Asociación Argentina
de Mecánica Computacional

Mecánica Computacional Vol XLII, pp. 1473-1481
A. Caggiano, G. Etse, P. Folino, M. Goldschmit, M. Pucheta, M. Storti (Eds.)
R. Lima, M. Piovan, G. Ruano, R. Sampaio (Issue eds.)
Buenos Aires, November 11-14, 2025

AN IMPROVED CAUSAL DECOMPOSITION PYTHON ALGORITHM WITH STATISTICAL CORROBORATION

Juan P. Muszkats[a,b], **Miguel E. Zitto**[b,c] **and Rosa Piotrkowski**[c,d]

[a]*Universidad Nacional del Noroeste de la Provincia de Buenos Aires, Roque Saenz Peña 456, Junín, Buenos Aires, Argentina, jpmuszkats@comunidad.unnoba.edu.ar*

[b]*Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Av. Medrano 951, Ciudad Autónoma de Buenos Aires, Argentina*

[c]*Facultad de Ingeniería, Universidad de Buenos Aires, Av. Paseo Colón 850, Ciudad Autónoma de Buenos Aires, Argentina*

[d]*Instituto de Tecnologías Emergentes y Ciencias Aplicadas (UNSAM-CONICET), Escuela de Ciencia y Tecnología, Av. 25 de mayo y Francia, San Martín, Buenos Aires, Argentina.*

Keywords: causal decomposition, EMD, complex systems

Abstract. Causal Decomposition based on Empirical Mode Decomposition (EMD) has proved to be a powerful tool for identifying causal relationships between time series. This method is based on the phase coherence of the respective oscillatory modes of the signals, known as Intrinsic Mode Functions (IMFs). Hence, a correct alignment of the respective modes of the signals is crucial. Unlike other methods, Causal Decomposition makes no assumption of linearity in the studied signals. Therefore, it is widely applicable to time series emerging from complex systems for which linearity hypothesis generally fail to hold. The decomposition in oscillatory modes is achieved with noise-assisted versions of EMD, which are known to improve the performance of the decomposition, reducing the mode mixing. However, adding noise introduces a stochastic element in the result, that is henceforth treated as a random variable. In the present work we introduce our Python version of the Causal Decomposition algorithm, which incorporates refinements for the selection of the decomposition based on energy considerations. These improvements aim to reduce the outlier results attributable to an incorrect mode alignment. The algorithm was tested on synthetic time series generated using a model of a mechanical oscillator with two masses and two modulated nonlinear forcing terms. A subsequent statistical analysis over multiple realizations showed less dispersion and fewer outliers compared to the previous version of the algorithm.

ISSN: 2591-3522 DOI: 10.70567/mc.v42.ocsid8221

1 INTRODUCTION

The seminal work by Huang et al. (1998) presented the idea of decomposing a signal in a data-driven fashion known as Empirical Mode Decomposition (EMD). Instead of traditional approaches such as Fourier analysis, in which the signal is projected into a predefined basis, EMD outputs a collection of Intrinsic Mode Functions (IMFs) which are directly obtained from the oscillations of the original signal. Hence, the resulting decomposition is expected to bear a more significant and physically sound relation to the data. Since then, EMD has undergone many developments in several directions, ranging from technical improvements of the basic algorithm to multivariate extensions. Among the former, we mention the noise-assisted versions of EMD. It was shown that the addition of noise to the signal and the subsequent average over multiple realizations emphasizes the dyadic filter properties of EMD (Flandrin et al., 2004). As to multivariate extensions, they allow the simultaneous decomposition of several signals. Among the variety of available techniques nowadays, we resorted to Noise-Assisted Multivariate EMD (NA-MEMD) (Rehman and Mandic, 2010, 2011).

There are several procedures for detecting and quantifying causality. The classical Granger causality, for instance, is a method that explores a linear process relating two time series. It quantifies a kind of predictive power and precedence relation between the variables. Other concepts of causality might be better suited when studying complex and highly non-linear systems. A recent proposal by Yang et al. (2018) has shown that EMD can be used to detect and quantify causal relations among signals. The method is based on the phase coherence of the signals under study. The influence of a particular oscillatory mode can be detected by comparison of phase coherence before and after its removal. Hence, the importance of a correct "mode alignment" of the respective IMFs in the studied signals. Given that phase coherence is calculated throughout the whole signal, the causality detected by the method does not imply a temporal precedence. Instead, it denotes an imbrication of the two signals which decreases when an oscillatory mode is absent. Therefore, this mode must somehow bear a causal effect on the other signal. We introduce our own Python version of the Causal Decomposition (CD) algorithm, adapted and translated from the original Matlab version (Yang, 2018). Given that decompositions are noise-assisted, each realization is slightly different from the others, making it necessary to perform a statistical analysis of the results. (Muszkats et al., 2024). Moreover, our new version of the algorithm does a previous selection of acceptable decompositions, leaving aside those that do not meet the mode alignment requirement. This improvement drastically reduces the number of outlier results.

The method is applied to synthetic series obtained from a classical model of a frictionless mechanical oscillator with two masses and non-linear external forcing. Given that the causal relations are known beforehand, this model provides a benchmark case for any causal detection and quantification scheme.

2 METHODS

2.1 Empirical Mode Decomposition

The basic EMD algorithm decomposes the input signal $x(t)$ into a sum of IMFs $c_j(t)$, plus a remainder term $r(t)$

$$x = \sum_{j=1}^{n} c_j + r \tag{1}$$

This process is known as *sifting* and usually requires several iterations until a stopping criterion is met. To be considered proper IMFs, the oscillatory modes must meet two basic requirements:

1. The number of extrema and the number of zeros must be the same, or at most differ by one.
2. The local maxima and minima determine their respective envelopes. The mean of these envelopes must be zero at any point.

Each IMF can be expressed as a cosine function modulated both in amplitude and phase:

$$c_j(t) = a_j(t) \cos\left[\varphi_j(t)\right] \tag{2}$$

This expression in turn allows a sensible definition of *instantaneous frequency* as the derivative of the phase: $\omega = \frac{d\varphi}{dt}$.

Among the various alternative and improved EMD algorithms since its first implementation, we adopted the Noise-Assisted Multivariate EMD (NA-MEMD). Multivariate techniques allow for the simultaneous decomposition of several signals, whereas noise assistance improves the filtering properties of the process. Both features greatly improve the results in Causal Decomposition. Instead of the original Matlab algorithm for NA-MEMD (Rehman and Mandic, 2011), we resourced to a Python translation (de Souza e Silva, 2018).

2.2 Causal Decomposition

The first step to establish a causal relation between two time series s_1 and s_2 is to define a measure of *phase coherence* between their respective IMFs s_{1j}, s_{2j}:

$$\text{coh}\left(s_{1j}, s_{2j}\right) = \frac{1}{T} \int_0^T e^{i\Delta\varphi_j(t)} dt \tag{3}$$

That is, the phase difference $\Delta\varphi_j(t) = \varphi_{2j}(t) - \varphi_{1j}(t)$ is summed all through the interval $[0, T]$ of the signals. If $\Delta\varphi_j(t)$ remains fairly constant, the result will approximate 1. If, instead, $\Delta\varphi_j(t)$ varies randomly, the sums will tend to cancel out and the result will approximate 0. Therefore, the closer phase coherence is to 1, the more related the signals are. To measure the influence that s_1 exerts upon s_2 at a particular IMF j, the jth IMF of s_2 is removed and the resulting signal is decomposed. The resulting IMFs are denoted s'_{2k}. If the jth IMF was influenced by s_1, phase coherence must diminish after removal. Hence, *causal strength* from s_1 over s_2 in the jth scale is defined as a weighted distance between coherences before and after the removal of the jth IMF:

$$D\left(s_{1j} \to s_{2j}\right) = \left\{ \sum_{k=1}^{n} \frac{\text{var}_{1k} \cdot \text{var}_{2k}}{\sum_{p=1}^{n} \text{var}_{1p} \cdot \text{var}_{2p}} \left[\text{coh}\left(s_{1k}, s_{2k}\right) - \text{coh}\left(s_{1k}, s'_{2k}\right)\right]^2 \right\}^{1/2} \tag{4}$$

where var_{ik} refers to the variance of the corresponding IMF. $D\left(s_{2j} \rightarrow s_{1j}\right)$ is defined analogously and both values are compared to decide if there is a differential causality.

For Eq. 4 to render reasonable results, the decomposition of both time series must be adequate in the sense of mode alignment. That is, the respective modes of both signals should be of similar frequency (at least on average). As already stated, mode alignment greatly improves by using multivariate decomposition. However, some decompositions could fail to meet this requirement, mostly after the removal of a particular IMF. It has been observed that some components of the signal "leak" to IMFs which were previously noisy low-amplitude components. To avoid these undesirable decompositions, we established energy thresholds that the IMFs must meet. For example, in Fig. 4 most of the energy (the sum of the squared components) of the original signals is allocated in IMFs $5, 6, 7$. Hence, every subsequent decomposition will be required to have at least 95% of its energy allocated in these same IMFs, or otherwise it will be discarded.

2.3 Statistical Analysis

Noise-assisted techniques imply that every realization of the decomposition is slightly different from the others. Hence, both $D\left(s_{1j} \rightarrow s_{2j}\right)$ and $D\left(s_{2j} \rightarrow s_{1j}\right)$ measurements are regarded as random variables. In the following, n observations of each one of these magnitudes will be abbreviated as $D_1, D_2, ...D_n$. Although the added noise is Gaussian, the resulting causal strengths undergo several transformations, and their distribution does not necessarily result normal. Without any further assumption about their actual distribution, the random variables $D_1, D_2, ...D_n$ can be safely treated as independent and identically distributed, because they emerge from different realizations of the same process. Their mean and variance will be respectively denoted μ and σ^2. What we are dealing with is their mean value $\bar{D}$ after n realizations of the process. If n is sufficiently large, this mean value is a new random variable derived from a large number sample. In these circumstances, the Central Limit Theorem implies that the distribution of $\bar{D}$ tends to be normal (Devore, 2009) with parameters $\mu_{\bar{D}} = \mu$ and $\sigma^2_{\bar{D}} = \sigma^2/n$. Given that the purpose of this work is to establish a differential causality between $D\left(s_{1j} \rightarrow s_{2j}\right)$ and $D\left(s_{2j} \rightarrow s_{1j}\right)$, we resourced to confidence intervals and hypothesis testing.

The confidence intervals for the sample mean with an approximate $100(1-\alpha)\%$ confidence level are

$$\left(\bar{d} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{d} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right) \tag{5}$$

where $\bar{d}$ is the actual sample mean, $z_{\alpha/2}$ is the value for which the normal standard distribution accumulates $1 - \alpha/2$ of the probability and s is the sample standard deviation.

The test of hypothesis is applied to a new random variable that measures the differential causality $D\left(s_{1j} \rightarrow s_{2j}\right) - D\left(s_{2j} \rightarrow s_{1j}\right)$ for each pair of decompositions. Once again, the Central Limit Theorem guarantees that the mean value of large samples of this variable tends to have a normal distribution. The null hypothesis H_0 is that $D\left(s_{1j} \rightarrow s_{2j}\right) - D\left(s_{2j} \rightarrow s_{1j}\right) = 0$ whereas the alternative hypothesis H_a is that $D\left(s_{1j} \rightarrow s_{2j}\right) - D\left(s_{2j} \rightarrow s_{1j}\right) > 0$. The test statistic is

$$Z = \frac{\bar{d} - 0}{s/\sqrt{n}} \tag{6}$$

where $\bar{d}$ and s are, respectively, the sample mean and standard deviation of the variable $D\left(s_{1j} \to s_{2j}\right) - D\left(s_{2j} \to s_{1j}\right)$. Considering an α significance level, the null hypothesis is rejected if the test statistic results greater than z_α (the value for which the normal standard distribution accumulates $1 - \alpha$ of the probability).

3 SYSTEM DESCRIPTION

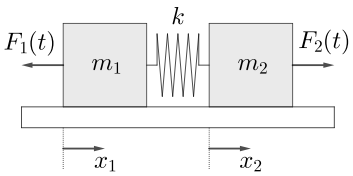

Figure 1: Mechanical oscillator with external forcing. The coordinates x_1 and x_2 measure displacement from equilibrium.

The CD algorithm is tested on the synthetical series emerging from the classical mechanical oscillator shown in Fig. 1. The governing equations model an undamped system with a linear spring:

$$\begin{aligned} m_1 \cdot x_1'' &= k\left(x_2 - x_1\right) + F_1(t) \\ m_2 \cdot x_2'' &= -k\left(x_2 - x_1\right) + F_2(t) \end{aligned} \tag{7}$$

The natural (angular) frequency of the system is $\omega_0 = \sqrt{\frac{k}{m_1} + \frac{k}{m_2}}$. The parameters are chosen so that the period for the free vibrations of the system is approximately 1:

$$m_1 = m_2 = 1 \qquad k = 20 \tag{8}$$

The external forces are modulated both in frequency and amplitude. As shown in Fig. 2, they behave like a non-linear influence that moves around a central frequency. The parameters are chosen so that F_1 is the high-frequency forcing, with a carrier frequency that doubles the natural frequency of the system, whereas the carrier frequency of F_2 halves it. In other words, F_1 has a period that fluctuates about 0.5, and F_2 has a period that fluctuates about 2.

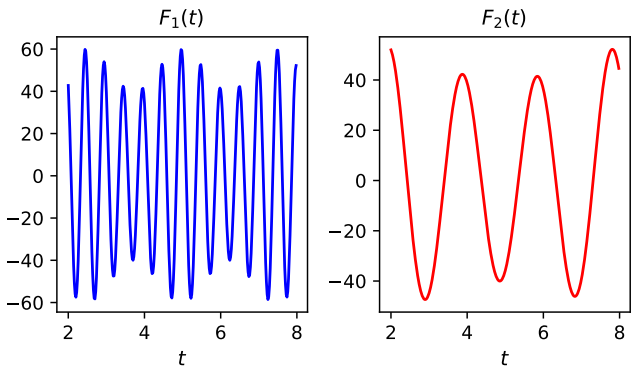

Figure 2: Forces are modulated both in amplitude and frequency. F_1 has a carrier frequency that doubles the natural frequency: $F_1(t) = \left[50 + 10\cos\left(0.4\omega_0 t\right)\right]\cos\left[2\omega_0 t + 0.5\sin\left(0.1\omega_0 t\right)\right]$. The carrier frequency of F_2 is half the natural frequency: $F_2(t) = \left[50 + 10\cos\left(0.1\omega_0 t\right)\right]\cos\left[0.5\omega_0 t + 0.5\sin\left(0.025\omega_0 t\right)\right]$

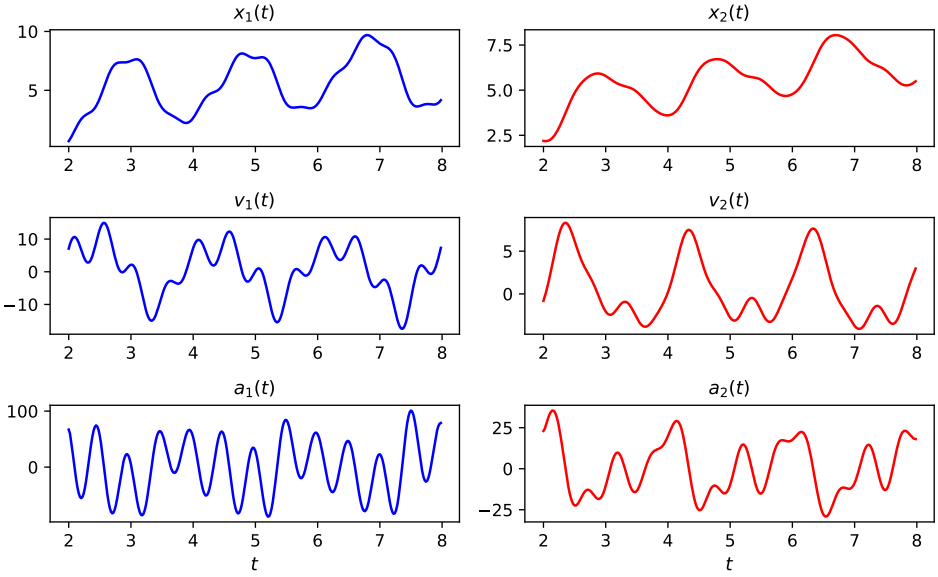

Figure 3: Numerical solution of the system. Resulting positions, velocities and accelerations.

4 RESULTS

The differential equations are solved numerically with a Runge-Kutta scheme, and the results are plotted in Fig. 3. The resulting accelerations exhibit high and low-frequency oscillations attributable to the respective forcings. These influences become patent once the signals are decomposed. As seen in Fig. 4, a_1 has a high-frequency component that neatly follows F_1. On the other hand, the influence of the low-frequency component F_2 seems to be milder. This fact is attributable to its indirect effect, mediated by the spring.

IMF	$D\left(F_1 \to a_1\right)$	$D\left(a_1 \to F_1\right)$	s $\left[D\left(F_1 \to a_1\right)\right]$	s $\left[D\left(a_1 \to F_1\right)\right]$
5	0.575	0.410	0.154	0.152
6	0.010	0.007	0.040	0.022
7	0.007	0.010	0.019	0.039

Table 1: Mean causal force and standard deviation s after 200 realizations. No energy considerations were taken into account. Notice that, as shown in Fig. 4, IMF 5 represents the high frequency component of the system (period 0.5), IMF 6 the free vibrations (period 1), and IMF 7 the low frequency (period 2).

IMF	$D\left(F_1 \to a_1\right)$	$D\left(a_1 \to F_1\right)$	s $\left[D\left(F_1 \to a_1\right)\right]$	s $\left[D\left(a_1 \to F_1\right)\right]$
5	0.609	0.450	0.115	0.136
6	0.002	0.005	0.001	0.002
7	0.004	0.004	0.002	0.002

Table 2: Mean causal force and standard deviation s after 200 realizations with the improved algorithm. Those decompositions in which relative energies depart from prescribed values were discarded. IMF 5 represents the high frequency component of the system (period 0.5), IMF 6 the free vibrations (period 1), and IMF 7 the low frequency (period 2).

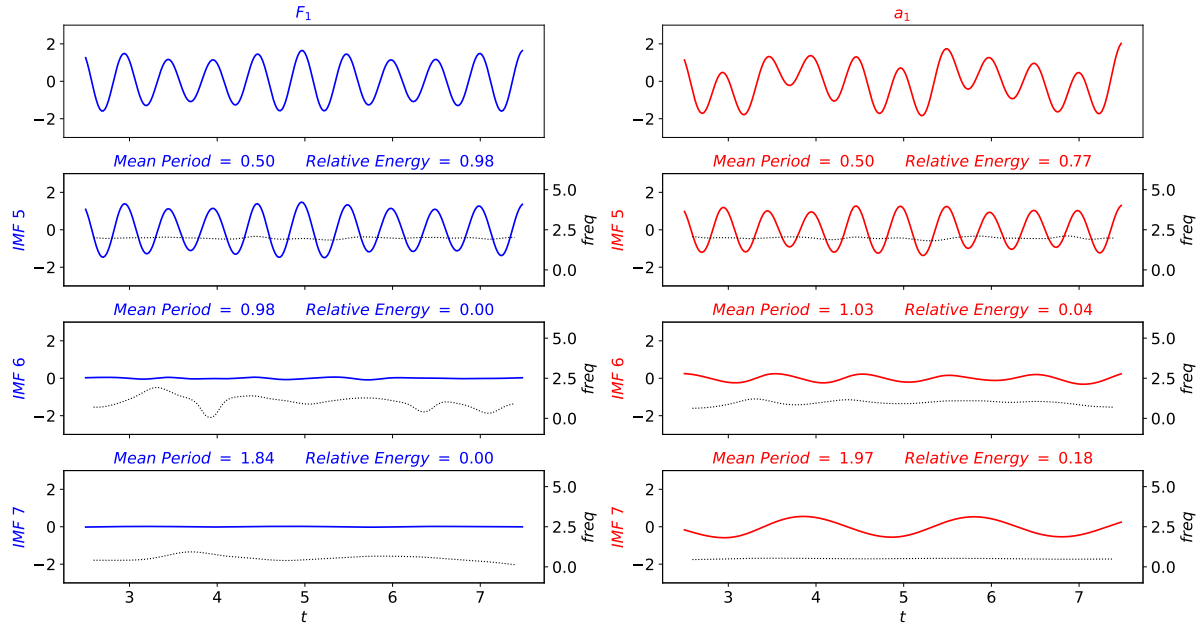

Figure 4: Multivariate (simultaneous) decomposition of F_1 and a_1. Some IMFs are omitted because they are low amplitude components attributable to numerical artifacts. Only the relevant IMFs (according to their energy) are shown. IMF 5 has a stable instantaneous frequency of about 2. It is associated with the high-frequency forcing and a_1 neatly reflects this fact. The influence of the other mass and force is observable in the rest of the a_1 components.

As expected, Causal Decomposition detects a differential causality between the respective IMFs 5 of F_1 and a_1. Table 1 shows the mean results after 200 realizations without energy considerations. Results exhibit a relatively high deviation due to outlier values, as explicitly seen in the boxplot of Fig. 5. Table 2 shows the same calculations with the improved algorithm, considering the relative energy of each IMF.

Finally, the differential causality was checked both with confidence intervals and hypothesis testing (after energy considerations). The confidence intervals in Fig. 6 exhibit a clear divide between $D\left(F_1 \to a_1\right)$ and $D\left(a_1 \to F_1\right)$, with a 99% confidence level. Moreover, hypothesis testing allows us to affirm that $D\left(F_1 \to a_1\right) - D\left(a_1 \to F_1\right) > 0$ with a 0.01 significance level. Therefore, we can be confident that force bears a differential causality upon acceleration.

Although not shown in detail, the influence of F_2 over a_2 parallels the already seen results. F_2 was neatly proven to bear a differential causality upon a_2. This relation expresses itself in the low-frequency component, that is, in IMF 7.

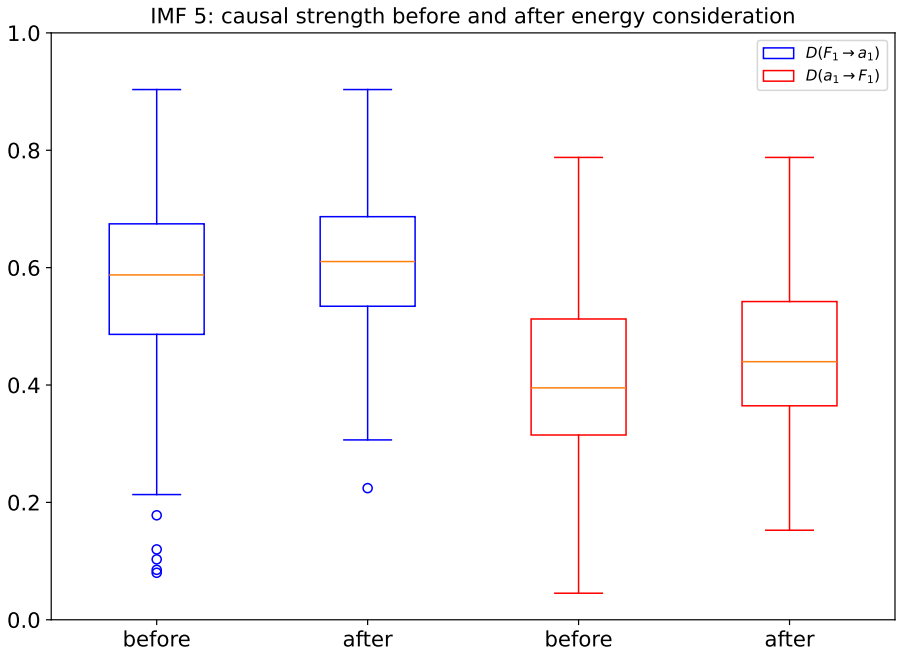

Figure 5: Comparison of results before and after energy considerations. The box signals the first and third quartiles, with an orange line for the median. The "whiskers" extend up to 1.5 of the interquartile range. Separate dots represent outliers, which happen to be quite common if no energy considerations are taken into account. The blue plots represent the distribution of $D\left(F_1 \to a_1\right)$, while the red ones stand for $D\left(a_1 \to F_1\right)$. In both cases dispersion reduces after energy considerations.

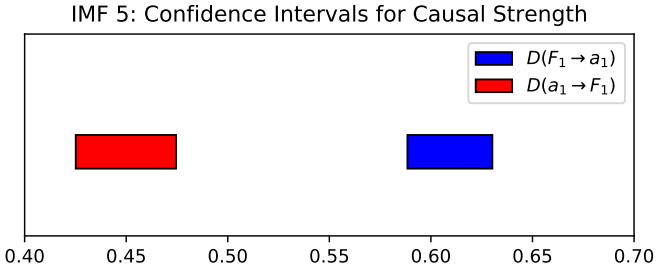

Figure 6: Confidence intervals for a large sample ($n = 200$) of causal strengths (with energy considerations). The confidence level is 99% and shows a neat differential causality of F_1 over a_1.

5 CONCLUSIONS

Given that causal relations are known beforehand, the mechanical oscillator has proven to be a reliable benchmark to test the original method and its improved version with energy considerations. While both methods detected the causal relations, the improved algorithm produced fewer outliers and a more robust result. The new method, based on an energy threshold criterion, discards decompositions leading to comparisons that make no sense.

We have restricted our study to the relation between a force and its nearest mass because it provides a straightforward example of causality and illustrates the power of the method. However, it is possible to study other interactions among a_1, a_2, F_1, and F_2. The relation between a_1

and a_2, for instance, conveys no obvious causality and therefore requires a subtler interpretation. These interactions resemble those in complex systems, where time series represent interwoven phenomena.

In light of these promising results, the improved algorithm will in future research be applied to time series emerging from complex interactions, such as the climate system.

ACKNOWLEDGEMENTS

This work has financial support and is part of the UBACyT project 20020220400001BA.

REFERENCES

de Souza e Silva M. Memd-python. https://github.com/mariogrune/MEMD-Python-, 2018. Accessed: 2025-07-06.

Devore J.L. *Probability and Statistics for Engineering and the Sciences*. CENGAGE Learning, 2009. ISBN 978-0-495-55744-9.

Flandrin P., Rilling G., and Goncalves P. Empirical mode decomposition as a filter bank. *IEEE signal processing letters*, 11(2):112–114, 2004. http://doi.org/10.1109/LSP.2003.821662.

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. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 454(1971):903–995, 1998. http://doi.org/10.1098/rspa.1998.0193.

Muszkats J., Muszkats S., Zitto M., and Piotrkowski R. A statistical analysis of causal decomposition methods applied to earth system time series. *Physica A: Statistical Mechanics and its Applications*, 641:129708, 2024. ISSN 0378-4371. http://doi.org/10.1016/j.physa.2024.129708.

Rehman N. and Mandic D.P. Multivariate empirical mode decomposition. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 466(2117):1291–1302, 2010. http://doi.org/10.1098/rspa.2009.0502.

Rehman N. and Mandic D.P. Filter bank property of multivariate empirical mode decomposition. *IEEE Trans. on Signal Process.*, 59(5):2421–2426, 2011. http://doi.org/10.1109/TSP.2011.2106779.

Yang A.C. Causal decomposition analysis. https://github.com/accyang/causal-decomposition-analysis, 2018. Accessed: 2025-07-06.

Yang A.C., Peng C.K., and Huang N.E. Causal decomposition in the mutual causation system. *Nat. Commun.*, 9(1):3378, 2018. http://doi.org/10.1038/s41467-018-05845-7.