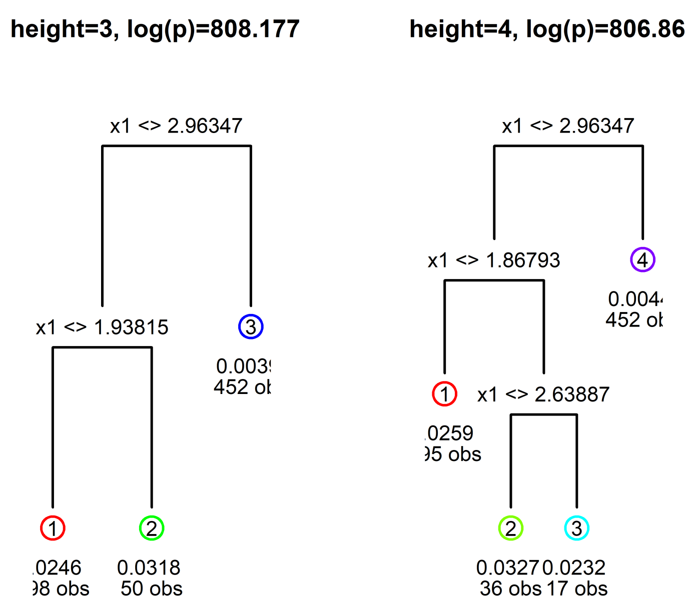


Figure 6. MAP trees for the Bayesian treed linear model: the two most probable are shown.

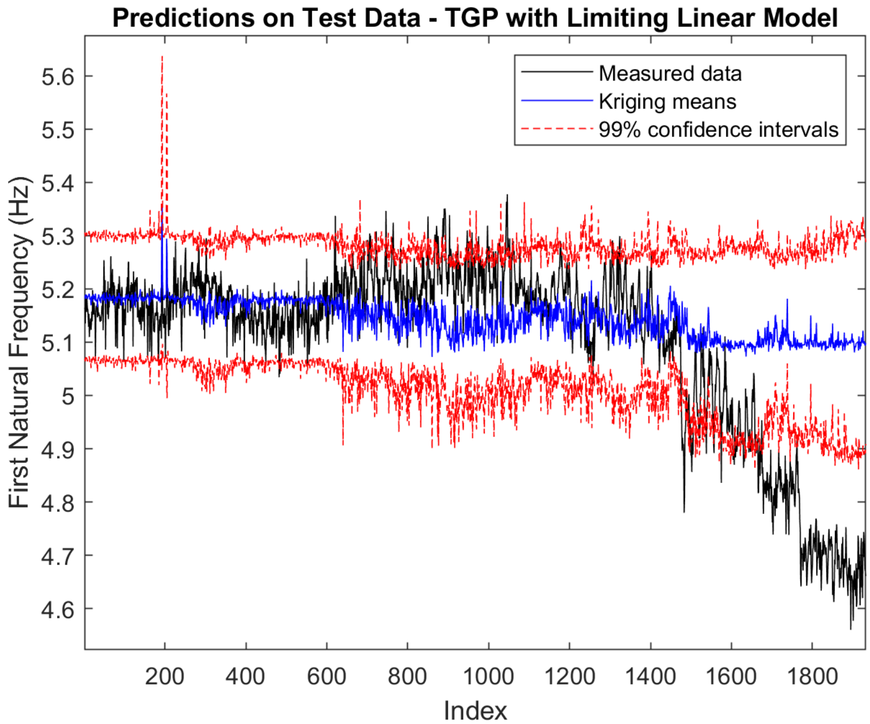


Figure 7. Cointegrated residual for a full TGP model (with limiting linear models): test data.

reason for the simpler tree is not clear, but could be a result of more complex behaviour immediately at the switching point.

## CONCLUSIONS

The aim of this paper is rather modest; in some ways, it could be regarded as 'completism'. Once one adopts an Engle-Granger approach to nonlinear cointegration, one could demonstrate results for *any* appropriately-powerful nonlinear regressor. However, the authors believe that some aspects of the model class here distinguish it from other nonparametric machine learners. In the first case, the TGPs can fit piecewise-smooth functions; this makes them optimal for any functions which do truly switch between physical regimes. Secondly, the estimated switching points can provide physical insight which would be absent if a smooth regressor were adopted. This insight is arguably somewhat indirect for the case study here, as the model switches on natural frequencies; however, with a little post-analysis effort, this can be converted to the point at which the model switches in terms of temperature.

## ACKNOWLEDGEMENTS

The authors of this paper gratefully acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC) via grant reference EP/W005816/1. For the purpose of open access, the authors has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

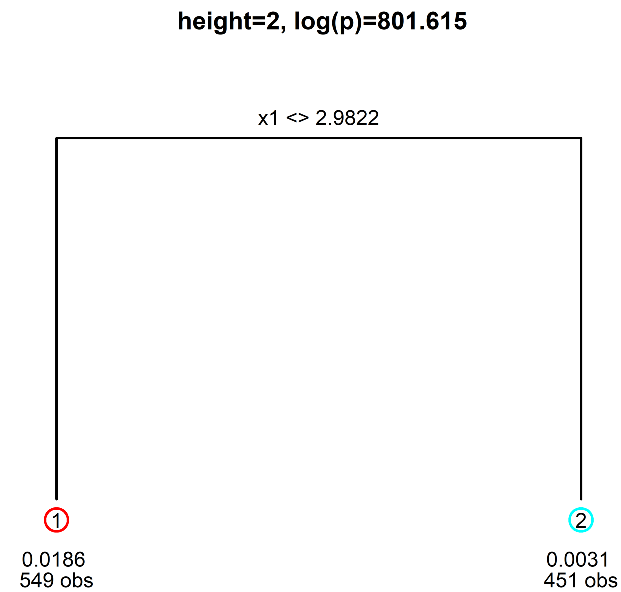


Figure 8. MAP tree for the full TGP model with limiting linear models.

## REFERENCES


1. Farrar, C. and K. Worden. 2013. *Structural Health Monitoring: A Machine Learning Perspective*, John Wiley and Sons.
2. Sohn, H. 2007. "Effects of environmental and operational variability on structural health monitoring," *Philosophical Transactions of the Royal Society - Series A*, 365:539–561.
3. Worden, K., T. Baldacchino, J. Rowson, and E. Cross. 2016. "Some recent developments in SHM based on nonstationary time series analysis," *Proceedings of the IEEE*, 104:1589–1603.
4. Perman, R. 1993. "Cointegration: an introduction to the literature," *Journal of Economic Studies*, 18: .
5. Cross, E., K. Worden, and Q. Chen. 2011. "Cointegration: a novel approach for the removal of environmental trends in structural health monitoring data," *Proceedings of the Royal Society - Series A*, 467:2712–2732.
6. Cross, E., G. Manson, K. Worden, and S. Pierce. 2012. "Features for damage detection with insensitivity to environmental and operational variations," *Proceedings of the Royal Society - Series A*, 468:4098–4122.
7. Johansen, S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.
8. Cross, E. and K. Worden. 2012. "Cointegration and why it works for SHM," in *Proceedings of $8^{th}$ International Conference on Modern Practice in Stress and Vibration, Glasgow, UK*.
9. Peeters, B. and G. De Roeck. 2001. "One-year monitoring of the Z24-Bridge: environmental effects versus damage events," *Earthquake Engineering and Structural Dynamics*, 30:149–171.

10. Worden, K. and E. Cross. 2018. “On switching response surface models, with applications to the structural health monitoring of bridges,” *Mechanical Systems and Signal Processing*, 98:139–156.
11. Breiman, L., J. Friedman, C. Stone, and R. Olsen. 1984. *Classification and Regression Trees*, Chapman and Hall/CRC.
12. Chipman, H., E. George, and R. McCulloch. 1998. “Bayesian CART model search,” *Journal of the American Statistical Association*, 93:935–948.
13. Chipman, H., E. George, and R. McCulloch. 2002. “Bayesian treed models,” *Machine Learning*, 48:299–320.
14. Gramacy, R. 2008. “Gaussian processes and limiting linear models,” *Computational Statistical Data Analysis*, 53:123–136.
15. Gramacy, R. 2007. “tgp: An R package for Bayesian nonstationary, semiparametric nonlinear regression and design by treed Gaussian process models,” *Journal of Statistical Software*, 19.
16. De Roeck, G. 2003. “The state-of-the-art of damage detection by vibration monitoring: the SIMCES experience,” *Journal of Structural Control*, 10:127–134.
17. Yan, A.-M., G. Kerschen, P. D. Boe, and J.-C. Golinval. 2005. “Structural damage diagnosis under varying environmental conditiona: Part II: local PCA for non-linear cases,” *Mechanical Systems and Signal Processing*, 17:865–880.