

Scale space multiresolution correlation analysis for time series data

L. Pasanen and L. Holmström

Department of Mathematical Sciences

University of Oulu, Finland

Abstract

We propose a new scale space method for the discovery of structure in the correlation between two time series. The method considers the possibility that correlation may not be constant in time and that it might have different features when viewed at different time scales. The time series are first decomposed into additive components corresponding to their features in different time scales. Temporal changes in correlation between pairs of such components are then explored by using weighted correlation within a sliding time window of varying length. Bayesian, sampling-based inference is used to establish the credibility of the correlation structures thus found and the results of analyses are summarized in scale space feature maps. The performance of the method is demonstrated using one artificial and two real data sets. The results underline the usefulness of the scale space approach when the correlation between the time series exhibit time-varying structure in different scales.

Keywords: time-varying correlation, time series decomposition, Bayesian inference, visualization

1 Introduction

This article is concerned with correlation between two time series in different time scales. The idea of correlated random variables dates back at least to Gauss who considered the normal surface of several correlated variates in 1823 (Rodgers and Nicewander, 1988). Gauss did not have any particular interest in a theoretical concept of correlation per se and the most widely used measure of correlation, the correlation coefficient that characterizes the linear dependency between two random variables was introduced by Karl Pearson only in 1895 (Rodgers and Nicewander, 1988). However, if the Pearson correlation coefficient is used to characterize the dependence between two time series, at least four problems need to be addressed.

The first problem is that if both time series contain a temporal trend, they might be deemed correlated even though they are not related in any way, a problem considered already in Yule (1926). One solution is to de-trend the time series before correlation analysis. One could subtract fitted linear trends from the two time series or difference them before the correlation coefficient is calculated. A different solution is offered by the co-integration methodology developed in the 1980s (Granger, 1981; Engle and Granger, 1987). It has become an important tool in particular in econometric analyses. The co-integration approach allows one to establish a linear relationship between time series that may include trends. For related earlier work, see Box and Tiao (1977). As examples of more recent developments, Podobnik and Stanley (2008) proposed an approach called de-trended cross-correlation analysis for determining the correlation between two non-stationary time series and Erdem et al. (2014) proposed a new correlation coefficient that takes into account the nonstationarity in the data.

The second problem is that correlation might change in time, in which case the Pearson correlation coefficient only reflects average correlation over the length of the time series. One solution is to use weighted correlation where, for each time point t , correlation is computed by weighting the time series values at times close to t more than the values far from it. Such an approach has been used for example in Pozzi et al. (2012) and Papadimitriou et al. (2006). We refer to such weighted correlation as *local correlation*, the function that determines the weights at the time points as the *kernel*, and the effective width of the kernel as the *window width*. The window width can be regarded as the temporal scale of the correlation.

The third problem is related to such a time window-related concept of scale and arises in a situation

where the data sets contain different correlation structures when viewed in different time horizons. An example of such a pair of random variables is money supply and interest rates (cf. Cochrane (1989)). In the short run they are considered to be negatively correlated while in the long run, because of inflation, the correlation is positive. The Pearson correlation coefficient cannot reveal the full dependency structure of such time series. In Cochrane (1989) the correlation structures were extracted in a specific scale by first applying a band-pass filter to the data. In this way the small scale features in the data do not hamper the detection of the large scale features, and vice versa. The use of a band-pass filter also reduces the effects of temporal long-term trends.

Another possible approach for the problem of scale-dependent structure in the data is to apply local correlation in several time scales. Local correlation for short and long window widths can reveal short and long range dependence structures. Therefore, a natural idea for finding structure in several scales is to consider local correlation for a range of window widths. Such an approach was taken by Nikolić et al. (2012) who considered time dependent cross-correlation in several scales. They summarized the results using a set of correlation plots with time on the horizontal axis, the time lag considered on the vertical axis, and the different scales presented in different plots.

The fourth problem is the difficulty of making valid statistical inference about the correlation. First of all, time series data usually contain autocorrelation. Therefore, the usual hypothesis test formulas for correlation that assume independent time points do not hold. Mudelsee proposed bootstrap based confidence intervals for serially dependent time series with an unknown distribution (Mudelsee, 2003). Later Ólafsdóttir and Mudelsee proposed a more accurate bootstrap based method (Ólafsdóttir and Mudelsee, 2014). Second, if local correlations are used, multiple tests are performed which then must be taken into account when inferences are made. One approach to the multiple testing problem is to apply the false discovery rate method (Benjamini and Hochberg, 1995).

To tackle the above four problems, we propose a new approach for discovering correlation structure in time series data that we call *scale space multiresolution correlation analysis*. The method consists of two steps. As in Cochrane (1989), the time series are first decomposed into a number of scale-dependent components. The decomposition method applied here is scale space multiresolution analysis first proposed for images in Holmström et al. (2011) and later for time series data in Pasanen et al. (2013). As the correlation structure between the extracted component pairs may well change in time, local correlation is used. Further, since the window widths corresponding to the underlying salient correlation structures are not known, just as in Nikolić et al. (2012), in the

second step of the method several window widths are applied in the correlation analysis.

In addition to exploring relevant correlation structure in the data, the method also identifies the time intervals and the time scales for which correlation is credibly positive or negative. Here Bayesian analysis is used because of the flexibility it offers in modeling and inference. For example, autocorrelation in the series can easily be accounted for and the multiple testing problem can be solved by finding sets of time points for which the joint posterior probability of the correlations being positive or negative is sufficiently high.

Our approach can be viewed as an instance of scale space methodology, first proposed in the computer vision literature (Witkin, 1983, 1984; Lindeberg, 1994). In scale space analysis a signal is viewed at several smoothing levels, each level considered to provide relevant information about the object of interest at a particular scale. The idea was introduced to statistics by Chaudhuri and Marron (1999) and subsequently extended into various directions (cf. Holmström (2010); Holmström and Pasanen (2016)).

The scale space interpretation of the proposed method is as follows. In scale space analysis, the small scales provide detailed information about the object of interest while the larger scales provide description in a coarser level. Similarly here, for each time series component pair the small window widths provide detailed information about the correlation over short time ranges, while analyses based on larger windows tell about overall correlation in longer time horizons. In the spirit of scale space analysis, the results are visualized using maps that display the strength of local correlation and summarize statistical inference.

When analyzing correlation between two time series, it is sometimes of interest to consider the possibility that the data exhibit dependence after a particular time shift is applied. Unlike in Nikolić et al. (2012), we do not include a time lag as a map dimension, but propose instead to construct the correlation and credibility maps for a range of time shifts and then examine the results for signs of lagged correlation. Even when traditional cross-correlation analysis is used to detect the lags that exhibit the strongest correlation, the scale space correlation maps can be used to add valuable information about how the correlation changes over time and at which scales correlation occurs.

Godtliebsen et al. (2012) applied the scale space approach for detecting differences between the slopes of two time series. If the data sets are normalized, difference of slopes can reveal the same kind of

features as local correlations. However, the fundamental difference between the two approaches is that while the difference of slopes approach works only if the signals are relatively smooth, the local correlation approach can discover also relations between signals that are not necessarily smooth.

Completely different approaches to local analysis of correlation structures in several time scales are the cross wavelet transform and wavelet coherence (Grinsted et al., 2004). Cross wavelet analysis reveals the common power and relative phase in the time-frequency space. Wavelet coherence, on the other hand, finds the degree of coherence between two time series even if their common power is low. Wavelet coherence can be thought of as a localized correlation coefficient in the time-frequency space. We will therefore compare the results obtained with scale space multiresolution correlation analysis to wavelet coherence.

The rest of the paper is organized as follows. In Section 2 the new method is described in detail. Local correlation analysis is explained in Section 2.1 and the multiresolution decomposition is discussed in Section 2.2. The use of Bayesian inference for credibility analysis is described in Section 2.3. In Section 3, the feasibility of the method is demonstrated using one artificial and one real data set and the analysis of lead-lag effects is considered. In Section 3.4 we compare our results to the results of wavelet coherence. Finally, the results are summarized in Section 4. In Online Resource 1, an analysis of a second real data set is provided.

2 Scale space correlation analysis

2.1 Local correlation in different time scales

Consider two time series, $\mathbf{x} = [x_1, \dots, x_n]^T$ and $\mathbf{y} = [y_1, \dots, y_n]^T$, observed at the same time points $t_1 < \dots < t_n$. The times t_k need not be equispaced. The temporal changes in the correlation between \mathbf{x} and \mathbf{y} can be captured by considering their local correlation at different times t_k , weighing the observations near t_k more than the ones far from it.

Let $\mathbf{w}_k = [w_{1k}, \dots, w_{nk}]^T$ be the weight vector for time t_k and denote by $\mathbf{W} = [w_{ik}]$ the $n \times n$

matrix of all weights. We assume that weights are normalized,

$$\sum_{i=1}^n w_{ik} = 1 \quad (1)$$

for each k . The local mean and variance of the time series \mathbf{x} at time t_k are defined as

$$\begin{aligned} \mathbf{E}_{\mathbf{W}}(\mathbf{x}; k) &= \mathbf{w}_k^T \mathbf{x}, \\ \text{Var}_{\mathbf{W}}(\mathbf{x}; k) &= \sum_{i=1}^n w_{ik} [x_i - \mathbf{E}_{\mathbf{W}}(\mathbf{x}; k)]^2. \end{aligned}$$

Local covariance and correlation between \mathbf{x} and \mathbf{y} at time t_k are defined analogously as

$$\begin{aligned} \text{Cov}_{\mathbf{W}}(\mathbf{x}, \mathbf{y}; k) &= \sum_{i=1}^n w_{ik} [x_i - \mathbf{E}_{\mathbf{W}}(\mathbf{x}; k)] [y_i - \mathbf{E}_{\mathbf{W}}(\mathbf{y}; k)], \\ \text{Corr}_{\mathbf{W}}(\mathbf{x}, \mathbf{y}; k) &= \frac{\text{Cov}_{\mathbf{W}}(\mathbf{x}, \mathbf{y}; k)}{\sqrt{\text{Var}_{\mathbf{W}}(\mathbf{x}; k)} \sqrt{\text{Var}_{\mathbf{W}}(\mathbf{y}; k)}}. \end{aligned} \quad (2)$$

The weight vectors can be chosen in many ways. A weight vector \mathbf{w}_k which is constant for indices close to k and zero elsewhere weighs uniformly observations in a neighborhood of t_k . This turns out to produce noisy results that are hard to interpret and we therefore use weights obtained from the so-called biweight kernel function,

$$w_{ik} \propto \begin{cases} \left[1 - \left(\frac{t_i - t_k}{\sigma}\right)^2\right]^2, & \text{if } \left|\frac{t_i - t_k}{\sigma}\right| \leq 1, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where $\sigma > 0$. The weight vector \mathbf{w}_k is scaled separately for each k to satisfy the condition (1).

The parameter σ in the weight vector determines the time horizon considered in the local correlation of \mathbf{x} and \mathbf{y} . As different time spans may reveal different structures, a range of values of σ is considered in a scale space fashion to find the salient structures in different scales. To illustrate how the weights depend on k and σ , Figure 1 shows plots of \mathbf{w}_k for three values of k and four values of σ . The time grid is the same as in the artificial data example of Section 3.1, that is, 900 equispaced points between 0 and 30. Note that since the weights in \mathbf{w}_k are scaled to sum to 1, the values of w_{1k} and w_{nk} are larger than $w_{n/2,k}$.

The results of multi-scale local correlation analysis are visualized using color maps where the hor-

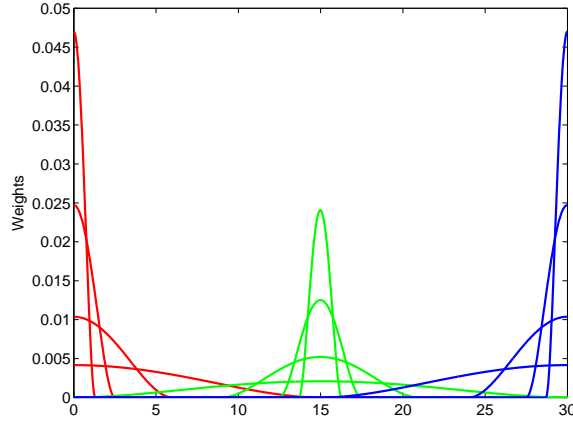


Figure 1: The weight vectors \mathbf{w}_k for three time points and four different values of the kernel scale parameter σ for the artificial example of Section 3.1. The three times are 0, 15 and 30 and the values of σ are 1.3, 2.5, 6 and 15. The weight vectors for time 0, 15 and 30 are colored red, green, and blue, respectively

horizontal axis represents the time and the vertical axis represents $\log_{10}(\sigma)$. The color of a pixel at $(t_k, \log_{10}(\sigma))$ represents the local correlation at time point t_k , where the degree of localness is determined by σ ; an example is shown in the lower panel of Figure 3. From now on such an image is called a *correlation map*.

2.2 Scale space multiresolution analysis

In the local correlation analysis described in the previous section, strong large scale dependence between two time series can make it difficult to detect small scale correlation structure, and vice versa. Therefore, before correlation analysis is carried out, the time series are decomposed into distinct scale-dependent components and correlation analysis is then performed on pairs of components corresponding to similar scales. The underlying assumption here is that the two time series contain interesting features in similar time scales and that our main goal is to discover the correlations of such features.

The decomposition uses the scale space multiresolution technique first proposed for images in Holmström et al. (2011) and later for time series in Pasanen et al. (2013). For the sake of completeness, we give here an outline of the method of Pasanen et al. (2013).

Denote by \mathbf{S}_λ a smoothing matrix corresponding to a smoothing level $\lambda \geq 0$. The scale-dependent components of time series \mathbf{x} are defined by differences $\mathbf{S}_{\lambda_i}\mathbf{x} - \mathbf{S}_{\lambda_{i+1}}\mathbf{x}$ of smooths, where the smoothing parameter values λ_i are chosen appropriately. The particular smoothing method used in Pasanen et al. (2013) is the discrete spline smoother

$$\mathbf{S}_\lambda = (\mathbf{I} + \lambda \mathbf{C}^T \mathbf{C})^{-1},$$

where $\mathbf{C} \in \mathbb{R}^{(n-2) \times n}$ defines second order differencing, that is, for $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{C}\mathbf{a} = \mathbf{b}$ with

$$b_i = \frac{a_{i+2} - a_{i+1}}{t_{i+2} - t_{i+1}} - \frac{a_{i+1} - a_i}{t_{i+1} - t_i}, \quad i = 1, \dots, n-2. \quad (4)$$

Note that, as $\lambda \rightarrow \infty$, $\mathbf{S}_\lambda \mathbf{x}$ converges to the linear regression line of the data (t_i, x_i) .

Let

$$0 = \lambda_1 < \lambda_2 < \dots < \lambda_{L-1} < \lambda_L \leq \infty. \quad (5)$$

Since $\mathbf{S}_{\lambda_1} \mathbf{x} = \mathbf{x}$, the time series \mathbf{x} can be decomposed into scale-dependent components as

$$\mathbf{x} = \sum_{j=1}^{L-1} (\mathbf{S}_{\lambda_j} - \mathbf{S}_{\lambda_{j+1}}) \mathbf{x} + \mathbf{S}_{\lambda_L} \mathbf{x} \equiv \sum_{j=1}^L \mathbf{u}_j, \quad (6)$$

where $\mathbf{u}_j = (\mathbf{S}_{\lambda_j} - \mathbf{S}_{\lambda_{j+1}}) \mathbf{x}$ for $j = 1, \dots, L-1$, and $\mathbf{u}_L = \mathbf{S}_{\lambda_L} \mathbf{x}$. The component \mathbf{u}_j can be interpreted as the structure of \mathbf{x} which is smoothed out when smoothing is increased from λ_j to λ_{j+1} . If $\lambda_L = \infty$, \mathbf{u}_L corresponds to the linear trend.

A good separation of the scale-dependent components in the decomposition (6) requires that the sequence (5) of smoothing levels is chosen carefully. While a trial and error approach could be used, we prefer the scale-derivative method suggested in Pasanen et al. (2013). The scale-derivative of \mathbf{x} is defined as

$$\mathbf{D}_\lambda \mathbf{x} = \lim_{\lambda' \rightarrow \lambda} \frac{\mathbf{S}_{\lambda'} \mathbf{x} - \mathbf{S}_\lambda \mathbf{x}}{\log \lambda' - \log \lambda} = -\lambda (\mathbf{I} + \lambda \mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{C} (\mathbf{I} + \lambda \mathbf{C}^T \mathbf{C})^{-1} \mathbf{x}.$$

In Pasanen et al. (2013) the so-called scale-derivative map was introduced to visualize $\mathbf{D}_\lambda \mathbf{x}$. It is a plot with time on the horizontal axis, $\log_{10}(\lambda)$ on the vertical axis, and the color of the pixels reflecting the value of the scale-derivative. Figure 4 displays examples of scale-derivative maps. The scale-dependent components of the signal \mathbf{x} show as oscillating bands of blue and red, suggesting

that the sequence of the smoothing levels λ_j in the multiresolution decomposition (6) can be chosen as the local minima of the function $\lambda \mapsto \|\mathbf{D}_\lambda \mathbf{x}\|$ (cf. Pasanen et al. (2013)).

Given two time series \mathbf{x} and \mathbf{y} , we want to use the same sequence of multiresolution levels λ_j for both. A compromise is therefore needed between the levels suggested by the scale-derivatives of \mathbf{x} and \mathbf{y} . This could be achieved by visual inspection, selecting the smoothing parameter sequence (5) so that the λ_j 's are located as nearly as possible between oscillating bands of blue and red in the scale-derivative maps of both \mathbf{x} and \mathbf{y} . However, as an automatic, data-driven method we propose to determine the levels λ_j as the local minima of

$$\lambda \mapsto \frac{\|\mathbf{D}_\lambda \mathbf{x}\|}{\|\mathbf{x}\|} + \frac{\|\mathbf{D}_\lambda \mathbf{y}\|}{\|\mathbf{y}\|}. \quad (7)$$

Given the multiresolution decompositions

$$\mathbf{x} = \sum_{j=1}^L \mathbf{u}_j, \quad \mathbf{y} = \sum_{j=1}^L \mathbf{v}_j, \quad (8)$$

local correlation analysis with varying window widths, as described in the previous section, is then applied to each component pair $(\mathbf{u}_j, \mathbf{v}_j)$.

2.3 Feature credibility analysis using Bayesian inference

The observed time series are usually noisy and statistical inference is needed to decide whether the structure suggested by correlation analysis is credible or just an artifact related to random errors in the data. For each multiresolution component pair, Bayesian inference is applied for a range of window widths σ to identify the time intervals where their local correlation is credibly positive or negative.

We assume that the observed time series consist of a true underlying signal and additive noise,

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\epsilon}, \quad \mathbf{y} = \boldsymbol{\nu} + \boldsymbol{\delta}, \quad (9)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are the noiseless signals whose correlation structure we wish to infer. Thus, using

the same notation for the multiresolution components as before, we consider the decompositions

$$\boldsymbol{\mu} = \sum_{j=1}^L \mathbf{u}_j, \quad \boldsymbol{\nu} = \sum_{j=1}^L \mathbf{v}_j \quad (10)$$

of the underlying true signals and analyze for each j the local correlations between \mathbf{u}_j and \mathbf{v}_j . To simplify the notation a little, we denote by $C(\mathbf{u}_j, \mathbf{v}_j; k, \sigma)$ the local correlation between \mathbf{u}_j and \mathbf{v}_j at time t_k and scale σ (cf. (2) and (3) in Section 2.1).

Inference is based on posterior sampling and consists of three steps: 1) generating a sample from the posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, 2) generating samples from the posterior distributions of the local correlations $C(\mathbf{u}_j, \mathbf{v}_j, k, \sigma)$, and 3) using the generated samples, finding for each multiresolution scale pair and each window width σ the time intervals for which the correlation is credibly positive or negative.

As in Godtliebsen et al. (2012), we simplify posterior modeling by assuming that $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are independent *a priori*, and also that $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are independent. With these assumptions, a sample from the joint posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ can be generated by drawing independent samples from their marginal posterior distributions.

Next, for each σ considered in the analysis, we need the joint posterior distribution of the local correlation coefficients $C(\mathbf{u}_j, \mathbf{v}_j, k, \sigma)$, $k = 1, \dots, n$. The multiresolution levels λ_j are determined using the scale-derivatives of the posterior means of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Thus, instead of (7), the levels are determined as the local minima of

$$\lambda \mapsto \frac{\|\mathbf{D}_\lambda \mathbf{E}(\boldsymbol{\mu}|\mathbf{x})\|}{\|\mathbf{E}(\boldsymbol{\mu}|\mathbf{x})\|} + \frac{\|\mathbf{D}_\lambda \mathbf{E}(\boldsymbol{\nu}|\mathbf{y})\|}{\|\mathbf{E}(\boldsymbol{\nu}|\mathbf{y})\|}. \quad (11)$$

We first sample from the posterior distributions of the components \mathbf{u}_j and \mathbf{v}_j . The posterior samples of \mathbf{u}_j and \mathbf{v}_j are obtained by applying the difference of smooths transformation to the sampled values of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Then the samples of \mathbf{u}_j and \mathbf{v}_j are further transformed to get a sample from the joint posterior distribution of the correlation coefficients $C(\mathbf{u}_j, \mathbf{v}_j, k, \sigma)$.

Finally, the sample generated from the joint posterior distribution of the correlation coefficients $C(\mathbf{u}_j, \mathbf{v}_j, k, \sigma)$ is used to identify the time intervals and window widths for which correlation is credibly positive or negative. This could be done simply by identifying for each σ the times t_k for

which the marginal posterior probability of the correlation being positive or negative exceeds some threshold value $0 < \alpha < 1$. However, such a point-wise inference is bound to result in a large number false positives, and we therefore use simultaneous inference over all times t_k and a fixed σ . The simultaneous inference technique applied is the method of highest pointwise probabilities (HPW), first described in Erästö and Holmström (2005). The HPW is a greedy algorithm that selects time points according to their descending order of marginal posterior probability until the joint posterior probability of correlations being positive or negative is at least α . In all examples we have taken $\alpha = 0.95$. The results of inference are summarized in a credibility version of the correlation map where each pixel is colored either white, black or gray, depending on whether the correlation is credibly positive, negative or neither (see e.g. the bottom panels of Figure 8).

3 Experiments

3.1 An artificial example

We first illustrate the proposed methods using an artificial example where the observed noisy time series \mathbf{x} and \mathbf{y} consist of two sine waves, linear trends and Gaussian noise,

$$x_i = \mu_i + \varepsilon_i = \begin{cases} 40 \sin(2t_i) + 40 \sin(0.75t_i) - 1.5t_i + \varepsilon_i, & i \leq 450, \\ -40 \sin(2t_i) - 40 \sin(0.75t_i) - 1.5t_i + \varepsilon_i, & i > 450, \end{cases}$$

$$y_i = \nu_i + \delta_i = -40 \sin(2t_i) + 40 \sin(0.75t_i) + 1.5t_i + \delta_i,$$

where $t_i = 30(i-1)/(900-1)$, the additive noise is independent and identically distributed, $\varepsilon_i, \delta_i \sim N(0, 50^2)$, and $i = 1, \dots, 900$. The observed time series \mathbf{x} and \mathbf{y} and the noiseless underlying truths $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are shown in Figure 2. The phase difference between the sine waves changes in the middle of the time interval. In the left part of the data, the small-scale waves are in antiphase, and the large-scale waves are in phase. The opposite is true in the right half of the data. This causes the Pearson correlation between \mathbf{x} and \mathbf{y} to be very close to zero, approximately -0.06, although the component waves clearly are correlated for certain window lengths.

Let us then try to investigate the situation by considering local correlation of \mathbf{x} and \mathbf{y} for several window widths. The result is shown in Figure 3. The presence of the underlying correlation

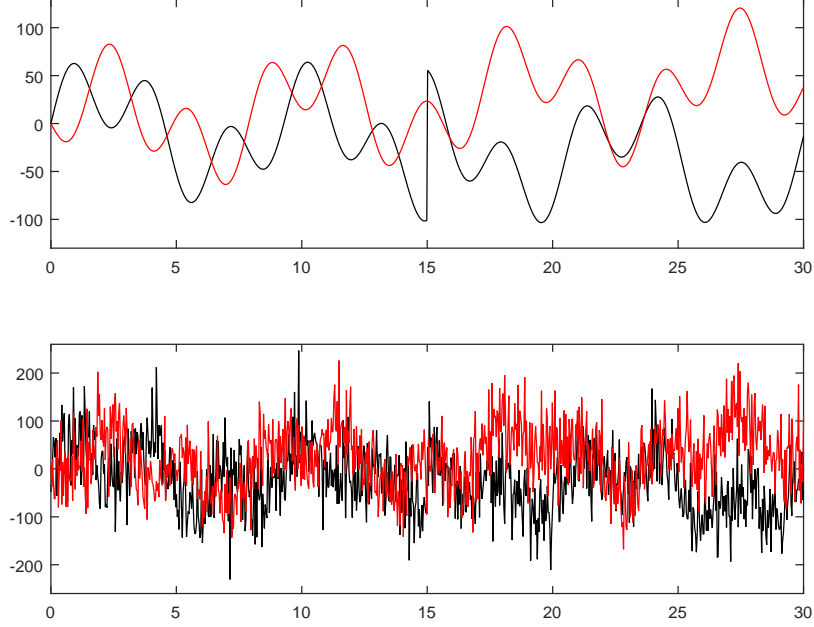


Figure 2: Upper panel: the noiseless time series. Lower panel: the noisy time series. Black: \mathbf{x} and $\boldsymbol{\mu}$. Red: \mathbf{y} and $\boldsymbol{\nu}$

structures remains unclear. To some extent, the correlation caused by the $\sin(2t_i)$ wave is detectable, but no other structure can be seen.

Thus, neither the Pearson correlation coefficient nor local multi-scale correlation analysis of the original time series \mathbf{x} and \mathbf{y} is able to properly capture the structure underlying the data. This emphasizes the importance of decomposing the time series \mathbf{x} and \mathbf{y} into components that reflect their features in distinct scales, before correlation analysis is performed. We use the decompositions (8) where the resolution levels λ_j are based on the scale-derivatives $\mathbf{D}_\lambda \mathbf{x}$ and $\mathbf{D}_\lambda \mathbf{y}$ visualized in the maps of Figure 4. The local minima of (7) are $\lambda_2 = 10^{0.085}$ and $\lambda_3 = 10^{2.57}$ and they are indicated by the black lines in the maps. The minima are located between oscillating bands of blue and red and hence should produce reasonable multiresolution decompositions. The sequence $[0, \lambda_2, \lambda_3, \infty]$ decomposes each of the original time series into $L = 4$ components (cf. (8)) and the correlation maps of the pairs $(\mathbf{u}_j, \mathbf{v}_j)$ are shown in Figure 5. Component 1, corresponding to the pair $(\mathbf{u}_1, \mathbf{v}_1)$, represents the noise and its correlation map is mostly close zero, except for very small local correlation window widths σ . The other components capture the salient correlation structure in the data. Component 4 reveals the negative correlation of the linear trends of the two time series and the underlying features related to the two sine waves are clearly visible in components 2 and

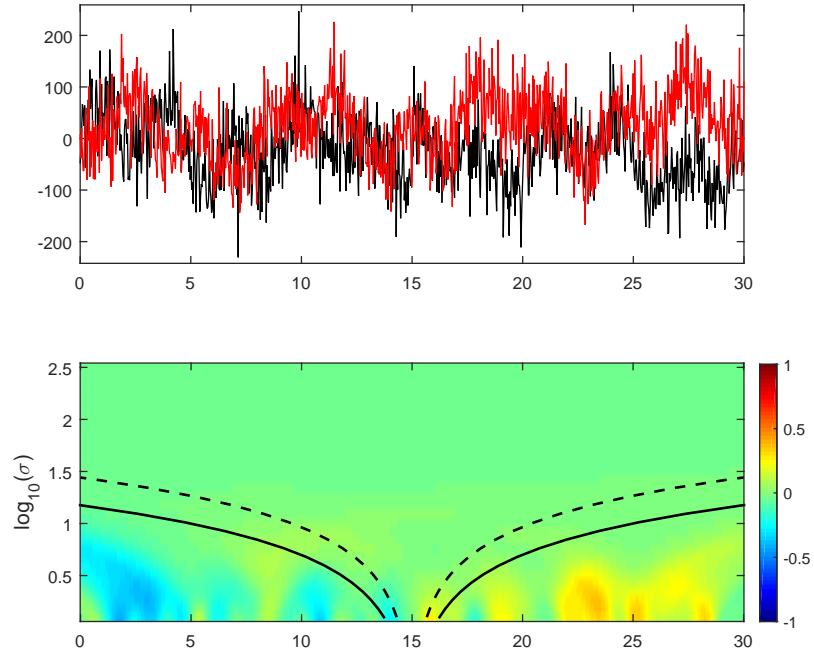


Figure 3: Correlation analysis of the artificial data set. Upper panel: the noisy time series. Black: \mathbf{x} ; Red: \mathbf{y} . Lower panel: the correlation map. The solid black lines indicate the width of a centrally positioned kernel and the dashed lines indicate the width of the interval where the kernel height has decreased to 50% of its maximum

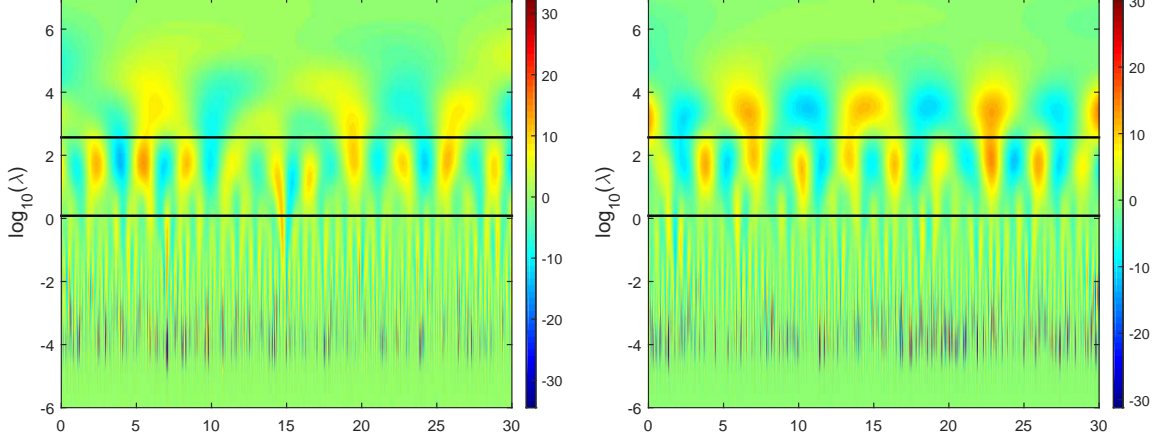


Figure 4: The scale-derivative maps of the time series \mathbf{x} and \mathbf{y} . The local minima of the function $\lambda \mapsto \|\mathbf{D}_\lambda \mathbf{x}\|/\|\mathbf{x}\| + \|\mathbf{D}_\lambda \mathbf{y}\|/\|\mathbf{y}\|$ are indicated by the black lines

3. In component 2, correlation is negative for the first half of the time series and positive for the second half, and the situation is reversed for component 3. For component 3, however, the lower left corner of the correlation map shows a short time interval where the correlation is negative. This is due to the fact that noise can have a strong effect on correlation computed for a very short time window. Except for component 4, the upper parts of the maps are green, indicating that the overall, Pearson-style correlation between the multiresolution components is close to zero. This again highlights the importance of measuring correlation in multiple time scales.

Finally, credibility of the suggested correlations is examined. We consider the model (9), where the additive noise is iid Gaussian and where the true underlying signals $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are assumed to be somewhat smoother than the observations \mathbf{x} and \mathbf{y} . The observations are assumed to be conditionally independent, given the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, which themselves are assumed to be independent, *a priori*. With these assumptions $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ will be independent also *a posteriori*. For \mathbf{x} , our basic time series model is

$$\begin{aligned} \mathbf{x}|\boldsymbol{\mu}, \sigma_x^2 &\sim \mathcal{N}(\boldsymbol{\mu}, \sigma_x^2 \mathbf{I}), \\ \sigma_x^2 &\sim \text{Inv-}\chi^2(\nu_{x0}, \sigma_{x0}^2), \\ p(\boldsymbol{\mu}|\kappa_x, \sigma_x^2) &\propto \left(\frac{\kappa_x}{\sigma_x^2}\right)^{\frac{n-2}{2}} \exp\left(-\frac{\kappa_x}{2\sigma_x^2} \|\mathbf{C}\boldsymbol{\mu}\|^2\right), \end{aligned}$$

where \mathbf{C} is defined in (4). The improper prior of $\boldsymbol{\mu}$ has a smoothing effect, the magnitude of which

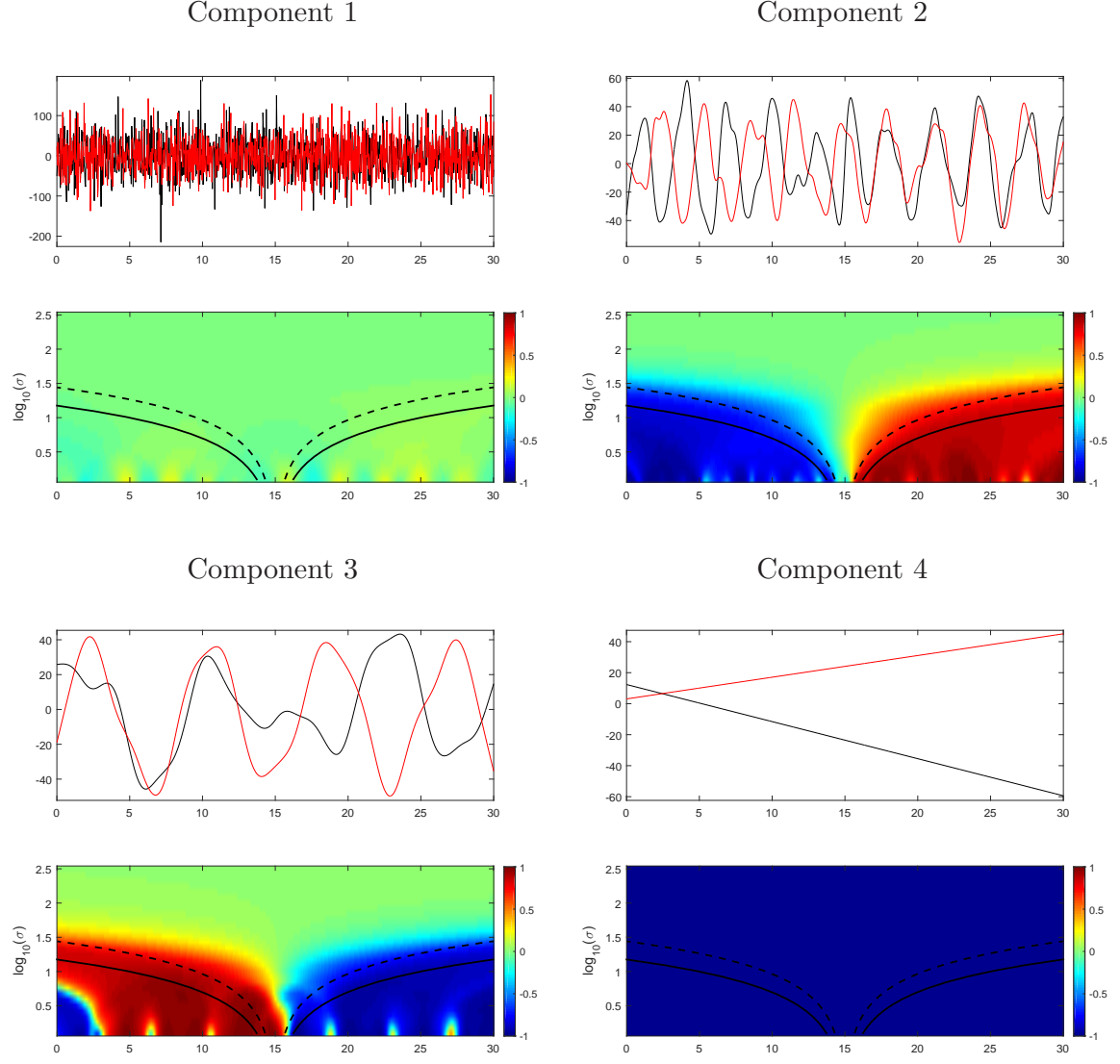


Figure 5: The correlation maps of the four multiresolution component pairs of the time series displayed in the lower panel of Figure 2. The upper panel of each component: the components \mathbf{u}_j and \mathbf{v}_j in the expansions (8). The lower panels: The correlation maps. For more information, see the caption of Figure 3

is controlled by the smoothing parameter κ_x . In this model, the posterior distribution of $\boldsymbol{\mu}$ is a multivariate-t,

$$\boldsymbol{\mu}|\mathbf{x} \sim t_{\nu_{x0}+n-2}(\mathbf{S}_{\kappa_x}\mathbf{x}, \boldsymbol{\Sigma}_{x0}),$$

where

$$\mathbf{S}_{\kappa_x}\mathbf{x} = (\mathbf{I} + \kappa_x \mathbf{C}^T \mathbf{C})^{-1} \mathbf{x}, \quad \boldsymbol{\Sigma}_{x0} = \left(\frac{\|\mathbf{x}\|^2 - \mathbf{x}^T \mathbf{S}_{\kappa_x} \mathbf{x} + \nu_{x0} \sigma_{x0}^2}{\nu_{x0} + n - 2} \right) \mathbf{S}_{\kappa_x}.$$

Further discussion and extensions of this model can be found in Erästö and Holmström (2005) and Pasanen et al. (2013). The model for the time series \mathbf{y} is defined similarly.

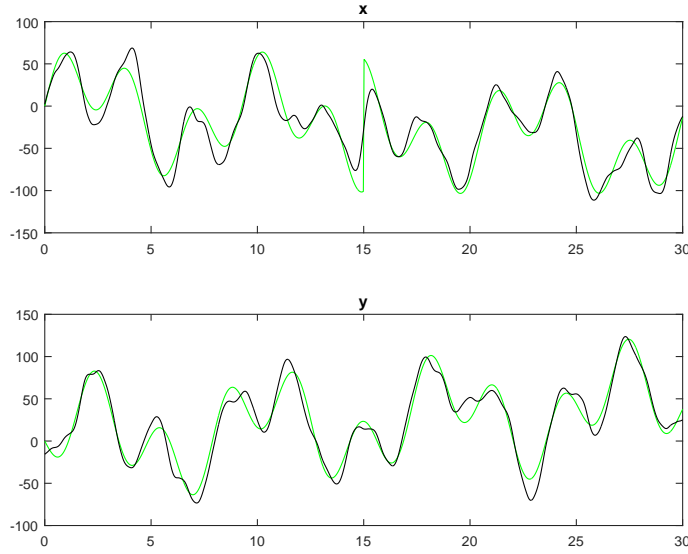


Figure 6: The posterior means and the true underlying signals for the artificial data displayed in Figure 2. Black: the posterior mean. Green: the noiseless signal

The results of inference will depend on the values of the parameters κ_x and κ_y , as well as on the choice of the hyperparameters σ_{x0}^2 , σ_{y0}^2 , ν_{x0} , and ν_{y0} , which all therefore need to be selected appropriately. We take $\nu_{x0} = \nu_{y0} = 10$, which makes the priors of σ_x^2 and σ_y^2 vague. As in Erästö and Holmström (2005), full Bayesian inference could be used by considering κ_x and κ_y as unknown parameters of the model. However, for computational efficiency we have here used an empirical Bayes approach and estimated κ_x and κ_y and the prior means of σ_x^2 and σ_y^2 from the data. The posterior mean of $\boldsymbol{\mu}$ is $E(\boldsymbol{\mu}|\mathbf{x}) = \mathbf{S}_{\kappa_x}\mathbf{x}$, and therefore a reasonable value for κ_x is one that produces a smooth \mathbf{x} which is close to $\boldsymbol{\mu}$. Such a smoothing level can be inferred from the scale-derivative map, since the smallest local minimum of $\lambda \mapsto \|\mathbf{D}_\lambda \mathbf{x}\|$ is expected to correspond to a smooth of \mathbf{x} in which the noise has been removed but the true signal is still largely intact. The value for κ_x obtained in this way is 1.10. The mean for the prior of σ_x^2 can then be estimated from the sample

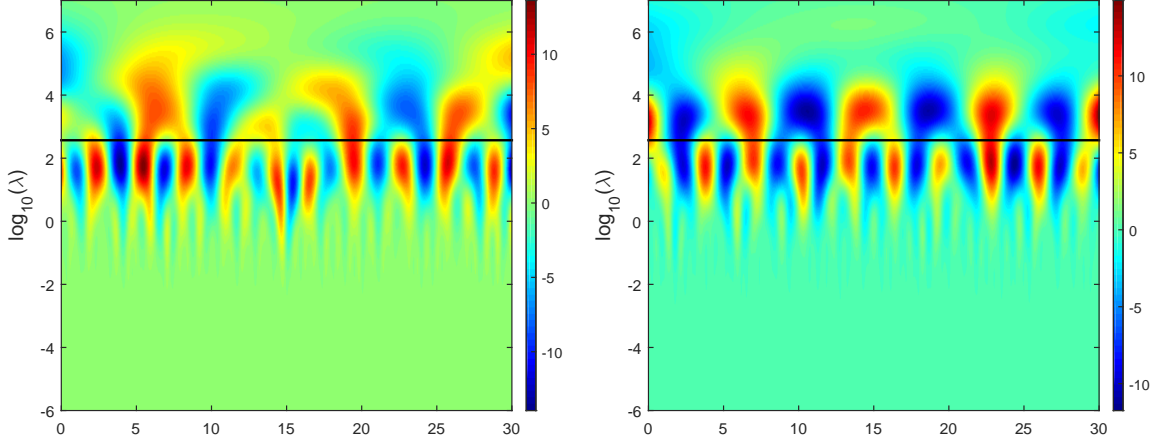


Figure 7: The scale-derivative maps of the posterior means of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. The local minima of (11) are denoted by black lines

variance of the residuals $\mathbf{x} - \mathbf{S}_{1.10}\mathbf{x}$ giving the value $E(\sigma_x^2) = 48.5^2$. Similar calculations are done for \mathbf{y} , giving $\kappa_y = 1.22$ and $E(\sigma_y^2) = 48.5^2$. Hence, $\sigma_{x0}^2 = \sigma_{y0}^2 = (8/10) \cdot 48.5^2$. Note that the estimates for the noise variance prior means are close to the true variance 50^2 . The sample size for posterior inference was 10^4 . Figure 6 shows that the posterior means $E(\boldsymbol{\mu}|\mathbf{x})$ and $E(\boldsymbol{\nu}|\mathbf{y})$ and the corresponding true signals $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are very similar.

The scale-derivative maps of the posterior means $E(\boldsymbol{\mu}|\mathbf{x})$ and $E(\boldsymbol{\nu}|\mathbf{y})$ are shown in Figure 7. The local minimum of (11) is $10^{2.57}$, giving the sequence $[0, 10^{2.57}, \infty]$ for scale space multiresolution decomposition. The posterior means of the correlation maps and the associated credibility maps for the two mid-scale multiresolution component pairs are displayed in Figure 8. In a correlation credibility map, white, black and gray indicate correlation which is credibly positive, negative or neither. The largest scale component shows credible negative correlation between the linear trends but to save space, the results of the analysis are not shown. As in the analysis of the raw observed time series \mathbf{x} and \mathbf{y} , the two displayed components capture the correlation structure caused by the sine waves. For the higher frequency oscillation, correlation is negative on the time interval $[0, 13]$ and positive on $[16, 30]$. For the lower frequency oscillation, correlation is positive on $[0, 14]$ and negative on $[17, 30]$. Note that the brief period of early negative correlation in the slower oscillation is not deemed credible, eliminating the artifact that appeared also in the correlation analysis of the raw observed data (cf. Figure 5). The short intervals of negative correlation in the midst of an area of positive correlation are not credible either, and the same holds for short intervals of positive correlation in the generally negative areas. The upper parts of the credibility maps are

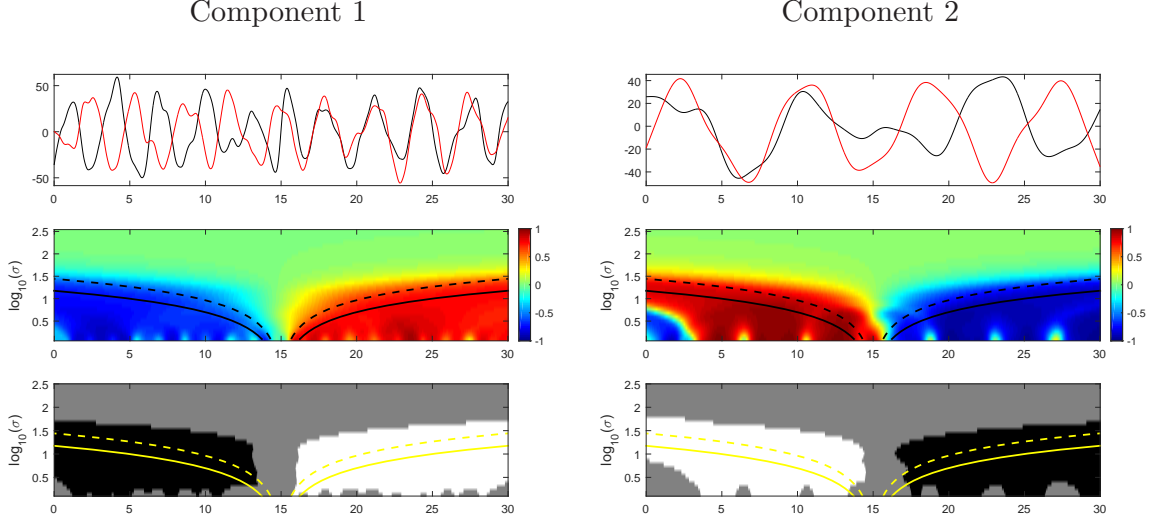


Figure 8: Posterior correlation analysis of the two mid-scale multiresolution component pairs $(\mathbf{u}_j, \mathbf{v}_j)$ in the decompositions (10) of the artificial data displayed in Figure 2. The upper panel of each component: the posterior means of the components. Black: μ ; Red: ν . The middle panels: The posterior means of the correlation maps. The lower panels: correlation credibility maps

gray suggesting that the overall correlation between the multiresolution components is not credibly different from zero.

3.2 The unemployment rate vs. the effective federal funds rate

As mentioned in the Introduction, Cochrane (1989) showed that in the case of interest rates and money supply, short term correlation is negative and long term correlation is positive. We consider here the effective U.S. federal funds rate and the unemployment rate (FEDFUNDS; UNRATE) and show their correlation structure to be similar. Correlation between these two time series is expected since, by affecting interest rates throughout financial markets, the federal funds rate has an effect on overall economic activity. The two time series are shown in Figure 9 and they indeed appear to exhibit both types of correlation. However, after 2000 also the large scale correlation might be negative.

The scale-derivative maps of the two time series are shown in Figure 10. The function (7) has a single local minimum at $\lambda_2 = 10^{4.2}$. The time series are therefore decomposed into three components using the multiresolution level sequence $[0, 10^{4.2}, \infty]$. The correlation maps for the two mid-scale

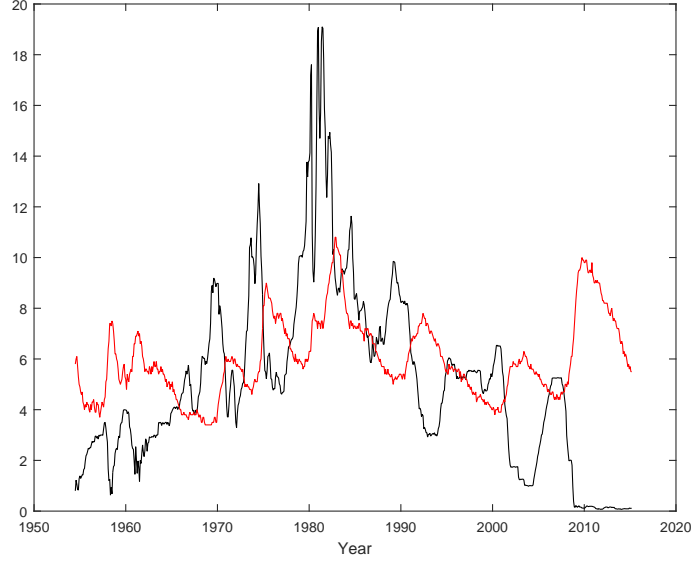


Figure 9: The U.S. economic times series analyzed for correlations. Black: the effective federal funds rate \mathbf{x} . Red: the unemployment rate \mathbf{y}

component pairs are shown in Figure 11. The linear trends in the data exhibit negative correlation (the results not shown). The presence of small-scale negative correlation is suggested by the mostly blue correlation map of component 1. However, around 1985 correlation is nearly zero for $\sigma = 10^1$, that is, when the local window length considered is roughly 20 years. The upper part of the correlation map of the larger scale component 2 is orange, suggesting a long-term positive correlation in the two time series. However, when the correlation of these components is viewed in shorter time scales, strong negative correlation appears both in 1954 -1968 and in the new millennium.

Next, we evaluate the credibility of the features suggested by the correlation maps. For the purpose of demonstration, the time series model adopted is identical to the one used for the artificial data in the previous section, although it might well be in this case too simplistic. The scale-derivative maps reveal no clear boundary between small scale structure that could be regarded as noise and other structure that could be considered to reflect the true underlying time series features. A different approach is therefore needed to estimate the values for the prior smoothing parameters κ_x and κ_y . As $E(\boldsymbol{\mu}|\mathbf{x}) = \mathbf{S}_{\kappa_x} \mathbf{x}$, we have that $\mathbf{x} = \mathbf{S}_{\kappa_x} \mathbf{x} + (\mathbf{x} - \mathbf{S}_{\kappa_x} \mathbf{x}) = E(\boldsymbol{\mu}|\mathbf{x}) + \mathbf{e}$. Since the errors in the time series model are assumed to be independent, a reasonable value for κ_x is one for which the residuals \mathbf{e} seem independent. The criterion for independence applied here is that the first autocorrelation of the residuals is close to zero. The values of the model parameters thus obtained are $\kappa_x = 0.0067$,

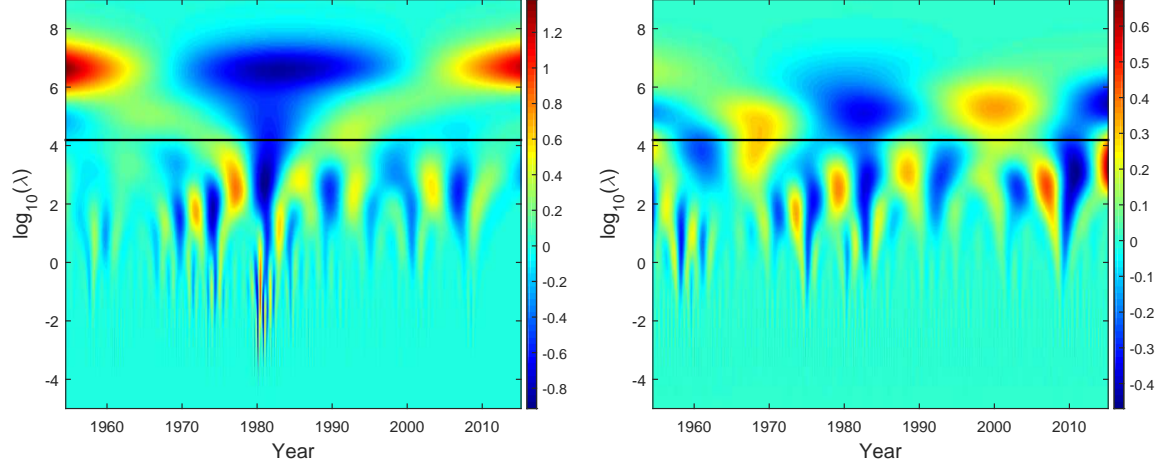


Figure 10: The scale-derivative maps of the effective federal funds rate \mathbf{x} and the unemployment rate \mathbf{y} . The local minimum of (7) is indicated by the black line

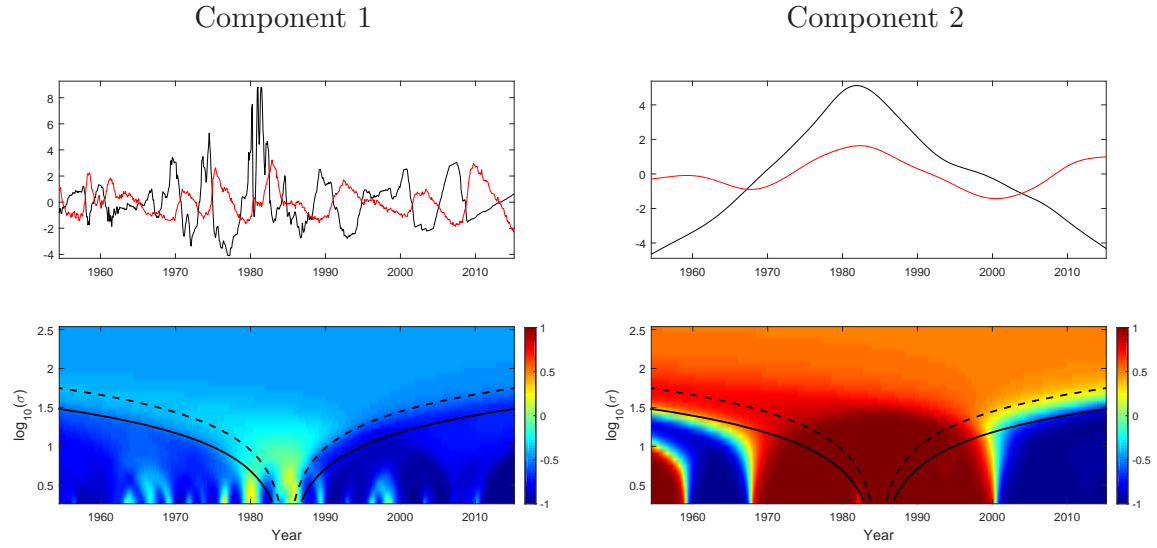


Figure 11: The correlation maps for the two mid-scale multiresolution component pairs of the U.S: economic data. The upper panel of each component: the components \mathbf{u}_j and \mathbf{v}_j in the decompositions (8). Black: the effective federal funds rate. Red: the unemployment rate. The lower panels: The correlation maps

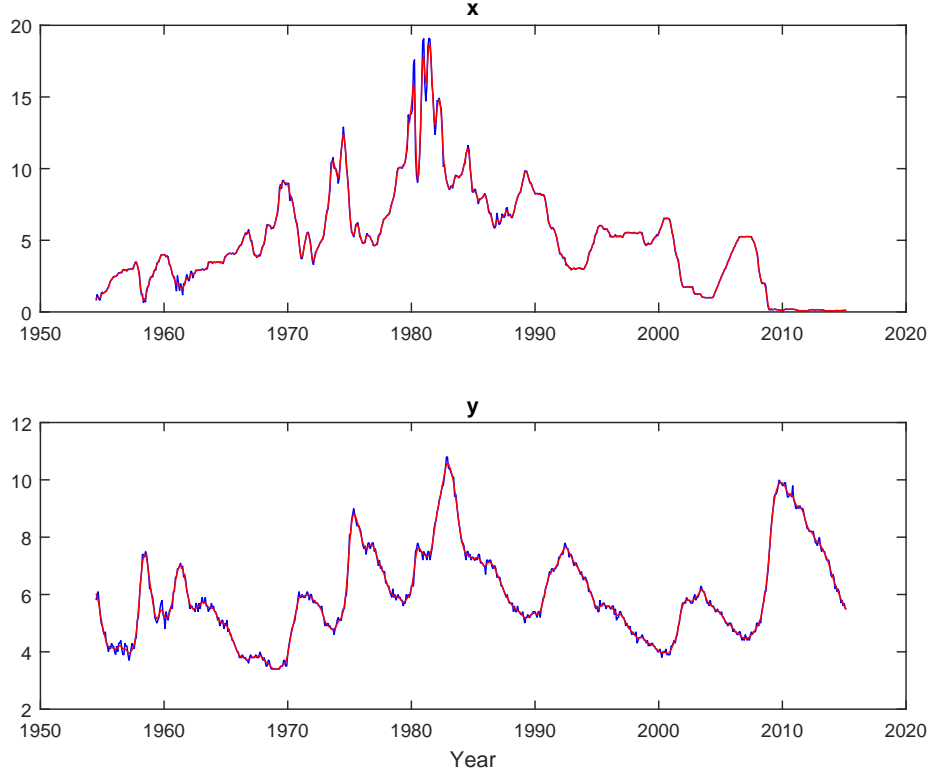


Figure 12: The posterior means and the original time series for the U.S. economic data. Red: the posterior mean. Blue: the original time series

$\kappa_y = 0.045$, $E(\sigma_x^2) = 0.2^2$, $E(\sigma_y^2) = 0.1^2$, and $\nu_{x0} = \nu_{y0} = 10$. The sample size in posterior inference is 10^4 . The posterior means and the original signals are compared in Figure 12. Close inspection of the figures shows that the posterior means are somewhat smoother than the original time series. Based on the scale-derivative maps of the posterior means of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ (maps not shown), the λ -sequence used in the multiresolution decomposition is $[0, 10^{4.2}, \infty]$. Inference about the correlation between the two multiresolution components is summarized in Figure 13. As the noise variance was estimated to be rather small, the credibility maps are mostly black and white. In component 1, the correlation is negative except for small-scale correlation around the years 1980 and 1985. The credibility map of component 2 confirms the findings made on the basis of the correlation maps computed using the multiresolution components of the original time series (Figure 11).

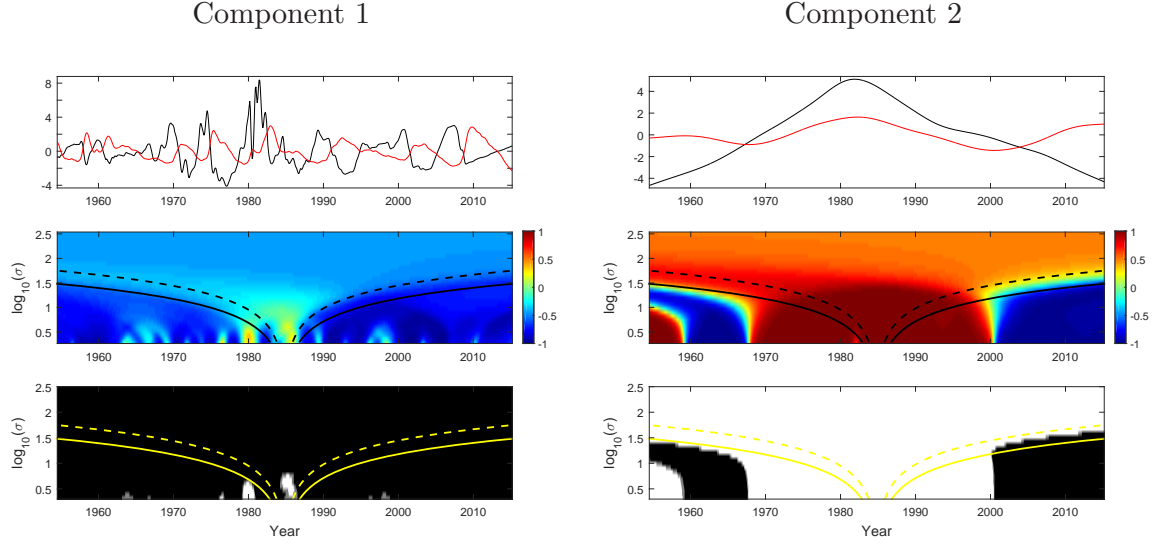


Figure 13: Posterior correlation analysis of the two multiresolution component pairs $(\mathbf{u}_j, \mathbf{v}_j)$ in the decompositions (10) of the U.S. economic data. The upper panel of each component: the posterior means of the components. Black: the effective federal funds rate $E(\mu|\mathbf{x})$. Red: the unemployment rate $E(\nu|\mathbf{y})$. The middle panels: The posterior means of the correlation maps. The lower panels: correlation credibility maps

3.3 Lead-lag effects

Finally, we consider the analysis of lead-lag effects in the proposed framework using the unemployment rate vs. the effective federal funds rate data discussed in Section 3.2 as an example. The small-scale negative correlation inferred for these time series could be the result of lagged positive correlation (cf. the left panels of Figure 11). To find out whether this is the case, we checked the correlation maps of their small-scale components \mathbf{u}_1 and \mathbf{v}_1 (cf. (8)) for lags $1, \dots, 150$, with either \mathbf{u}_1 leading and \mathbf{v}_1 lagging, or vice versa. The strongest positive correlation, that is, a mostly red correlation map was obtained by applying a lag of 25 months to the unemployment rate component \mathbf{v}_1 . The map is shown in Figure 14 and it suggests a lagged connection where unemployment follows federal funds rate by approximately two years. However, prior to 1967 and after 1993, this connection is not as strong as in the years between 1967 and 1993. On the other hand, the lower right panel of Figure 11 shows that for the larger scale components \mathbf{u}_2 and \mathbf{v}_2 , the correlation is mostly positive during the time interval considered. Further, none of the non-zero lags considered seemed to produce a better match between these components. Thus, at least between 1970 and 2000, the long-term increase and decrease in the federal funds rate seems to be associated with similar trends in the unemployment rate. Superimposed on this large-scale picture, periodic increases

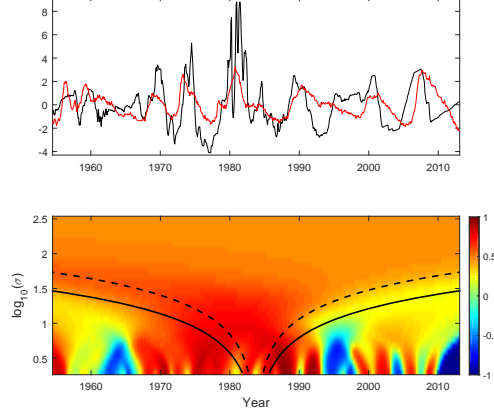


Figure 14: The lagged correlation map for the smallest multiresolution components in the U.S. economic data. The upper panel: the small scale components. Black: the effective federal funds rate component \mathbf{u}_1 . Red: the unemployment rate component \mathbf{v}_1 lagged by 25 months. The lower panel: The correlation map.

and decreases in the federal funds rate appear to have led to corresponding upward and downward movements in the unemployment rate after a time lag of approximately two years.

For comparison, the ordinary sample cross-correlation plot of the original time series is shown in Figure 15. The strongest positive cross-correlation is obtained for lag 25. Hence, the roughly two year lagged connection observed in the small scale correlation map is evident also in the cross-correlation of the original time series. Inspecting the correlation map of the original time series for lag -70, we observed that the positive sample cross-correlation around this lag (cf. Figure 15) corresponds to positive local correlation in long time horizons. When the maps for the multiresolution components were then examined, this correlation appeared to stem mostly from a match between the lagged large-scale components (result not shown). For lags beyond the range $[-150, 150]$, the correlation was mostly negative (result not shown). Note that for lag 0, the sample cross-correlation almost vanishes. Therefore, the large-scale positive correlation revealed by the scale space multiresolution correlation analysis is missed if only cross-correlation is considered. Of course, the cross-correlation coefficient also cannot detect temporal changes in correlation.

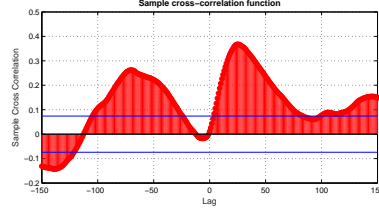


Figure 15: The sample cross-correlation plot for the U.S. economic data. The blue lines indicate the approximate 95% confidence intervals assuming that the two variables are independent.

3.4 Comparison with wavelet coherence

Wavelet coherence (WTC) can be thought of as a localized correlation coefficient in the time-frequency space and it is therefore natural to compare it with the results obtained with the scale space multiresolution correlation method. Comparisons were made using the cross wavelet and wavelet coherence Matlab toolbox written by Alan Grinsted and available at

<http://www.glaciology.net/wavelet-coherence>. For more information on wavelet coherence, see Grinsted et al. (2004).

The left panel of Figure 16 shows the results of WTC analysis for the artificial time series of Section 3.1. The arrows in the map show the phase difference of the signals and the color indicates the strength of the coherence. The statistical significance of coherence is inferred using a Monte Carlo method against AR(1) noise and the significant wavelet coherence areas are surrounded by black lines. The smaller wavelength sinusoidal oscillation is detected, but the larger wavelength is not. The right panel of Figure 16 shows the results of WTC analysis for the unemployment rate vs. effective federal funds rate discussed in Section 3.2. Wavelet coherence is able to detect the small scale negative correlation, but not the positive correlation present in larger scales.

The continuous wavelet transformation produces edge artifacts because the wavelet is not completely localized in time. The part of the WTC illustrations where edge effects may distort the results is called the cone of influence (COI) and in the plots it is shown in lighter shade colors. In the above analysis of WTC results we considered only scales for which at least some time points fall outside the COI. Should larger scales be included, results that can be interpreted as corresponding to our scale space correlation analyses can be inferred (plots not shown). Thus, at scale 2048 the signals of the artificial data example are in antiphase, which can be interpreted to arise from the negative correlation of

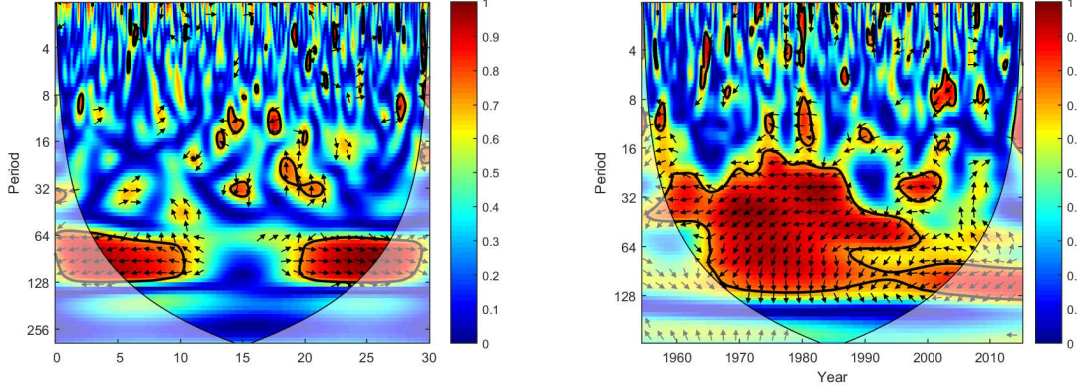


Figure 16: Wavelet coherence analysis of the time series considered in Section 3. Left: analysis of the artificial example considered in Section 3.1. Right: analysis of the unemployment rate vs. effective federal funds rate data considered in Section 3.2. The horizontal axis represents the time and the vertical axis represents the Fourier period which is related to scale

their linear trends. Similarly, the economic time series at scale 512 are in phase reflecting their long-term positive correlation. However, it should be noted that making such interpretations requires consideration of parts of the WTC map that fall completely within the COI making any conclusions susceptible to edge effects.

These results suggest that scale space multiresolution correlation analysis can be more sensitive than wavelet coherence in the detection of the salient features in the correlation between two time series. Its graphical summary of the statistical analyses may also be easier to interpret.

4 Summary

We proposed a new scale space approach for the discovery of correlation structures between two time series. The method takes into consideration the possibility that correlation may not be constant in time and that it might have different features when viewed at different time scales.

The method also provides sampling-based inference about the credibility of the correlation features. The posterior distributions of the time series are used to detect the credibly positive or negative correlations at different scales. Since posterior modeling and correlation analysis are separate steps, the method is very flexible allowing different noise and prior structures in the model.

The performance of the method was demonstrated using three examples, one artificial and two real data sets; the second real data example is included in Online Resource 1. All three examples showed the importance of using multiresolution components of the time series in the correlation analysis and also the need for local correlation analysis with a wide range of scales.

We showed how lead-lag effects in the data can be identified with scale space multiresolution correlation analysis by considering a range of lags and then identifying those that produce the strongest correlations between the multiresolution components. Compared to the traditional cross-correlation analysis, such an approach provides information about temporal changes in correlation and it is also able to resolve superposed, scale-related correlation structures in the data.

Our method was compared with wavelet coherence. Scale space multiresolution correlation analysis appeared to better detect the relevant correlation structures and its results were easier to interpret. Another advantage of the proposed method is that, as opposed to wavelet coherence, the time grid in the series does not have to be equispaced.

Acknowledgement

We are grateful to Antti Ojala and Ilkka Launonen for providing us with the time series and the posterior samples for the varved lake sediment data example in Online Resource 1. We also thank Alan Grinsted for making the wavelet coherence Matlab toolbox publicly available. The work of both authors was supported by the Academy of Finland grant no. 250862.

References

- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995. ISSN 00359246.
- G. E. P. Box and G. C. Tiao. A canonical analysis of multiple time series. *Biometrika*, 64(2): 355–365, 1977.

- P. Chaudhuri and J. S. Marron. SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, 94(447):807–823, 1999.
- J. H. Cochrane. The return of the liquidity effect: A study of the short-run relation between money growth and interest rates. *Journal of Business & Economic Statistics*, 7(1):75–83, 1989.
- R. F. Engle and C. W. J. Granger. Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2):251–276, 1987.
- P. Erästö and L. Holmström. Bayesian multiscale smoothing for making inferences about features in scatter plots. *Journal of Computational and Graphical Statistics*, 14(3):569–589, 2005.
- O. Erdem, E. Ceyhan, and Y. Varli. A new correlation coefficient for bivariate time-series data. *Physica A: Statistical Mechanics and its Applications*, 414:274 – 284, 2014. ISSN 0378-4371.
- FEDFUNDS. Board of governors of the federal reserve system (US), effective federal funds rate retrieved from fred, federal reserve bank of st. louis. <https://fred.stlouisfed.org/series/FEDFUNDS>, 2015. Accessed: 2015-06-8.
- F. Godtlielsen, L. Holmström, A. Miettinen, P. Erästö, D. V. Divine, and N. Koc. Pairwise scale space comparison of time series with application to climate research. *Journal of Geophysical Research: Oceans*, 117(C3):n/a–n/a, 2012. ISSN 2156-2202.
- C. W. J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16(1):121 – 130, 1981.
- A. Grinsted, J. C. Moore, and S. Jevrejeva. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11(5/6):561–566, 2004.
- L. Holmström. Scale space methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2):150–159, 2010. ISSN 1939-0068.
- L. Holmström and L. Pasanen. Statistical scale space methods. To appear in *International Statistical Review*, 2016.

- L. Holmström, L. Pasanen, R. Furrer, and S. R. Sain. Scale space multiresolution analysis of random signals. *Computational Statistics & Data Analysis*, 55(10):2840 – 2855, 2011.
- T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.
- M. Mudelsee. Estimating pearson’s correlation coefficient with bootstrap confidence interval from serially dependent time series. *Mathematical Geology*, 35(6):651–665, 2003. ISSN 0882-8121.
- D. Nikolić, R. C. Mureşan, W. Feng, and W. Singer. Scaled correlation analysis: a better way to compute a cross-correlogram. *European Journal of Neuroscience*, 35(5):742–762, 2012.
- K. Ólafsdóttir and M. Mudelsee. More accurate, calibrated bootstrap confidence intervals for estimating the correlation between two time series. *Mathematical Geosciences*, 46(4):411–427, 2014. ISSN 1874-8961.
- S. Papadimitriou, J. Sun, and P. S. Yu. Local correlation tracking in time series. In *Data Mining, 2006. ICDM ’06. Sixth International Conference on*, pages 456–465, Dec 2006.
- L. Pasanen, I. Launonen, and L. Holmström. A scale space multiresolution method for extraction of time series features. *Stat*, 2(1):273–291, 2013.
- B. Podobnik and H. E. Stanley. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Phys. Rev. Lett.*, 100(8), Feb 2008.
- F. Pozzi, T. Di Matteo, and T. Aste. Exponential smoothing weighted correlations. *The European Physical Journal B*, 85(6), 2012. ISSN 1434-6028.
- J. L. Rodgers and W. A. Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.
- UNRATE. Us. bureau of labor statistics, civilian unemployment rate, retrieved from fred, federal reserve bank of st. louis. <https://research.stlouisfed.org/fred2/series/UNRATE/>, 2015. Accessed: 2015-06-8.

- A. P. Witkin. Scale-space filtering. In *8ht International Joint Conference of Artificial Intelligence*, pages 1019–1022, August 1983.
- A. P. Witkin. Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, volume 9, pages 150–153, Mar 1984.
- G. U. Yule. Why do we sometimes get nonsense-correlations between time-series?—A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89(1):1–63, 1926. ISSN 09528385.

An on-line supplement to

Scale space multiresolution correlation analysis for time series data

L. Pasanen and L. Holmström

Department of Mathematical Sciences

University of Oulu, Finland

1 Scale space multiresolution correlation analysis of varved lake sediment records

This example considers climate change, more specifically changes in the past climate since the last Ice Age. Of particular interest in paleoclimate research are periodic or quasiperiodic fluctuations in the climate driven for example by solar variability and ocean-atmosphere circulation patterns. Spectral methods can detect the presence of such phenomena but techniques that work in the time domain can also capture transient phenomena and trends in the data. We therefore believe that scale space multiresolution correlation analysis can be a useful tool in the analysis of paleoclimate reconstructions.

Annually accumulating sediment layers in lakes that have been undisturbed for thousands of years can serve as a proxy for the natural cycles and trends that force climate variability in different time scales (Ojala et al, 2015). We analyze the correlation between logarithmic clastic (inorganic) lamina thickness in sediment records from lakes Nautajärvi and Korttajärvi, located in Finland, during the

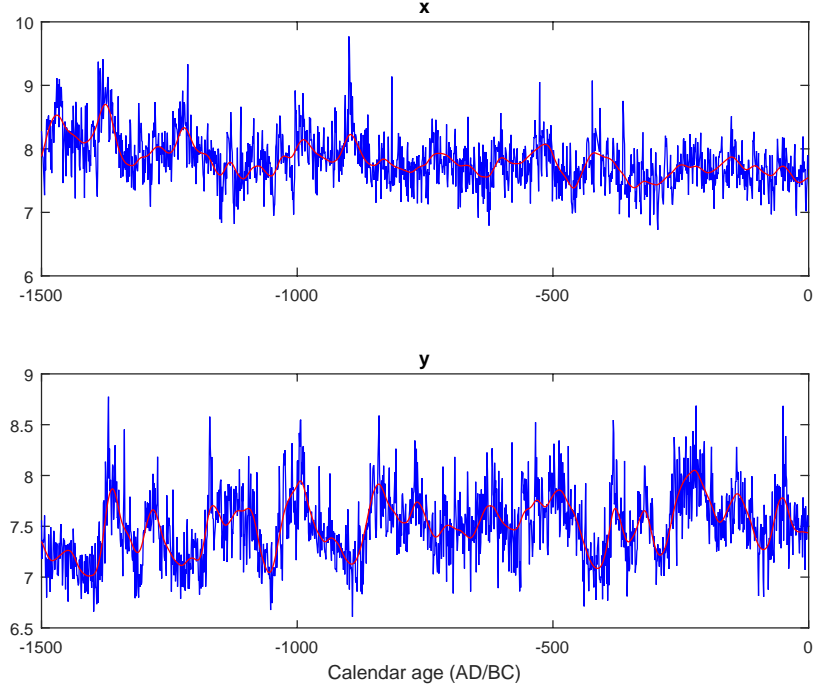


Figure S.1: Time series of clastic lamina thickness in bottom sediments of lakes Korttajärvi and Nautajärvi. The time series are shown in blue and the posterior means of the noiseless signals are shown in red. Horizontal axis: time before present with 0 indicating the present. Vertical axis: logarithm of clastic lamina thickness

time period 1950-3450 cal BP (calendar years before present). The two time series are shown in Figure S.1.

In the examples of the main paper, noise in the observed time series was assumed to be iid Gaussian. As shown in Ojala et al (2015), this is not a valid model for the logarithmic clastic lamina thickness data. Instead, the proper model for additive noise is a first order Gaussian autoregressive process and the Bayesian time series model has to be modified accordingly. Such a modification is described in Ojala et al (2015) where AR(1) additive noise was assumed, considering all parameters of the noise model and the underlying signal prior as unknown.

The posterior sample was drawn using hybrid Gibbs sampling by picking every 100th sample after a burn-in period of 10^5 for lake Nautajärvi and $4 \cdot 10^5$ samples for lake Korttajärvi, the final sample size for both being 2000. There were 7 consecutive missing observations after calendar year -417 in the Korttajärvi data. As our method assumes that the time series considered are observed at the same time points, the missing 7 observations were imputed by linear interpolation. The

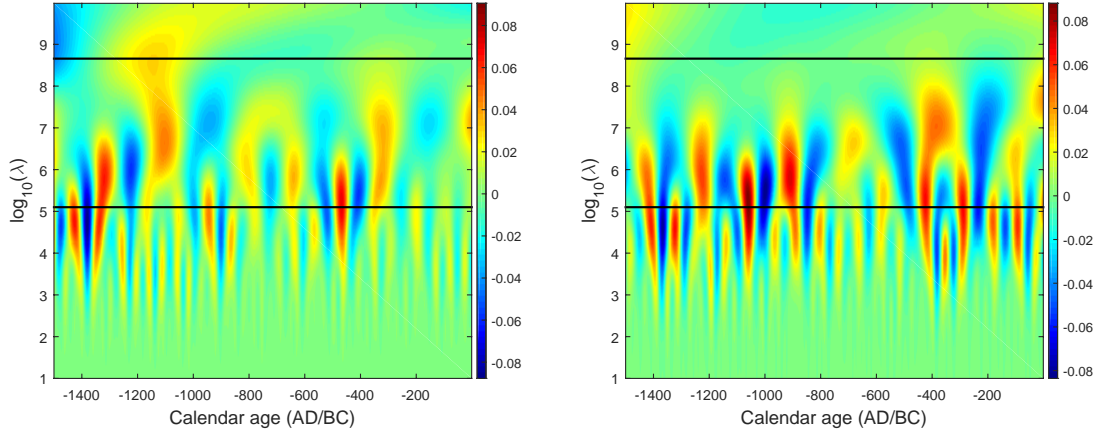


Figure S.2: The scale-derivative maps of the posterior means for the Korttajärvi and Nautajärvi time series. The black lines indicate the smoothing parameter values λ_j used in the multiresolution decomposition (6) of the main paper

posterior means of the underlying signals are shown in Figure S.1.

The posterior mean scale-derivative maps are shown in Figure S.2. Both maps seem to contain two oscillating bands of blue and red, above and below $\lambda = 10^{5.1}$. However, since the bands overlap to some extent, there is no local minimum in the function (11) of the main paper between the two bands and the line $\lambda = 10^{5.1}$ was added to the map on the basis of this visual judgment only. Hence, the multiresolution level sequence applied is $[0, 10^{5.1}, 10^{8.91}]$; due to the large value of the local minimum $10^{8.91}$, the value ∞ is omitted. The posterior correlation maps of the multiresolution components are displayed in Figure S.3. The components of each time series consist two oscillating wave forms and a long-scale monotone trend. The trend components show a one-unit difference between lakes Nautajärvi and Korttajärvi at the beginning of the time interval considered which gradually disappears towards the end of the interval. This gives rise to negative correlation in large time scales.

In Ojala et al (2015), the authors searched for cyclical features in these data sets using posterior singular spectrum analysis (PSSA) and the Lomb-Scargle periodogram. They discovered that both time series contained several cyclical components but only oscillations of wavelengths 70-100 and 150-170 years were found in both of them. These components are shown in the lower panel of Figure 7 of Ojala et al (2015) and they seem to correspond to our components 1 and 2. For example, the 150-170 year oscillations of Ojala et al. contain 9-10 cycles and our component 2 contains approximately 7-10 cycles. Also, Ojala et al.'s oscillations are in antiphase, except for the time interval $[-300, 0]$

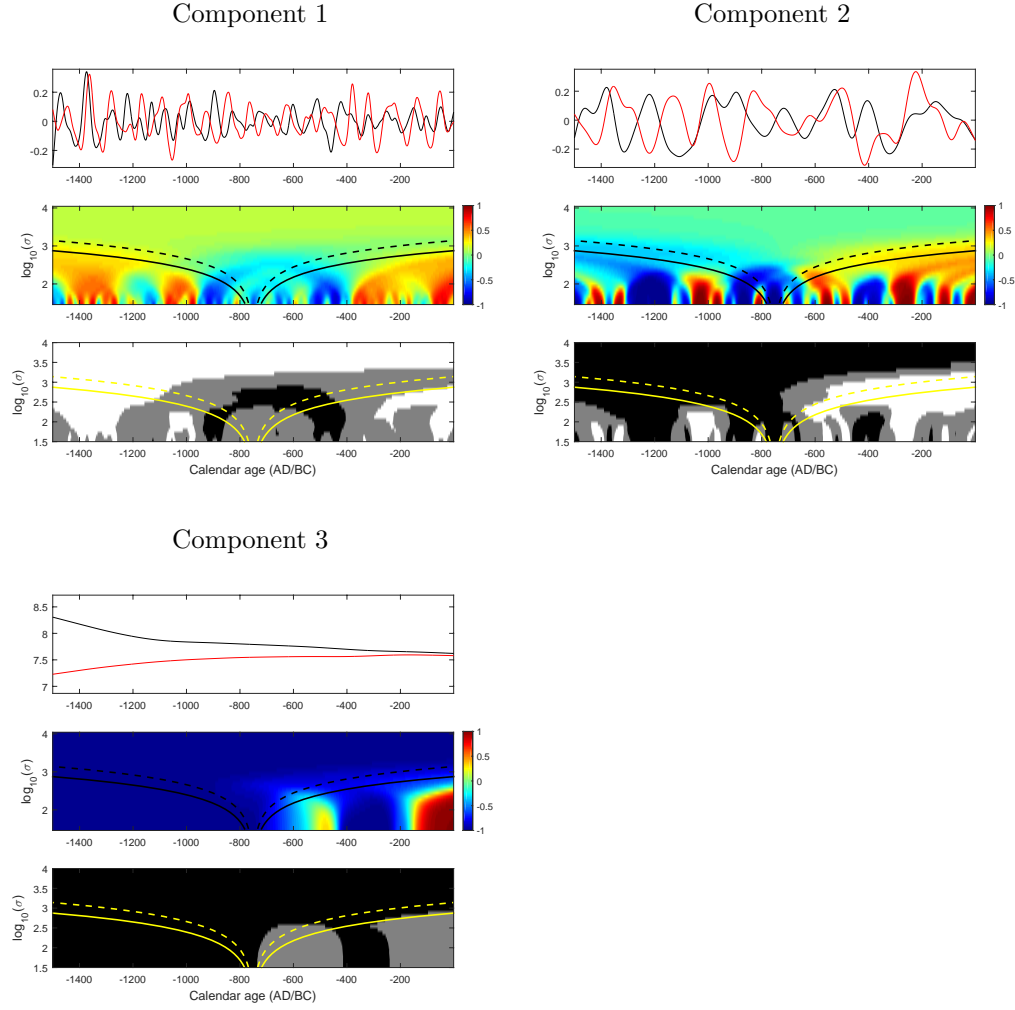


Figure S.3: Posterior correlation analysis of the three multiresolution component pairs $(\mathbf{u}_j, \mathbf{v}_j)$ in the decompositions (10) of the main paper for the Korttajärvi and Nautajärvi clastic lamina thickness data. The upper panel of each component: the posterior means of the components. Black: Korttajärvi $E(\boldsymbol{\mu}|\mathbf{x})$. Red: Nautajärvi $E(\boldsymbol{\nu}|\mathbf{y})$. The middle panels: The posterior means of the correlation maps. The lower panels: correlation credibility maps

where they are in phase. Similarly, the centennial scale correlation of component 2 in the two time series is credibly negative from -1400 until about -700 and changes sign after that, being credibly positive from around -600 onward. The average, long-term correlation between these components is close to zero (-0.09), but still credibly negative over the 1500 year time period considered. Ojala et al. note that this type of periodicity has been rarely recorded and it is either caused by a random feature created during sediment sampling, varve analyses or data treatment or that, alternatively, there is an additional local mechanism pacing sedimentation in these lakes through an unidentified site-specific threshold.

There is also similarity between Ojala et al.'s 70-100 year cycles and our component 1. However, the similarity is not as clear as for component 2 since the oscillations found in Ojala et al (2015) contain 15-16 cycles, whereas our component 1 of lake Nautajärvi contains 18 and that of Korttajärvi contains 21 cycles. Component 1 of Korttajärvi in particular seems a bit noisier in appearance than the corresponding component of Nautajärvi. This might be explained by the fact that Ojala et al. also detected a more rapid 47 year oscillation in the Korttajärvi data, and this might be mixed in our component 1 causing it to have more cycles. However, comparing the phase difference of the oscillations found in Ojala et al (2015) and the sign of the correlation between our components, the results are reasonably similar. Their 70-100 year oscillation is in phase in $[-1500, -1300]$, around -1000, and then again in $[-500, -1]$, and more or less in antiphase elsewhere. For component 1, correlation is inferred to be positive in approximately these time intervals and negative in $[-900, -400]$, where the oscillations of Ojala et al (2015) are mostly in antiphase. The overall correlation over the whole 1500 year period of these components is close to zero (0.1), but still credibly positive. According to Ojala et al., the periodicity exhibited by this oscillation might be related to the 88-year Gleissberg solar cycle.

As in the examples of the main paper, all relevant features in the dependence structure between the two time series cannot be found by local correlation analysis that uses only one fixed window length. This emphasizes the importance of a scale space approach that considers a whole range of window widths, instead.

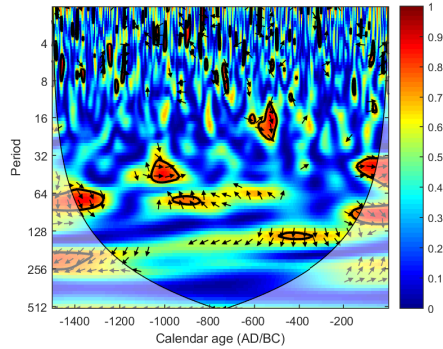


Figure S.4: Wavelet coherence analysis of varved lake sediment records. The horizontal axis represents the time and the vertical axis represents the Fourier period which is related to scale

2 Wavelet coherence analysis of the varved lake sediment records

Figure S.4 shows the results of wavelet coherence analysis for the varved lakes sediment data. Wavelet coherence is able to detect some of the features discovered with scale space multiresolution correlation analysis and posterior SSA. However the results are more difficult to interpret.

References

Ojala A, Launonen I, Holmström L, Tiljander M (2015) Effects of solar forcing and north atlantic oscillation on the climate of continental scandinavia during the holocene. *Quaternary Science Reviews* 112:153 – 171