Decomposition of time series: An application to simulated time series

Maria Almeida Silva

Instituto Superior Técnico

December 2018

Outline

Introduction - Basic theory about stochastic processes

Let $(\Omega, \mathcal{A}, P)$ be a probability space, where Ω is the sample space, $\mathcal{A}$ is a σ-algebra and P is a probability measure function. Let $(E, \boldsymbol{\varepsilon})$ be a measurable space and $\mathcal{T}$ a set of indexes.

- A stochastic process indexed by $\mathcal{T}$ is a family of random variables $(X_t, t \in \mathcal{T})$ defined over $(\Omega, \mathcal{A}, P)$ and with values in $(E, \boldsymbol{\varepsilon})$.
- $\mathcal{T}$ is called the time-space.
- E is called the state space of the process.
- For each $w \in \Omega$, fixed, $(X_t(w), t \in \mathcal{T})$ defines a realisation or a trajectory of the process.

Introduction - Basic theory about stochastic processes

- A stochastic process is strictly stationary if:

$$\forall k, \forall t_1, \ldots, t_k \in \mathbb{Z}, \forall h \in \mathbb{Z},\ (X_{t_1}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h}).$$

Introduction - Basic theory about stochastic processes

- A stochastic process is strictly stationary if:

$$\forall k, \forall t_1, \ldots, t_k \in \mathbb{Z}, \forall h \in \mathbb{Z},\ (X_{t_1}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h}).$$

- A stochastic process is weakly stationary (or just stationary) if:
  - $E(X_t) = m, \forall t \in \mathbb{Z}$,
  - $Cov(X_t, X_{t+h}) = \gamma(h), \forall t \in \mathbb{Z}$.

Introduction - Basic theory about stochastic processes

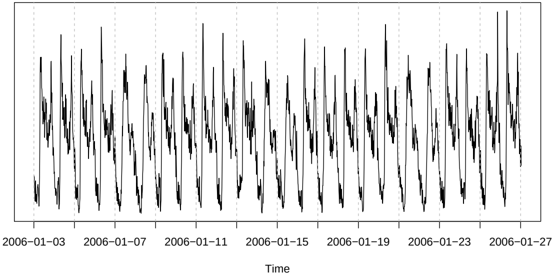
- A stochastic process is strictly stationary if:

$$\forall k, \forall t_1, \ldots, t_k \in \mathbb{Z}, \forall h \in \mathbb{Z}, \ (X_{t_1}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h}).$$

- A stochastic process is weakly stationary (or just stationary) if:
  - $E(X_t) = m, \forall t \in \mathbb{Z}$,
  - $Cov(X_t, X_{t+h}) = \gamma(h), \forall t \in \mathbb{Z}$.
- White noise is a random series $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ such that:
  - $E[\varepsilon_t] = 0, \ \forall t \in \mathbb{Z}$,
  - $cov(\varepsilon_s, \varepsilon_t) = 0, \ \forall s \neq t$,
  - $Var(\varepsilon_t) = \sigma^2$, independent of t.

Introduction - Basic theory about stochastic processes

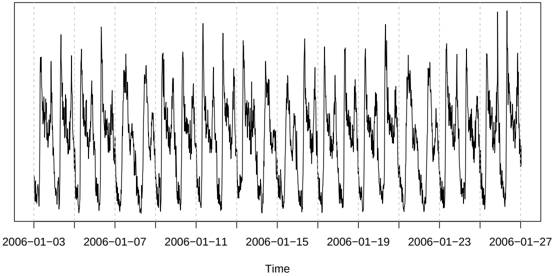
- Considering that the time is discrete, *e.g.* $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{Z}$, then the stochastic process is called a stochastic process in discrete time or a time series.



Example of a time series.

Introduction - Basic theory about stochastic processes

- Considering that the time is discrete, *e.g.* $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{Z}$, then the stochastic process is called a stochastic process in discrete time or a time series.
- The lag operator (L) operates on an element of a time series to produce the previous element, *i.e.*, $LX_t = X_{t-1}, \forall t \in \mathbb{Z}$ (or $t > 1$).



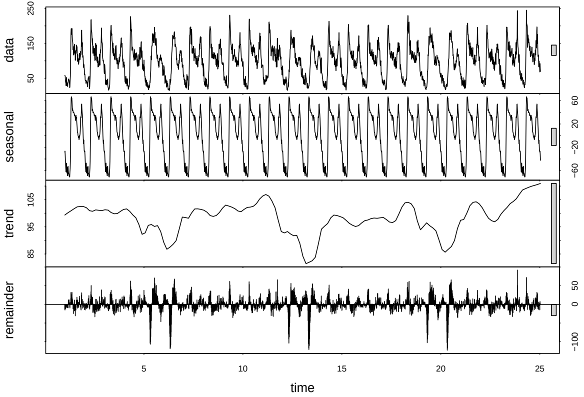
Example of a time series.

Introduction - Decomposition of time series

- In general, time series are decomposed into their natural components (trend, seasonal, cyclical and irregular components) before further analysis.
- The irregular component is the hardest component to model, which makes of it the most interesting.

Introduction - Decomposition of time series

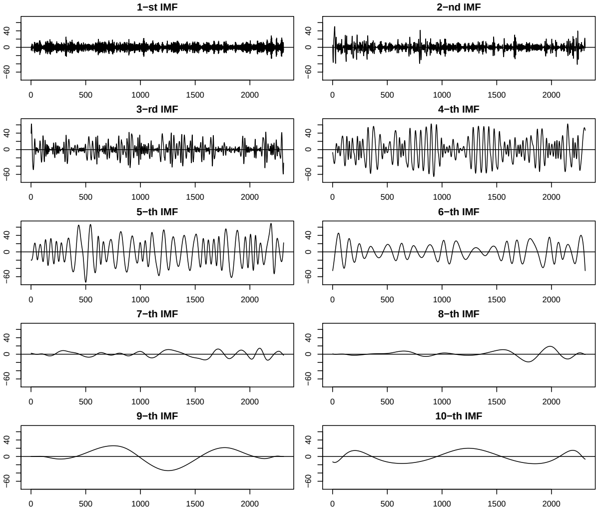
- In general, time series are decomposed into their natural components (trend, seasonal, cyclical and irregular components) before further analysis.
- The irregular component is the hardest component to model, which makes of it the most interesting.



Example of a time series decomposition into seasonal, trend and irregular components.

Introduction - Decomposition of time series

- Nevertheless, a decomposition into unusual components can also be useful in several areas of research.



Example of a time series decomposition into unusual components.

Seasonal-Trend decomposition process based on LOESS (STL)

STL is a classical method for decomposing time series into trend, seasonal and remainder components, proposed by Cleveland *et al.* (1990) [1]. This method uses iterative LOESS smoothing to obtain an estimate of the trend and then LOESS smoothing again to extract a changing additive seasonal component. LOESS smoothing denotes a method that is also known as locally weighted polynomial regression.

This method is not fully automated and a set of parameters must be manually set to guarantee the performance.

Discrete Fourier Transform (DFT)

The Fourier Transform (FT) is one of the techniques used to decompose a signal into a sum of sinusoids.

DFT is a classical technique used to perform Fourier analysis in many practical applications. DFT converts a time series into the frequency domain.

Discrete Fourier Transform (DFT)

The Fourier Transform (FT) is one of the techniques used to decompose a signal into a sum of sinusoids.

DFT is a classical technique used to perform Fourier analysis in many practical applications. DFT converts a time series into the frequency domain.

The assumption behind Fourier analysis is that a time series can be decomposed into a set of linear, stationary and harmonic components. The number of harmonics required to describe a time series increases when the non-linearity and non-stationarity of the time series increases.

Discrete Fourier Transform (DFT)

The Fourier Transform (FT) is one of the techniques used to decompose a signal into a sum of sinusoids.

DFT is a classical technique used to perform Fourier analysis in many practical applications. DFT converts a time series into the frequency domain.

The assumption behind Fourier analysis is that a time series can be decomposed into a set of linear, stationary and harmonic components. The number of harmonics required to describe a time series increases when the non-linearity and non-stationarity of the time series increases.

Although DFT is valid under extremely general conditions, there are some crucial restrictions for a Fourier spectral analysis:

- the system must be linear;
- the data must be strictly periodic or stationary.

Otherwise, the resulting spectrum will make little physical sense.

Discrete Fourier Transform (DFT)

The Fourier Transform (FT) is one of the techniques used to decompose a signal into a sum of sinusoids.

DFT is a classical technique used to perform Fourier analysis in many practical applications. DFT converts a time series into the frequency domain.

The assumption behind Fourier analysis is that a time series can be decomposed into a set of linear, stationary and harmonic components. The number of harmonics required to describe a time series increases when the non-linearity and non-stationarity of the time series increases.

Although DFT is valid under extremely general conditions, there are some crucial restrictions for a Fourier spectral analysis:

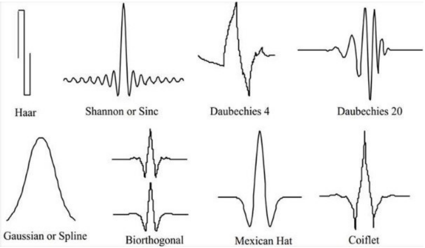
- the system must be linear;
- the data must be strictly periodic or stationary.

Otherwise, the resulting spectrum will make little physical sense.

DFT cannot accurately model a linear trend or non-linear abnormality.

Discrete Wavelet Transform (DWT)

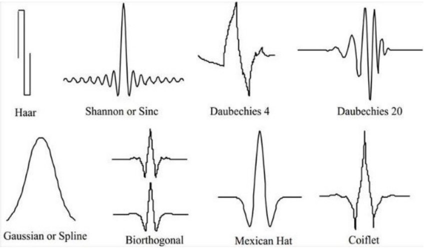
A wavelet is a mathematical function, that can be viewed as a wave-like oscillation.



Examples of wavelet functions. [Source: Raj Endiran]

Discrete Wavelet Transform (DWT)

A wavelet is a mathematical function, that can be viewed as a wave-like oscillation.

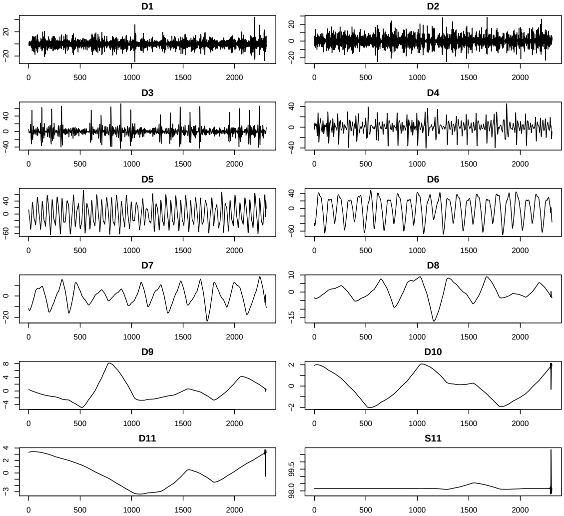


Examples of wavelet functions. [Source: Raj Endiran]

DWT is a tool for decomposing a signal by location and frequency. It is similar to DFT. However, DWT makes it possible to analyse time series not only in the frequency domain, like DFT, but also locally in the time domain. Thus, DWT is very helpful for non-stationary processes.

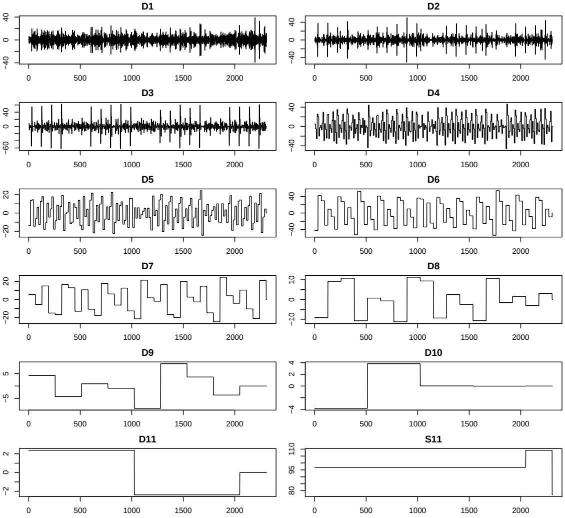
Discrete Wavelet Transform (DWT)

The decomposition is represented in the form of a set of detail series (related with the high-pass filter) and one approximation series (related to the low-pass filter).



Example of a time series decomposition using DWT.

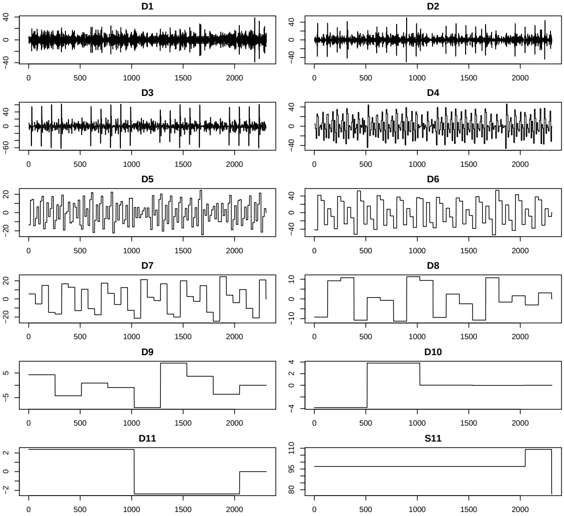
Discrete Wavelet Transform (DWT)



Example of a time series decomposition using DWT.

Discrete Wavelet Transform (DWT)

The choice of the basic wavelet function has a significant influence on the wavelet decomposition results, because the essence of DWT is to discover the similarity between the analysed series and the wavelet used.



Example of a time series decomposition using DWT.

Wavelet Packet Transform (WPT)

An extension of DWT is the Wavelet Packet Decomposition (WPD) [3].

With WPD, besides the approximation coefficients, the detail series is also decomposed, resulting in a set of detail series and a set of approximation series.

Wavelet Packet Transform (WPT)

An extension of DWT is the Wavelet Packet Decomposition (WPD) [3].

With WPD, besides the approximation coefficients, the detail series is also decomposed, resulting in a set of detail series and a set of approximation series.

WPD achieves better frequency resolution for the decomposed signal than DWT since the last one may miss important information in higher frequency components.

Wavelet Packet Transform (WPT)

An extension of DWT is the Wavelet Packet Decomposition (WPD) [3].

With WPD, besides the approximation coefficients, the detail series is also decomposed, resulting in a set of detail series and a set of approximation series.

WPD achieves better frequency resolution for the decomposed signal than DWT since the last one may miss important information in higher frequency components.

The decomposition results are still very dependent on the basic wavelet function.

Empirical Mode Decomposition (EMD)

Huang *et al.* [2] proposed in 1998 a new method for analysing nonlinear and non-stationary data, named Hilbert-Huang Transform (HHT). EMD is a key part of the HHT, that decomposes any data set into Intrinsic Mode Functions (IMFs).

Empirical Mode Decomposition (EMD)

Huang *et al.* [2] proposed in 1998 a new method for analysing nonlinear and non-stationary data, named Hilbert-Huang Transform (HHT). EMD is a key part of the HHT, that decomposes any data set into Intrinsic Mode Functions (IMFs).

An IMF is a function that satisfies two conditions:

Empirical Mode Decomposition (EMD)

Huang *et al.* [2] proposed in 1998 a new method for analysing nonlinear and non-stationary data, named Hilbert-Huang Transform (HHT). EMD is a key part of the HHT, that decomposes any data set into Intrinsic Mode Functions (IMFs).

An IMF is a function that satisfies two conditions:

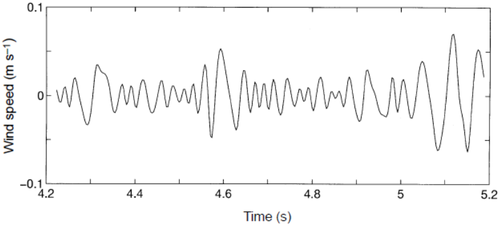
- in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one;

Empirical Mode Decomposition (EMD)

Huang *et al.* [2] proposed in 1998 a new method for analysing nonlinear and non-stationary data, named Hilbert-Huang Transform (HHT). EMD is a key part of the HHT, that decomposes any data set into Intrinsic Mode Functions (IMFs).

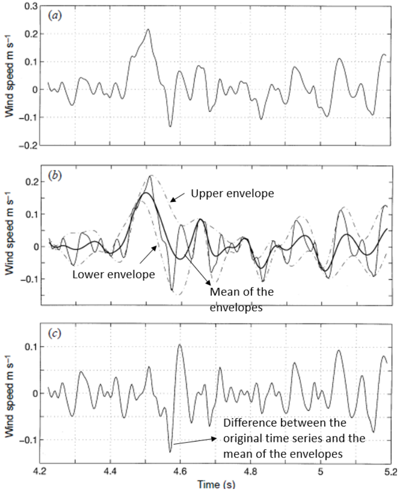
An IMF is a function that satisfies two conditions:

- in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one;
- at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.



Example of an IMF. Source: [2]

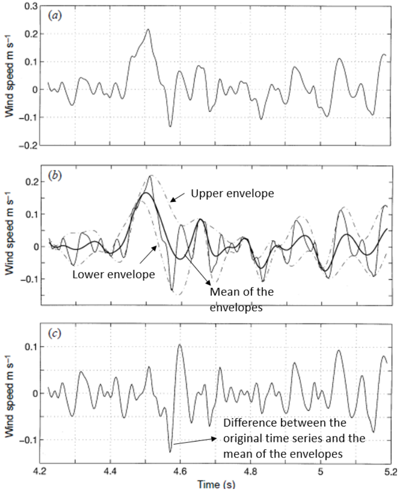
Empirical Mode Decomposition (EMD)



Example of the sifting process. Source: [2] (adapted)

- Identify all the local extrema.

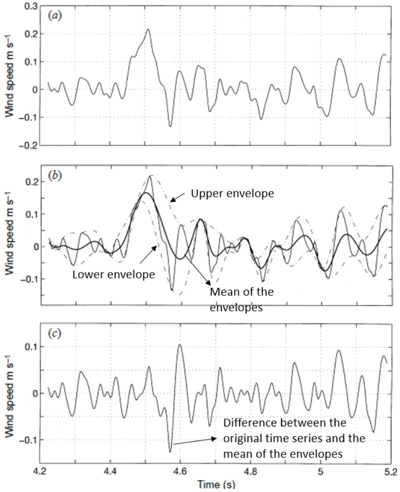
Empirical Mode Decomposition (EMD)



Example of the sifting process. Source: [2] (adapted)

- Identify all the local extrema.
- Connect all the local maxima (minima) by a cubic spline line as the upper (lower) envelope.

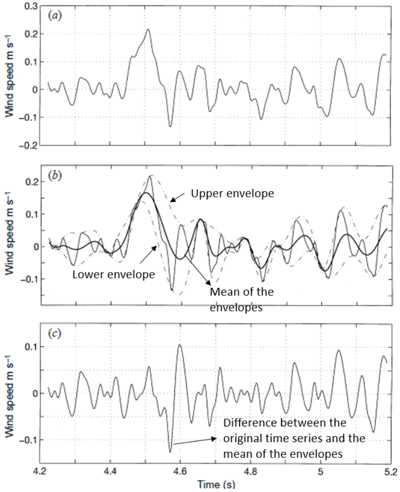
Empirical Mode Decomposition (EMD)



Example of the sifting process. Source: [2] (adapted)

- Identify all the local extrema.
- Connect all the local maxima (minima) by a cubic spline line as the upper (lower) envelope.
- Compute the mean of the upper and lower envelopes, designated as m_1.

Empirical Mode Decomposition (EMD)

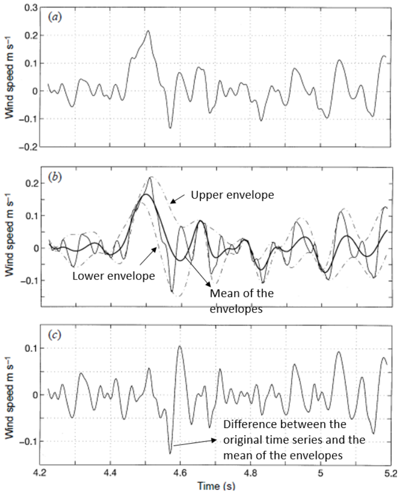


Example of the sifting process. Source: [2] (adapted)

- Identify all the local extrema.
- Connect all the local maxima (minima) by a cubic spline line as the upper (lower) envelope.
- Compute the mean of the upper and lower envelopes, designated as m_1.
- The difference between the data and m_1 is the first component, h_1:

$$X(t) - m_1 = h_1.$$

Empirical Mode Decomposition (EMD)



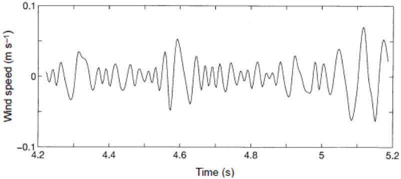
Example of the sifting process. Source: [2] (adapted)

- Identify all the local extrema.
- Connect all the local maxima (minima) by a cubic spline line as the upper (lower) envelope.
- Compute the mean of the upper and lower envelopes, designated as m_1.
- The difference between the data and m_1 is the first component, h_1:

$$X(t) - m_1 = h_1.$$

- If h_1 is not an IMF, then the process is repeated replacing the original time series by h_1, until to obtain an IMF, c_1.

Empirical Mode Decomposition (EMD)

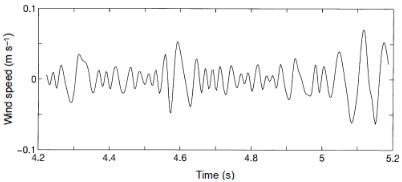


First IMF. Source: [2]

Empirical Mode Decomposition (EMD)

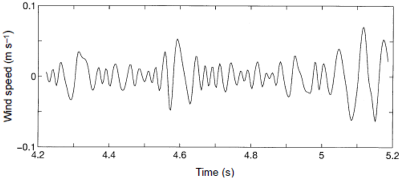
- Consider the residue r_1 obtained as:

$$X(t) - c_1 = r_1.$$



First IMF. Source: [2]

Empirical Mode Decomposition (EMD)



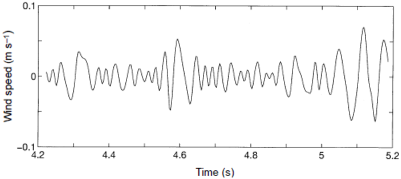
First IMF. Source: [2]

- Consider the residue r_1 obtained as:

$$X(t) - c_1 = r_1.$$

- The residue r_1 is treated as the new data and subjected to the same sifting process as describe before.

Empirical Mode Decomposition (EMD)



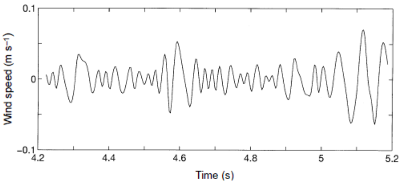
First IMF. Source: [2]

- Consider the residue r_1 obtained as:

$$X(t) - c_1 = r_1.$$

- The residue r_1 is treated as the new data and subjected to the same sifting process as describe before.
- The process stops when the residue r_n becomes a monotonic function from which no more IMF can be extracted.

Empirical Mode Decomposition (EMD)



First IMF. Source: [2]

- Consider the residue r_1 obtained as:

$$X(t) - c_1 = r_1.$$

- The residue r_1 is treated as the new data and subjected to the same sifting process as describe before.
- The process stops when the residue r_n becomes a monotonic function from which no more IMF can be extracted.
- Then

$$X(t) = \sum_{j=1}^{n} c_j + r_n.$$

Empirical Mode Decomposition (EMD)

EMD has been tested and validated but only empirically. The true physical meanings in many of the data examined is an advantage of this method.

Empirical Mode Decomposition (EMD)

EMD has been tested and validated but only empirically. The true physical meanings in many of the data examined is an advantage of this method.

The advantage of EMD, compared to DWT, is that the EMD is a data-driven algorithm: it decomposes a time series into a natural way without prior knowledge about the signal of interest embedded in the data.

Empirical Mode Decomposition (EMD)

EMD has been tested and validated but only empirically. The true physical meanings in many of the data examined is an advantage of this method.

The advantage of EMD, compared to DWT, is that the EMD is a data-driven algorithm: it decomposes a time series into a natural way without prior knowledge about the signal of interest embedded in the data. Moreover, EMD works in temporal space directly rather than in the corresponding frequency space.

Empirical Mode Decomposition (EMD)

EMD has been tested and validated but only empirically. The true physical meanings in many of the data examined is an advantage of this method.

The advantage of EMD, compared to DWT, is that the EMD is a data-driven algorithm: it decomposes a time series into a natural way without prior knowledge about the signal of interest embedded in the data. Moreover, EMD works in temporal space directly rather than in the corresponding frequency space.

Despite the EMD method has been widely adopted to decompose time series, a problem was pointed out when the same IMF has very different amplitudes along time or different IMFs have similar oscillations in amplitudes, what is named as mode mixing.

Ensemble Empirical Mode Decomposition (EEMD)

Aiming solving the mode mixing problem, a new method called Ensemble Empirical Mode Decomposition (EEMD) was developed by Wu and Huang in 2009 [5], in which the original signal is added by random white Gaussian noises.

$$X_i(t) = X(t) + W_i(t)$$

Adding white noise improves the accuracy of the decomposed signal and preserves the original information of the signal.

Ensemble Empirical Mode Decomposition (EEMD)

Aiming solving the mode mixing problem, a new method called Ensemble Empirical Mode Decomposition (EEMD) was developed by Wu and Huang in 2009 [5], in which the original signal is added by random white Gaussian noises.

$$X_i(t) = X(t) + W_i(t)$$

Adding white noise improves the accuracy of the decomposed signal and preserves the original information of the signal.

EEMD depends on the amplitude of the added noise and the ensemble times:

- When the amplitude of the added white noise is too low, the mode mixing problem cannot be suppressed, while if the amplitude is too high, more pseudo components will appear.

Ensemble Empirical Mode Decomposition (EEMD)

Aiming solving the mode mixing problem, a new method called Ensemble Empirical Mode Decomposition (EEMD) was developed by Wu and Huang in 2009 [5], in which the original signal is added by random white Gaussian noises.

$$X_i(t) = X(t) + W_i(t)$$

Adding white noise improves the accuracy of the decomposed signal and preserves the original information of the signal.

EEMD depends on the amplitude of the added noise and the ensemble times:

- When the amplitude of the added white noise is too low, the mode mixing problem cannot be suppressed, while if the amplitude is too high, more pseudo components will appear.
- Relative to the ensemble times, if the noise is added in the EEMD more times, then the noise of the average result is smaller and the result is closer to the real value.

Independent Component Analysis (ICA)

Independent Component Analysis (ICA) has been introduced in order to decompose data into statistically Independent Components (ICs). *A priori* knowledge is not necessary to apply this method, which is an advantage of ICA.

ICA aims to reveal hidden factors and components in data.

Independent Component Analysis (ICA)

Independent Component Analysis (ICA) has been introduced in order to decompose data into statistically Independent Components (ICs). *A priori* knowledge is not necessary to apply this method, which is an advantage of ICA.

ICA aims to reveal hidden factors and components in data.

For $t = 1, 2, \ldots, T$, the ICA model is defined as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t),$$

where $\mathbf{x}(t)$ is a column vector of the observations of a time series set at time instant t, $\mathbf{A}$ is an unknown square matrix and $\mathbf{s}(t)$ is an unknown column vector of the independent components at time instant t.

Independent Component Analysis (ICA)

Independent Component Analysis (ICA) has been introduced in order to decompose data into statistically Independent Components (ICs). *A priori* knowledge is not necessary to apply this method, which is an advantage of ICA.

ICA aims to reveal hidden factors and components in data.

For $t = 1, 2, \ldots, T$, the ICA model is defined as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t),$$

where $\mathbf{x}(t)$ is a column vector of the observations of a time series set at time instant t, $\mathbf{A}$ is an unknown square matrix and $\mathbf{s}(t)$ is an unknown column vector of the independent components at time instant t.

In practice, data should be centred and whitened, before the application of any algorithm.

Independent Component Analysis (ICA)

In general, the ICs are considered as statistically uncorrelated and non-Gaussian.

Independent Component Analysis (ICA)

In general, the ICs are considered as statistically uncorrelated and non-Gaussian.

However, in the time series case, the hypothesis of non-Gaussianity is not necessary. Alternatively, it is assumed that the ICs have different autocovariances, in particular, they are all different from zero. Note that the covariance between ICs is zero due to the independence, however if the data have time-dependences, the autocovariances are often different from zero.

Independent Component Analysis (ICA)

In general, the ICs are considered as statistically uncorrelated and non-Gaussian.

However, in the time series case, the hypothesis of non-Gaussianity is not necessary. Alternatively, it is assumed that the ICs have different autocovariances, in particular, they are all different from zero. Note that the covariance between ICs is zero due to the independence, however if the data have time-dependences, the autocovariances are often different from zero. Another approach is to consider that the variances of the ICs are non-stationary.

Independent Component Analysis (ICA)

In general, the ICs are considered as statistically uncorrelated and non-Gaussian.

However, in the time series case, the hypothesis of non-Gaussianity is not necessary. Alternatively, it is assumed that the ICs have different autocovariances, in particular, they are all different from zero. Note that the covariance between ICs is zero due to the independence, however if the data have time-dependences, the autocovariances are often different from zero. Another approach is to consider that the variances of the ICs are non-stationary.

Moreover, instead of estimate $\mathbf{A}$, a matrix $\mathbf{W}$ is estimated such that:

$$\mathbf{W}\mathbf{z}(t) = \mathbf{s}(t),$$

where $\mathbf{Z}$ is the whitened data.

Since the data is whitened, then $\mathbf{W}$ is an orthogonal matrix, which reduces the complexity of the estimation problem.

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   In this approach, the time-lagged covariance matrix $\mathbf{C}_\tau^\mathbf{z} = E[\mathbf{z}(t)\mathbf{z}(t-\tau)^T]$ is used, where τ is some lag.

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   In this approach, the time-lagged covariance matrix $\mathbf{C}_\tau^\mathbf{z} = E[\mathbf{z}(t)\mathbf{z}(t-\tau)^T]$ is used, where τ is some lag.The motivation for the use of these matrices is, for the ICs, the lagged covariances are all zero due to the independence.

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   In this approach, the time-lagged covariance matrix $\mathbf{C}_\tau^{\mathbf{z}} = E[\mathbf{z}(t)\mathbf{z}(t-\tau)^T]$ is used, where τ is some lag.The motivation for the use of these matrices is, for the ICs, the lagged covariances are all zero due to the independence.

   1.1 Using one lag:

   Since $\mathbf{W}$ is an orthogonal matrix,
   $\overline{\mathbf{C}}_\tau^{\mathbf{z}} = \frac{1}{2}(\mathbf{C}_\tau^{\mathbf{z}} + (\mathbf{C}_\tau^{\mathbf{z}})^T)$

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   In this approach, the time-lagged covariance matrix $\mathbf{C}^{\mathbf{z}}_{\tau} = E[\mathbf{z}(t)\mathbf{z}(t-\tau)^T]$ is used, where τ is some lag.The motivation for the use of these matrices is, for the ICs, the lagged covariances are all zero due to the independence.

   1.1 Using one lag:

   Since $\mathbf{W}$ is an orthogonal matrix, $\overline{\mathbf{C}}^{\mathbf{z}}_{\tau} = \frac{1}{2}(\mathbf{C}^{\mathbf{z}}_{\tau} + (\mathbf{C}^{\mathbf{z}}_{\tau})^T) = \mathbf{W}^T\mathbf{D}\mathbf{W}$, where $\mathbf{D} = \mathbf{C}^{\mathbf{s}}_{\tau}$.

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   In this approach, the time-lagged covariance matrix $\mathbf{C}_\tau^\mathbf{z} = E[\mathbf{z}(t)\mathbf{z}(t-\tau)^T]$ is used, where τ is some lag.The motivation for the use of these matrices is, for the ICs, the lagged covariances are all zero due to the independence.

   1.1 Using one lag:

      Since $\mathbf{W}$ is an orthogonal matrix, $\overline{\mathbf{C}}_\tau^\mathbf{z} = \frac{1}{2}(\mathbf{C}_\tau^\mathbf{z} + (\mathbf{C}_\tau^\mathbf{z})^T) = \mathbf{W}^T\mathbf{D}\mathbf{W}$, where $\mathbf{D} = \mathbf{C}_\tau^\mathbf{s}$. $\mathbf{D}$ is diagonal due to the independence of the ICs and, then, $\mathbf{W}$ is part of the eigenvalue decomposition of $\overline{\mathbf{C}}_\tau^\mathbf{z}$, *i.e.*, the rows of $\mathbf{W}$ are given as the eigenvectors of $\overline{\mathbf{C}}_\tau^\mathbf{z}$.

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   In this approach, the time-lagged covariance matrix $\mathbf{C}_\tau^{\mathbf{z}} = E[\mathbf{z}(t)\mathbf{z}(t-\tau)^T]$ is used, where τ is some lag.The motivation for the use of these matrices is, for the ICs, the lagged covariances are all zero due to the independence.

   1.1 Using one lag:

   Since $\mathbf{W}$ is an orthogonal matrix, $\overline{\mathbf{C}}_\tau^{\mathbf{z}} = \frac{1}{2}(\mathbf{C}_\tau^{\mathbf{z}} + (\mathbf{C}_\tau^{\mathbf{z}})^T) = \mathbf{W}^T\mathbf{D}\mathbf{W}$, where $\mathbf{D} = \mathbf{C}_\tau^{\mathbf{s}}$.
   $\mathbf{D}$ is diagonal due to the independence of the ICs and, then, $\mathbf{W}$ is part of the eigenvalue decomposition of $\overline{\mathbf{C}}_\tau^{\mathbf{z}}$, *i.e.*, the rows of $\mathbf{W}$ are given as the eigenvectors of $\overline{\mathbf{C}}_\tau^{\mathbf{z}}$.
   This algorithm only works when the eigenvectors of $\overline{\mathbf{C}}_\tau^{\mathbf{z}}$ are uniquely defined, otherwise ICs cannot be estimated.

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   1.2 Using several lags:

   Using several lags, we want to simultaneously diagonalize all the corresponding lagged covariance matrices.

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   1.2 Using several lags:

   Using several lags, we want to simultaneously diagonalize all the corresponding lagged covariance matrices. One approach for measuring the nondiagonality is: for any positive-definite matrix $\mathbf{M}$, the nondiagonality of $\mathbf{M}$ can be measure by

$$F(\mathbf{M}) = \sum_i \log(m_{ii}) - \log|\det \mathbf{M}|.$$

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   1.2 Using several lags:

   Using several lags, we want to simultaneously diagonalize all the corresponding lagged covariance matrices. One approach for measuring the nondiagonality is: for any positive-definite matrix $\mathbf{M}$, the nondiagonality of $\mathbf{M}$ can be measure by

$$F(\mathbf{M}) = \sum_i \log\,(m_{ii}) - \log\,|\det\,\mathbf{M}|.$$

   Summing these measures for different time lags and considering whitened data, we obtain the following objective function to minimize:

$$J(\mathbf{W}) = \frac{1}{2}\sum_{\tau\in S} F(\mathbf{W}\overline{\mathbf{C}}_{\tau}^{\mathbf{z}}\mathbf{W}^T) = \sum_{\tau\in S}\sum_i \frac{1}{2}\,\log\,(\mathbf{w}_i(t)\overline{\mathbf{C}}_{\tau}^{\mathbf{z}}\mathbf{w}_i) + \text{constant}.$$

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

    1.2 Using several lags:

    Using several lags, we want to simultaneously diagonalize all the corresponding lagged covariance matrices. One approach for measuring the nondiagonality is: for any positive-definite matrix $\mathbf{M}$, the nondiagonality of $\mathbf{M}$ can be measure by

$$F(\mathbf{M}) = \sum_i \log\,(m_{ii}) - \log\,|\det\,\mathbf{M}|.$$

    Summing these measures for different time lags and considering whitened data, we obtain the following objective function to minimize:

$$J(\mathbf{W}) = \frac{1}{2}\sum_{\tau\in S} F(\mathbf{W}\overline{\mathbf{C}}_\tau^{\mathbf{z}}\mathbf{W}^T) = \sum_{\tau\in S}\sum_i \frac{1}{2}\,\log\,(\mathbf{w}_i(t)\overline{\mathbf{C}}_\tau^{\mathbf{z}}\mathbf{w}_i) + \text{constant}.$$

    Thus, the gradient descent algorithm can be applied, followed by the orthogonalization of $\mathbf{W}$ in each iteration.

Independent Component Analysis (ICA)

1. **Estimation of ICs through the autocovariances:**

   1.2 Using several lags:

   Using several lags, we want to simultaneously diagonalize all the corresponding lagged covariance matrices. One approach for measuring the nondiagonality is: for any positive-definite matrix $\mathbf{M}$, the nondiagonality of $\mathbf{M}$ can be measure by

$$F(\mathbf{M}) = \sum_i \log\,(m_{ii}) - \log\,|\det\,\mathbf{M}|.$$

   Summing these measures for different time lags and considering whitened data, we obtain the following objective function to minimize:

$$J(\mathbf{W}) = \frac{1}{2}\sum_{\tau \in S} F(\mathbf{W}\overline{\mathbf{C}}_\tau^{\mathbf{z}}\mathbf{W}^T) = \sum_{\tau \in S}\sum_i \frac{1}{2}\,\log\,(\mathbf{w}_i(t)\overline{\mathbf{C}}_\tau^{\mathbf{z}}\mathbf{w}_i) + \text{constant}.$$

   Thus, the gradient descent algorithm can be applied, followed by the orthogonalization of $\mathbf{W}$ in each iteration.

The estimation of ICs using autocovariances has a basic limitation that cannot be avoided: if the ICs have identical autocovariances, they cannot be estimated.

Independent Component Analysis (ICA)

2. **Estimation of ICs through the nonstationarity of variances:**

   In this approach, it is assumed that the variances of the ICs change smoothly in time.

Independent Component Analysis (ICA)

2. **Estimation of ICs through the nonstationarity of variances:**

   In this approach, it is assumed that the variances of the ICs change smoothly in time.

   2.1 Using local autocorrelations:

   If we find a matrix $\mathbf{W}$ such that the components of $\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t)$ are uncorrelated at every time point t, we have estimated the ICs.

Independent Component Analysis (ICA)

2. **Estimation of ICs through the nonstationarity of variances:**

   In this approach, it is assumed that the variances of the ICs change smoothly in time.

   2.1 Using local autocorrelations:

   If we find a matrix $\mathbf{W}$ such that the components of $\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t)$ are uncorrelated at every time point t, we have estimated the ICs. The (local) uncorrelatedness of $\mathbf{y}(t)$ could be measured using the same measures of nondiagonality:

$$Q(\mathbf{W}, t) = \sum_i \log\left(E_t[y_i(t)^2]\right) - \log\left(E_t[\mathbf{y}(t)\mathbf{y}(t)^T]\right),$$

   where the expectations are around the time point t.

Independent Component Analysis (ICA)

2. **Estimation of ICs through the nonstationarity of variances:**

   In this approach, it is assumed that the variances of the ICs change smoothly in time.

   2.1 Using local autocorrelations:

   If we find a matrix $\mathbf{W}$ such that the components of $\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t)$ are uncorrelated at every time point t, we have estimated the ICs. The (local) uncorrelatedness of $\mathbf{y}(t)$ could be measured using the same measures of nondiagonality:

$$Q(\mathbf{W}, t) = \sum_i \log\left(E_t[y_i(t)^2]\right) - \log\left(E_t[\mathbf{y}(t)\mathbf{y}(t)^T]\right),$$

   where the expectations are around the time point t. Expressing this as a function of $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)^T$ and considering that $\mathbf{W}$ is orthogonal, the objective function to minimize is given by:

$$J(\mathbf{W}) = \sum_{i,t} \log\left(E_t[(\mathbf{w}_i^T \mathbf{z}(t))^2]\right) + \text{constant}.$$

Independent Component Analysis (ICA)

2. **Estimation of ICs through the nonstationarity of variances:**

   In this approach, it is assumed that the variances of the ICs change smoothly in time.

   2.1 Using local autocorrelations:

   If we find a matrix $\mathbf{W}$ such that the components of $\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t)$ are uncorrelated at every time point t, we have estimated the ICs. The (local) uncorrelatedness of $\mathbf{y}(t)$ could be measured using the same measures of nondiagonality:

$$Q(\mathbf{W}, t) = \sum_i \log\, (E_t[y_i(t)^2]) - \log\, (E_t[\mathbf{y}(t)\mathbf{y}(t)^T]),$$

   where the expectations are around the time point t. Expressing this as a function of $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)^T$ and considering that $\mathbf{W}$ is orthogonal, the objective function to minimize is given by:

$$J(\mathbf{W}) = \sum_{i,t} \log\, (E_t[(\mathbf{w}_i^T \mathbf{z}(t))^2]) + \text{constant}.$$

   Thus, the gradient descent algorithm can be applied, followed by the symmetric orthogonalization of $\mathbf{W}$ in each iteration.

Independent Component Analysis (ICA)

2. **Estimation of ICs through the nonstationarity of variances:**

   2.2 Using non-linear autocorrelations:

   The variance nonstationarity of a signal $Y(t)$ could be measured using a measure based on the fourth-order cross-cumulant:

$$cum(y(t), y(t), y(t-\tau), y(t-\tau)) = E[y(t)^2 y(t-\tau)^2] - E[y(t)^2]E[y(t-\tau)^2] - 2(E[y(t)t(t-\tau)])^2.$$

Independent Component Analysis (ICA)

2. **Estimation of ICs through the nonstationarity of variances:**

   2.2 Using non-linear autocorrelations:

   The variance nonstationarity of a signal $Y(t)$ could be measured using a measure based on the fourth-order cross-cumulant:

$$cum(y(t), y(t), y(t-\tau), y(t-\tau)) = E[y(t)^2 y(t-\tau)^2] - E[y(t)^2]E[y(t-\tau)^2] - 2(E[y(t)t(t-\tau)])^2.$$

   Using the principle of fixed-point iteration, $\mathbf{w}$ is updated as follows:

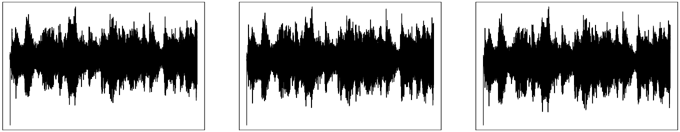
$$\mathbf{w} \leftarrow E[\mathbf{z}(t)\mathbf{w}^T\mathbf{z}(t)(\mathbf{w}^T\mathbf{z}(t-\tau))^2] + E[\mathbf{z}(t-\tau)\mathbf{w}^T\mathbf{z}(t-\tau)(\mathbf{w}^T\mathbf{z}(t))^2] - 2\mathbf{w} - 4\overline{\mathbf{C}}_\tau^{\mathbf{z}}\mathbf{w}(\mathbf{w}^T\overline{\mathbf{C}}_\tau^{\mathbf{z}}\mathbf{w}).$$

   After each iteration, $\mathbf{w}$ is normalised.

   To estimate several ICs, deflationary orthogonalization or symmetric orthogonalization can be used.

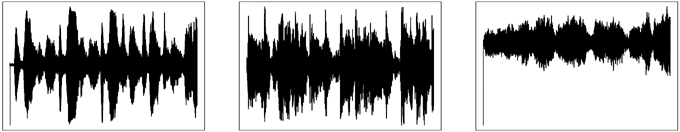
Example of ICA with sounds

Data



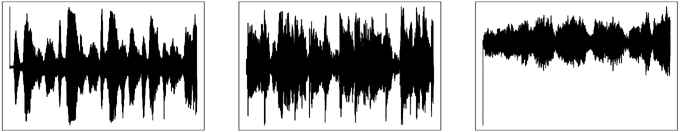
play stop play stop play stop

Structure



play stop play stop play stop

FastICA



play stop play stop play stop

Independent Component Analysis (ICA) for single-channel signal

For the original application of ICA, a set of time series is necessary. Generally, the number of sensors must be no less than that of the sources to acquire information to support the signal decomposition.

However, in real cases, frequently only one single measure of a certain specific physical variable is available.

In this case, EMD and ICA can be combined. The proposal of Yu *et al.* in 2018 [6] consists of the following steps:

- EMD is firstly applied to decompose the time series;
- the first IMF is excluded, since it mainly concentrates the high-frequency noise;
- ICA is applied to the set consisting of the original time series and the IMFs less correlated with the original time series.

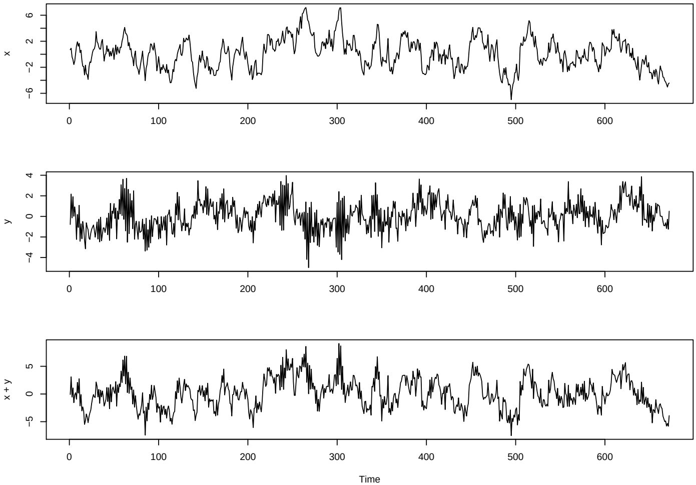
Independent Component Analysis (ICA) for single-channel signal

ICA can also be combined with EEMD.

Mijovic *et al.* in 2010 [4] proposed a technique that applies EEMD and ICA sequentially:

- EEMD is firstly applied to decompose the time series;
- ICA is applied to the set of all IMFs.

Application to a simulated time series

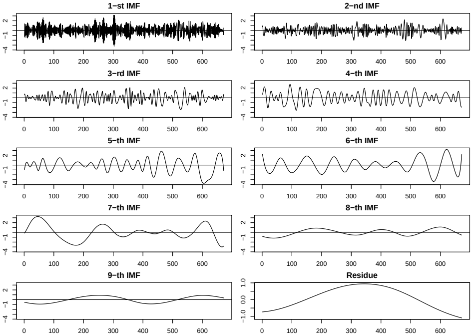


Simulated time series and their sum.

$$Z(t) = X(t) + Y(t) = 0.9X(t-1) + \varepsilon_X(t) + 0.7Y(t-2) + \varepsilon_Y(t)$$

where $\varepsilon_X \overset{d}{=} \varepsilon_Y \sim \mathcal{N}(0,1)$.

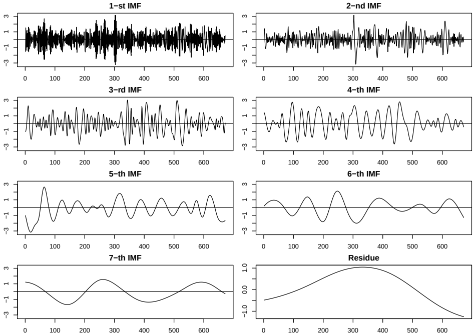
Application to a simulated time series - EMD decomposition



EMD decomposition of the simulated time series.

- 9 IMFs and the residue, which is a high number of components compared with the real one.
- The original components of the time series Z may not be represented by IMFs.
- However, there are many possible combinations of IMFs that can represent the original components of Z.

Application to a simulated time series - EEMD decomposition

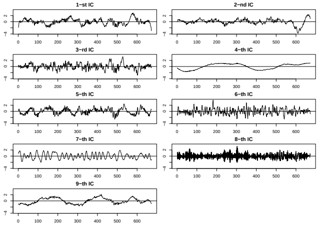


EEMD decomposition of the simulated time series.

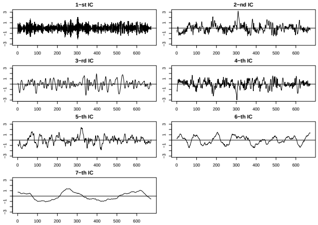
- 7 averaged IMFs and the residue, which is still a high number of components compared with the real one.
- There are also many possible combinations of IMFs that can represent the original components of the time series Z.

Application to a simulated time series - (E)EMD and ICA decomposition

- Applying ICA to the set of all IMFs obtained with EMD or EEMD, the number of Independent Components (ICs) is also high compared with the real number of components that exist in Z.
- However, the ICs can have a behaviour more similar to the real behaviour of the Z components.

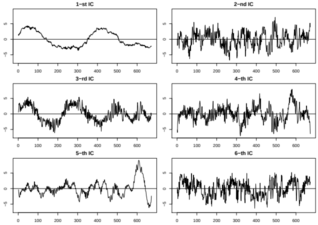


EMD and ICA decomposition of the simulated time series, considering the set of all IMFs.

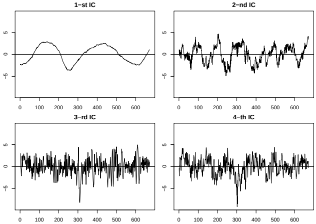


EEMD and ICA decomposition of the flow time series, considering the set of all averaged IMFs.

Application to a simulated time series - (E)EMD and ICA decomposition



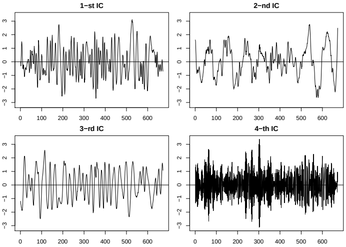
EMD and ICA decomposition of the flow time series, considering the set of the original time series and the IMFs less correlated with the original time series (2^{nd}, 5^{th}, 7^{th}, 8^{th} and 9^{th} IMFs).



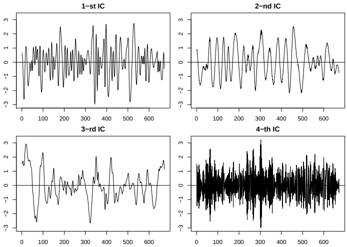
EEMD and ICA decomposition of the flow time series, considering the set of the original time series and the IMFs less correlated with the original time series (2^{nd}, 6^{th} and 7^{th} IMFs).

- Unfortunately, no component of Z was obtained when combinations of ICs were tested.

Application to a simulated time series - (E)EMD and ICA decomposition



EMD and ICA decomposition of the flow time series, considering the set of the IMFs more correlated with the original time series (1^{st}, 3^{rd}, 4^{th} and 6^{th} IMFs).



EEMD and ICA decomposition of the flow time series, considering the set of the IMFs more correlated with the original time series (1^{st}, 3^{rd}, 4^{th} and 5^{th} IMFs).

- Considering $A = IC_2 + IC_4$,
- Considering $B = IC_2 + IC_4$,

then

$$A(t) = 0.08A(t-1) + 0.70A(t-2) + 0.04$$

$$B(t) = 0.08B(t-1) + 0.70B(t-2) + 0.00$$

Application to a simulated time series - (E)EMD and ICA decomposition

$$A(t) = 0.08A(t-1) + 0.70A(t-2) + 0.04$$

$$B(t) = 0.08B(t-1) + 0.70B(t-2) + 0.00$$

Application to a simulated time series - (E)EMD and ICA decomposition

$$A(t) = 0.08A(t-1) + 0.70A(t-2) + 0.04$$
$$B(t) = 0.08B(t-1) + 0.70B(t-2) + 0.00$$

The original time series components were simulated as:

$$X(t) = 0.9X(t-1) + Z_X(t)$$

and

$$Y(t) = 0.7Y(t-2) + Z_Y(t)$$

where Z_X and Z_Y are white noise.

Then, Y can be recovered using these methods. This is the component with lower variance, which can be the reason why this was the only one that was recovered. More studies need to be run to obtain more conclusions and to recover the other component.

Concluding remarks

- This work presented a set of time series decomposition methods.
- Two approaches can be followed when the purpose is to decompose time series: the first approach decomposes the time series into its natural components (seasonality, trend and noise); the second one decomposes the time series into unusual components.
- STL is the most usual method for the first approach.
- The second approach has been widely applied in recent years with EMD and ICA the most used methods.
- Some of the methods presented in this work were applied to a simulated time series.
- The original component with lower variation was possible to recover combining (E)EMD and ICA.
- The original component with higher variation was not possible to recover with the methods and combinations tested.

Future work

- In the future, more simulations need to be run to obtain more knowledge about the components that can be recovered with these combinations of methods.
- Additionally, new combinations of methods need to be studied to recover the remain components.
- Moreover, these methods will be applied to flow time series. The decomposition results should be analysed to understand if they decompose the time series into metered consumption, unmetered consumption, base losses and pipe breaks.
- This decomposition is essential for water utilities to improve the water supply system management and to reduce water losses.

References I

Cleveland, R., Cleveland, W., and Terpenning, I.
STL: A seasonal-trend decomposition procedure based on loess.
Journal of Official Statistics 6, 1 (1990), 3.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H.
The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis.
In *Proceedings of the Royal Society of London A: Mathematical, Phisycal and Engineering Sciences* (1998), vol. 454, pp. 903–995.

Kevric, J., and Subasi, A.
Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system.
Biomedical signal processing and control 31 (2017), 398–406.

Mijovic, B., De Vos, M., Gligorijevic, I., Taelman, J., and Van Huffel, S.
Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis.
IEEE Transactions on biomedical engineering 57, 9 (2010), 2188–2196.

Wu, Z., and Huang, N. E.
Ensemble empirical mode decomposition: a noise-assisted data analysis method.
Advances in adaptive data analysis 1, 01 (2009), 1–41.

Yu, H., Guo, J., Kong, Q., and Chen, X.
Gravity tides extracted from relative gravimeter data by combining empirical mode decomposition and independent component analysis.
Pure and Applied Geophysics 175 (2018), 1683–1697.