Empirical Discovery of Multi-Scale Transfer of Information in Dynamical Systems

Christopher W. Curtis[1] and Erik M. Bollt[2,3]

[1]Department of Mathematics and Statistics , San Diego State University, 5500 Campanile Dr., San Diego, CA, 92182
[2]Department of Electrical and Computer Engineering, Clarkson University, 8 Clarkson Avenue, Potsdam, New York 13699, USA
[3]Clarkson Center for Complex Systems Science, Clarkson University, 8 Clarkson Avenue, Potsdam, New York 13699, USA

Abstract

In this work, we quantify the timescales and information flow associated by multiscale energy transfer in a weakly turbulent system through a novel new interpretation of transfer entropy. Our goal is to provide a detailed understanding of the nature of complex energy transfer in nonlinear dispersive systems driven by wave mixing. Further, we present a modal decomposition method based on the empirical wavelet transform that produces a relatively small number of nearly decorrelated, scale separated modes. Using our method, we are able to track multiscale energy transfer using only scalar time series measurements of a weakly turbulent system. This points to our approach being of broader applicability in real-world data coming from chaotic or turbulent dynamical systems.

1 Introduction

The question of causality, or perhaps more broadly information flow and coupling, in time series is a central one. By addressing the question in linear time series coming from econometric data, Clive Granger famously won a Nobel prize in 2003. Building off of this ground-breaking work, methods using information theory to determine significant couplings between variables in nonlinear time series have been developed; see in particular [1] which introduced the metric of *transfer entropy* (also called conditional mutual information). Furthermore, we have shown that conditioning on tertiary effects by what we called causation entropy (CE) [2, 3, 4, 5], allows for an effective means of identifying causal chains across large numbers of measured variables, by an algorithm that we called optimal causation entropy (oCSE), thereby accurately generating networks of information flow among multiple time series. Readily available and dedicated software libraries, such as IDTxl [6, 7], now make the generation of these networks increasingly straightforward.

However, the question of information flow in the physical sciences is still a relatively unexplored and immature topic. For example, very recent work in [8] shows how transfer entropy can provide a more sophisticated understanding of the measurement of energy cascades of fluid turbulence. Likewise, studies of information flow in chaotic and turbulent dynamical systems have appeared with regard to modeling error quantification and fluctuation-dissipation methods have appeared; see [9] and related work. Preliminary work exploring how information theory helps describe atmospheric and ionospheric dynamics has appeared in [10, 11]. Nevertheless, motivated by this existing work, much remains to be explored in this area.

Therefore in this work, we explore how information theory is able to track multiscale energy transfer in the Majda-Mclaughlin-Tabak (MMT) model [12]. This model is particularly interesting since despite its relative simplicity of being only a 1+1 dimensional nonlinear dispersive wave equation, it is known to exhibit weak-wave turbulence (WWT) [12, 13]. For the MMT model, both forward and inverse cascades are present. Using then a modification of the measurements of energy transfer in [8], we track the most significant energy transfer across scales using the IDTxl library. This is a nontrivial task since recent results from [14, 15] have shown that while WWT can be characterized by a statistically stationary average energy distribution, energy is not moved in a directly cascading way but instead transported in a more intricate fashion via multi-wave mixing. Our results further illustrate this point, though they also detect a relatively clear dichotomy in which forward energy transfer typically proceed at a markedly faster rate than inverse cascades. However, fast inverse transfers do occur, potentially illustrating the point of recent work exploring the complexity of multiscale energy transfer in wave-mixed systems [15].

We also address in this work the question of how we might detect multiscale energy transfer from limited measurements. This is a foundational question in the physical sciences where full multidimensional resolution of complex processes is rarely available outside controlled laboratory conditions. Our approach to answering this dilemma is to use an extension of the empirical wavelet transform (EWT) as developed by [16]. In our modification, we use Otsu's method [17] to find a pre-selected number of optimal separations of a signal in frequency space. This then produces a limited number, again chosen by the user, of nearly time decorrelated, scale separated modes. We call this tool the Otsu EWT (OEWT). With the OEWT in hand then, using the IDTxl library, we look at information transfer across the scale separated components which result from the OEWT method. While the couplings are not as intricate as when we have access to more sophisticated measurements of MMT dynamics, we are nevertheless able to still capture multiscale energy transfer thereby showing our approach allows for the detection of cascade phenomena in otherwise limited, scalar measurements.

The present work then provides a unique methodology for analyzing chaotic up to turbulent time series and gives insight into the complexity of stationary cascade formation in multi-wave mixing systems. We have shown both the utility of using transfer entropy to characterize multiscale coupling and information flow in a new context, and we have also developed a new and convenient multiscale decomposition method for tracking information flow from scalar time series. Natural next questions for this work are how it performs in more classically turbulent problems coming from fluid mechanics, and how it fares with noisy and incomplete real world measurement. These are both questions of active research by our group.

The structure of the paper is as follows. In Section 2, we present an explanation of transfer entropy and the algorithm underlying the IDTxl library. We likewise look at a typical example of its use. We then present our first results on WWT in the MMT model. In Section 3, we present development of the OEWT method, and then show how it can be used to detect energy transfer in the MMT model using only a scalar time series. In Section 4, we provide summary discussion and suggest several further directions of research.

2 Determining Information Flow through Transfer Entropy

Given a multidimensional time series, $\{\mathbf{x}_j\}_{j=1}^{N_T}$, with $\mathbf{x}_j \in \mathbb{R}^m$ with vector components denoted as $x_{k,j}$, it is a basic question to determine the extent to which a time series along one dimension *causes*, or more broadly *informs*, another. Motivated by the now celebrated *Granger causality* test, *cf.* [18], in linear time series, [1] introduced the notion of *transfer entropy* (TE) to determine the causal relationship between two time series. The TE from $x_{l,j}$ to $x_{k,j}$, say $T_{x_l \to x_k}(j)$ is defined in [1] to be

$$T_{x_l \to x_k} = H\left(x_{k,j+1} \middle| \, x_{k,j}\right) - H\left(x_{k,j+1} \middle| \, x_{k,j}, x_{l,j}\right) \equiv I(x_{k,j+1}, x_{l,j} | x_{k,j}),$$

where $H(Y|X)$ is the conditional entropy between two random variables X and Y defined as

$$H(Y|X) = \int p(y,x) \log p(y|x) dx dy.$$

Note, if $x_{k,j+1}$ is independent of $x_{l,j}$, then $H(x_{k,j+1}|x_{k,j}, x_{l,j}) = H(x_{k,j+1}|x_{k,j})$ so that $T_{x_l \to x_k} = 0$.

This initial concept of transfer entropy has given rise to a host of modifications and improvements, see in particular [19] and [2], which has ultimately lead to sophisticated software libraries being developed which can determine networks of interactions between time series that accurately account for confounding variables and non-Markovian influences of past states. In this work, we use the library [6] given its wide modeling capabilities and relatively rigorous hypothesis testing features.

The backbone of the method couples the power of non-uniform embeddings of time series [20, 21, 19], with greedy-algorithm optimization routines which seek out those time series models which provide the most transfer entropy. The algorithm generates two models. One is for *sources* in which we find the maximum information flow to $x_{k,j+1}$ from $x_{k,\boldsymbol{\ell}_s}$, where $\boldsymbol{\ell}_s$ represents an optimal choice of some u lags, say $\boldsymbol{\ell}_s = (\ell_1, \cdots, \ell_u)$ so that

$$x_{k,\boldsymbol{\ell}_s} = (x_{k,\ell_1}, \cdots, x_{k,\ell_u}).$$

The other model the method generates is for *targets* across all complimentary dimensions say $\mathbf{x}_{\mathbf{k}_c,\boldsymbol{\ell}_t}$ where

$$\mathbf{x}_{\mathbf{k}_c,\boldsymbol{\ell}_t} = \left\{(x_{l,\ell_{l,1}}, \cdots, x_{l,\ell_{l,u_l}})\right\}_{l\neq k}.$$

The choice of target lags can vary from target dimension to target dimension, and thus the algorithm is able to find sophisticated non-uniform time embeddings in order to determine information flow within multi-dimensional time series. Each model generation consists of two phases, the first being a BUILD phase, the second being a PRUNE phase. Throughout, we also track the transfer entropy for each chosen lag between dimensions say l and k, which for a given chosen lag ℓ_{ch} we denote as $T_{l\to k}(\ell_{ch})$. We then define $\ell_{ch,*}$ so that

$$\ell_{ch,*} = \arg\max_{\ell_{ch}\in\boldsymbol{\ell}_t} T_{l\to k}(\ell_{ch}).$$

and $T^M_{l\to k}(\ell_{ch,*}) = T_{l\to k}(\ell_{ch,*})$. All of these processes are summarized in Algorithm 1; for full details see [7].

Note, while for brevity we only report the lag $\ell_{ch,*}$ which gives the largest target to source transfer entropy, i.e. $T^M_{l\to k}(\ell_{ch,*})$, there are still other lagged versions of the target which significantly contribute information to the source dynamics. In part, the difficulty of reporting results for this method is a reflection of the underlying greedy-algorithm. This means that we can only report results relative to their appearance in a particular run of the method. See [7] and [22] for further details on this point.

To briefly explore the use of IDTxl and its related issues, we study a common problem from the affiliated literature, which is the coupled Lorenz–

Algorithm 1 IDTxl Algorithm

for Dimension k **do**
    procedure GENERATE SOURCE MODEL FOR $x_{k,j}$
        INITIALIZE: Set $\boldsymbol{\ell}_s^{(k)} = \{\varnothing\}$, $\boldsymbol{\ell}_r = \{1 \cdots d\}$.
        procedure BUILD
            while $\boldsymbol{\ell}_r \neq \{\varnothing\}$ **do**
                Given $\boldsymbol{\ell}_s = \{\ell_1 \cdots \ell_c\}$ and $\boldsymbol{\ell}_r = \{1, \cdots, d\} \setminus \boldsymbol{\ell}_s$
                $\ell_* \leftarrow \arg\max_{\ell_{c+1} \in \boldsymbol{\ell}_r} I\left(x_{k,j+1}, x_{k,\boldsymbol{\ell}_s^{(k)} \cup \ell_{c+1}} \middle| x_{k,\boldsymbol{\ell}_s^{(k)}}\right)$
                if ℓ_* is statistically significant **then**
                    $\boldsymbol{\ell}_s^{(k)} \leftarrow \boldsymbol{\ell}_s^{(k)} \cup \{\ell_*\}$
                end if
            end while
        end procedure
        procedure PRUNE
            INITIALIZE: Set $S \equiv$ True
            while S **do**
                $\tilde{\ell}_* \leftarrow \arg\min_{l_c \in \boldsymbol{\ell}_s^{(k)}} I\left(x_{k,j+1}, x_{k,\boldsymbol{\ell}_s^{(k)} \setminus \ell_c} \middle| x_{k,\boldsymbol{\ell}_s^{(k)}}\right)$
                if $\tilde{\ell}$ is statistically insignificant **then**
                    $\boldsymbol{\ell}_s^{(k)} \leftarrow \boldsymbol{\ell}_s^{(k)} \setminus \left\{\tilde{\ell}_*\right\}$
                else
                    $S \equiv$ False
                end if
            end while
            RETURNS: $\boldsymbol{\ell}_s$
        end procedure
    end procedure
    procedure GENERATE TARGET MODEL FOR $x_{k,j}$
        for Dimension $l \neq k$ **do**
            INITIALIZE: Set $\boldsymbol{\ell}_t^{(l)} = \{\varnothing\}$, $\boldsymbol{\ell}_r = \{1 \cdots d\}$, $T_{l \to k}(\ell_{ch}) = 0$.
            procedure BUILD
                Build (as above) $\boldsymbol{\ell}_t^{(l)}$ from $x_{l,j}$ conditioned on $\boldsymbol{\ell}_s$.
                Compute $T_{l \to k}(\ell_{ch})$ for $\ell_{ch} \in \boldsymbol{\ell}_t^{(l)}$.
            end procedure
            procedure PRUNE
                Prune (as above) $\boldsymbol{\ell}_t^{(l)}$ conditioned on $\boldsymbol{\ell}_s^{(k)}$.
                RETURNS: $\boldsymbol{\ell}_t^{(l)}$, $T_{l \to k}^{M}(\ell_{ch,*})$
            end procedure
        end for
    end procedure
    RETURNS: $\boldsymbol{\ell}_s^{(k)}$, $\cup_{l \neq k} \boldsymbol{\ell}_t^{(l)}$, $\left\{T_{l \to k}^{M}(\ell_{ch,*})\right\}_{l \neq k}$
end for

Rössler system of the form

$$\begin{aligned}
\dot{x}_0 =& \sigma(x_1 - x_0)\\
\dot{x}_1 =& x_0(\rho - x_2) - x_1 + Cx_4^2\\
\dot{x}_2 =& x_0x_1 - \beta x_2\\
\dot{x}_3 =& -6(x_4 + x_5)\\
\dot{x}_4 =& 6(x_3 + \alpha x_4)\\
\dot{x}_5 =& 6(\gamma + x_5(x_3 - \delta))
\end{aligned}$$

Here we let $\sigma = 10$, $\rho = 28$, $\beta = 8/3$, $\alpha = .2$, $\gamma = .2$, $\delta = 5.7$. C can be varied so as to enhance the driving effect of the Rössler system on the Lorenz system, though the effect of this can be surprising, especially when looked at over the whole network; see Figure 1 for details. See Throughout our tests, we use trajectories found via a 4th-order Runge–Kutta scheme using a time step of $\delta t = .01$ run out to a total time of 150 units of non-dimensional time. The first 100 units of time are ignored so as to remove any transient phenomena from our data set.

To compute the TE/CMI, we use nearest-neighbor estimators developed in [23], which we label the KSG estimator. While one of the most popular choices for estimators, we note that there are small pathological quirks that must be managed. In particular, each stage of the IDTxl method has an affiliated significance test and a corresponding p-value which is set to $p = .05$. In the PRUNE phase, the smallest values of I are typically of the order of 10^{-3}, and the use of the KSG estimator often leads to negative values of conditional mutual information. This should be theoretically impossible, and thus it is a consequence of the estimation technique. What to do with these very small but negative values is not entirely clear, but we have found that automatically rejecting them as significant leads to the best results by minimizing false-positive links.

Setting the coupling $C = 1$, letting the maximum lag in time be $d = 4$, and normalizing the data to have zero average and unit variance, we get the result in Figure 2. As we can see, the flow of information largely moves as we would expect. There is a false positive link from x_4 to x_2, albeit lagged behind the correct link between x_4 and x_1. Thus, the method struggles to not confound links across different time lags, though we note that $T^M_{41}(1) = .3058$ while $T^M_{42}(2) = .02332$, so that the transfer entropy corresponding to the coupling link between the systems is ten times larger than the false positive. We also then could stand to have a more stringent hypothesis test in place, though the computational overhead that results is significant. Nevertheless, we see our results are very good, with the method even capturing the more multi-scale nature of the Rössler system by way of the greater difference in lag values throughout dimensions x_3, x_4, and x_5.

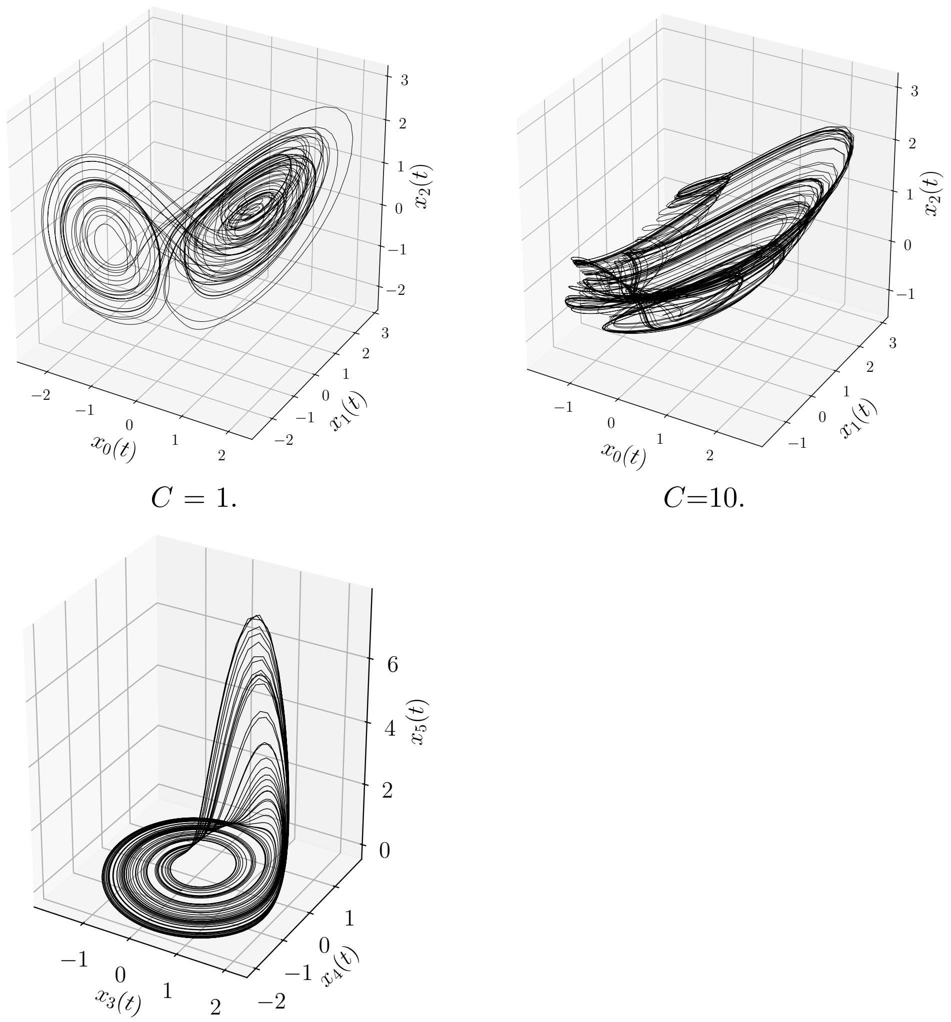


Figure 1: Plot of Lorenz–Rössler system. Top Figures: Lorenz dynamics for $C = 1$ and $C = 10$. Bottom Figure: Rössler dynamics.

2.1 Tracking Multiscale Energy Transfer in a Weakly Turbulent System via Information Theory

We now explore using the IDTxl library on data coming from the Majda-McLaughlin-Tabak (MMT) model [12]. The particular MMT model we study is of the form

$$i\partial_t \psi = |\partial_x|^{1/2} \psi - |\psi|^2 \psi + i\epsilon^2 \left(f - \left(\frac{|\partial_x|}{k_+} \right)^{d_+} - \left(\frac{k_-}{|\partial_x|} \right)^{d_-} \right) \psi,$$

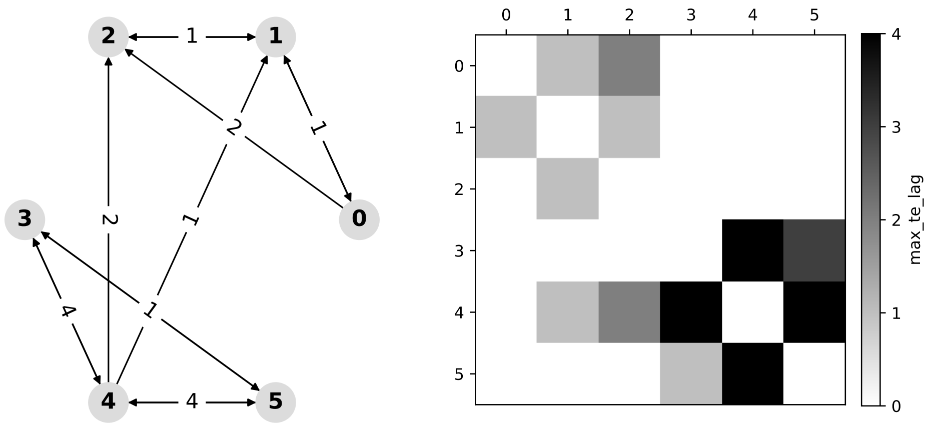


Figure 2: Left Figure: Maximum lag coupling between dimensions in the coupled Lorenz–Rössler system with $C = 1$. Right Figure: Time corresponding to the optimal lags for the maximum transfer entropy between dimensions in the coupled Lorenz–Rössler system. The hypothesis testing threshold is $p = .05$.

where the forcing f is defined so that

$$\widehat{f\psi}(k,t) = \left(\sum_{j=1}^{n} \hat{\delta}(k-k_j)\right) \hat{\psi}(k,t), \; k_l \leq k_j \leq k_h,$$

where $\hat{\delta}(k) = 1$ for $k = 0$ and is zero otherwise. The range of wave numbers between k_l and k_h define the *forcing* regime. Likewise, we damp long waves for $|k| < k_-$ and short waves for $|k| > k_+$. Those wave numbers that are sufficiently greater than k_h but smaller than k_+ define the *inertial range.*

Our interest then in this model comes from the fact that it is a *weakly turbulent* system, which means that it generates spatio-temporally chaotic dynamics which, in a properly identified inertial range, can be described by a mean energy cascade profile. This means that by defining

$$n(k,t) = \left\langle \left|\hat{\psi}(k,t)\right|^2 \right\rangle,$$

one can show [13] in the long time limit that $n(k,t) \to C|k|^{-1}$. Within this equilibrium distribution, we should anticipate both inverse and forward cascades by looking at the *particle number* and *energy*, given respectively by the sums

$$\sum_k n(k,t), \; \sum_k |k|^{1/2} n(k,t).$$

Both are otherwise conserved quantities in the unforced and undamped case, and so within the inertial range, they explain the limiting tendency towards a statistically steady state.

However, as explored in [15] and [14], the process by which statistically stationary conditions is achieved is intricate. One can see this by ignoring forcing and damping, which is appropriate within the inertial regime, and then passing to a Fourier representation of the MMT model written as

$$i\partial_t \hat{\psi}(k,t) = |k|^{1/2}\hat{\psi}(k,t) + \sum_{k_1,k_2,k_3} \hat{\psi}_1(t)\hat{\psi}_2(t)\hat{\psi}_3^*(t)\delta(k_1+k_2-k_3-k),$$

where $\hat{\psi}_j(t) = \hat{\psi}(k_j,t)$. Defining $\omega_j = |k_j|^{1/2}$ and using the substitution $\hat{\phi}_j(t) = \hat{\psi}_j(t)e^{i\omega_j t}$, we get the equivalent system

$$i\partial_t \hat{\phi}(k,t) = \sum_{k_1,k_2,k_3} \hat{\phi}_1(t)\hat{\phi}_2(t)\hat{\phi}_3^*(t)\delta(k_1+k_2-k_3-k)e^{-i(\omega_1+\omega_2-\omega_3-\omega)t}.$$

Thus, in the long time limit, a stationary phase argument shows us that those wave numbers that lead to, or nearly to, *4-wave mixing*, i.e.

$$k_1+k_2-k_3-k=0,\ \omega_1+\omega_2-\omega_3-\omega=0,$$

drive the process of convergence and maintenance of a statistically steady state. Therefore, we can have significant multiscale energy transfer across otherwise widely separated scales. This greatly complicates the question of tracking information flow, and having some quantitative sketch of this process is of interest.

Throughout the remainder of this work, we always choose the initial condition

$$\hat{\psi}(k,0) = \frac{\epsilon}{|k|}\hat{z}_k,\ \hat{z}_{k,r/i} \sim \mathcal{N}(0,1)$$

and parameters

$$k_l = 6,\ k_h = 9,\ d_- = d_+ = 8,\ k_- = 5,\ k_+ = 1000,\ \epsilon = .5.$$

We take the inertial range to be $50 < k < 500$. We fix the space domain to be $[0,2\pi]$. Following the analysis in [12], per our choice of ϵ, the nonlinearity acts over time scales on the order of $1/\epsilon^2 = 4$ non-dimensional units of time. Using a pseudo-spectral in space and 4th order Runge–Kutta in time discretization scheme, we generate data up to $t_f = 4k_+/\epsilon^2$, thereby allowing for nonlinearity to induce several turnovers of energy within the inertial range; see Figure 3 for a plot of $|\psi(x,t)|$ for $2k_+/\epsilon^2 < t < 2k_+/\epsilon^2 + 160$. We keep the last $t_{kp} = 2k_+/\epsilon^2$ length of data, sampled at a rate of $\delta_s = .2$ units of non-dimensional time.

Averaging over t_{kp}, we generate the following approximation of $n(k)$ seen in Figure 4 Thus we see that we are generating dynamics consistent with the time and length scale requirements in WWT theory.

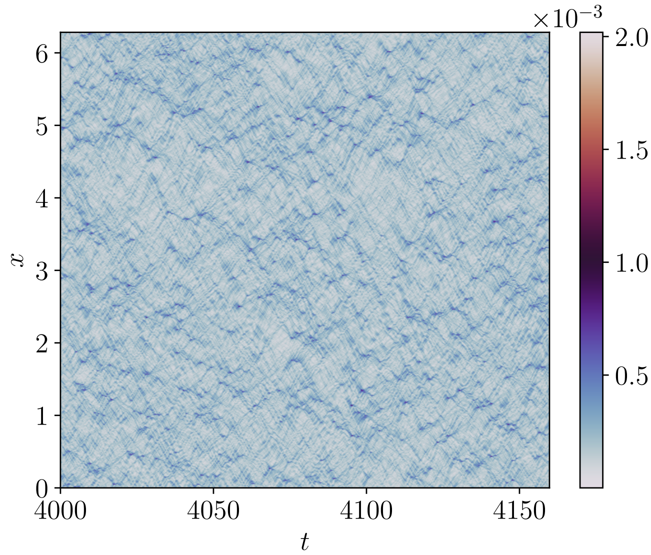


Figure 3: Plot of $|\psi(x,t)|$ for $2k_+/\epsilon^2 < t < 2k_+/\epsilon^2 + 160$.

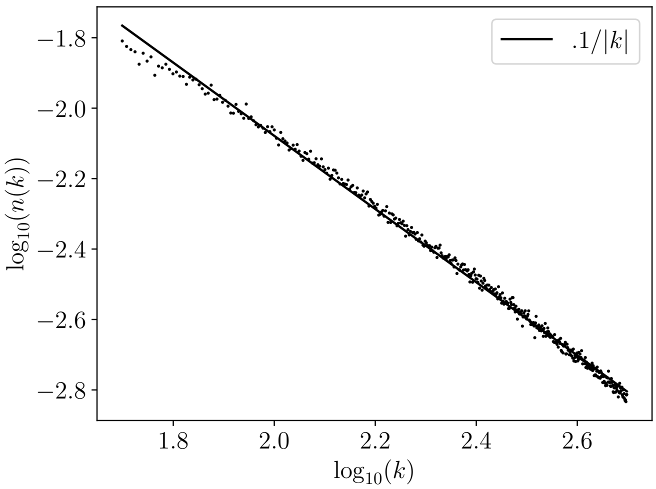


Figure 4: Plot of $n(k)$ for $50 < k < 500$ that we see is approximated by a fit of $n(k) \approx .1/|k|$.

To characterize the transfer of information across scales, similar to the choices made in [8], we separate the inertial range into four overlapping intervals, say $\Delta_j(k)$, with $\Delta_j(k) = [50, 50 + j(500-50)/4]$, $j = 0, 1, 2, 3$. As in [24], we can separate the MMT dynamics into mean and fluctuation components, say

$$\psi = \bar{\psi}_j + \psi'_j$$

where

$$\bar{\psi}_j(x,t) = \sum_{k\in\Delta_j(k)} \left[\hat{\psi}_k(t)\right] e^{ikx},$$

and

$$\left[\hat{\psi}_k(t)\right] = \frac{1}{W}\int_t^{t+W} \hat{\psi}(k,\tau)d\tau.$$

Given that $\delta_s = .2$, we choose the window of time averaging, W, so that we smooth over a time scale of $1/2\epsilon^2$, which is half the length over which nonlinear effects are significant. Thus we are avoiding aliasing in the time series of the mechanism of multi-scale energy transfer.

One can then show, using the quasi-Gaussian closure approximation that $\overline{\left|\psi_j'\right|^2 \psi_j'} \approx 0$, that we can separate the energy across $\Delta_j(k)$ so that

$$\frac{d}{dt}\int_0^{2\pi} |\partial_x|^{1/2} \left|\bar{\psi}_j\right|^2 dx = F_j(t),$$

where

$$F_j(t) = \text{Im}\left\{\int_0^{2\pi} |\partial_x|^{1/2} \left(\bar{\psi}_j^*\right)^2 \overline{\left(\psi_j'\right)^2} dx\right\}.$$

Given the nesting of the intervals $\Delta_j(k)$, i.e. $\Delta_0(k) \subset \Delta_1(k)$ etc..., we see each fluctuation ψ_j' represents higher wave numbers than the corresponding average so that the average energy transfer function $F_j(t)$ tracks the mean coupling between longer and shorter wavelengths, thereby allowing us to characterize energy cascade phenomena.

To compute the lagged transfer of information across the multiscale energy transfer functions $F_j(t)$, starting from the raw data sets $\{F_j(t_k)\}_{k=1}^{N_{tot}}$, where $t_{k+1} - t_k = .2$, we deprecate the data further by a factor of 20 making the defacto sampling rate $\delta_s = 4$. Finally, we apply a low-pass filter to each term $F_j(t)$ so as to isolate the most meaningful portions of the signal which is quantified through autocorrelation. We see the results of this in Figure 5. In particular, we see that deprecation of the time series and the use of the low-pass filter brings out correlations on time scales that are feasible to examine via IDTXL.

Having sufficiently processed the data, we now examine via the IDTXL library how the energy transfer functions $F_j(t)$ do or do not exhibit causative relationships, thereby illustrating how energy moves to maintain the statistically stationary cascade distribution. Using a maximum lag length of 50 deprecated times steps, thus corresponding to a maximum lag time of 200 units of non-dimensional time which is the characteristic time scale for nonlinearity to have a significant effect on wavenumbers at the left of our chosen inertial range, we produce the results of Figure 6. Computational limitations prevent us from exploring larger lag choices. We plot the lag between scales which corresponds to the maximum transfer entropy, and we plot results

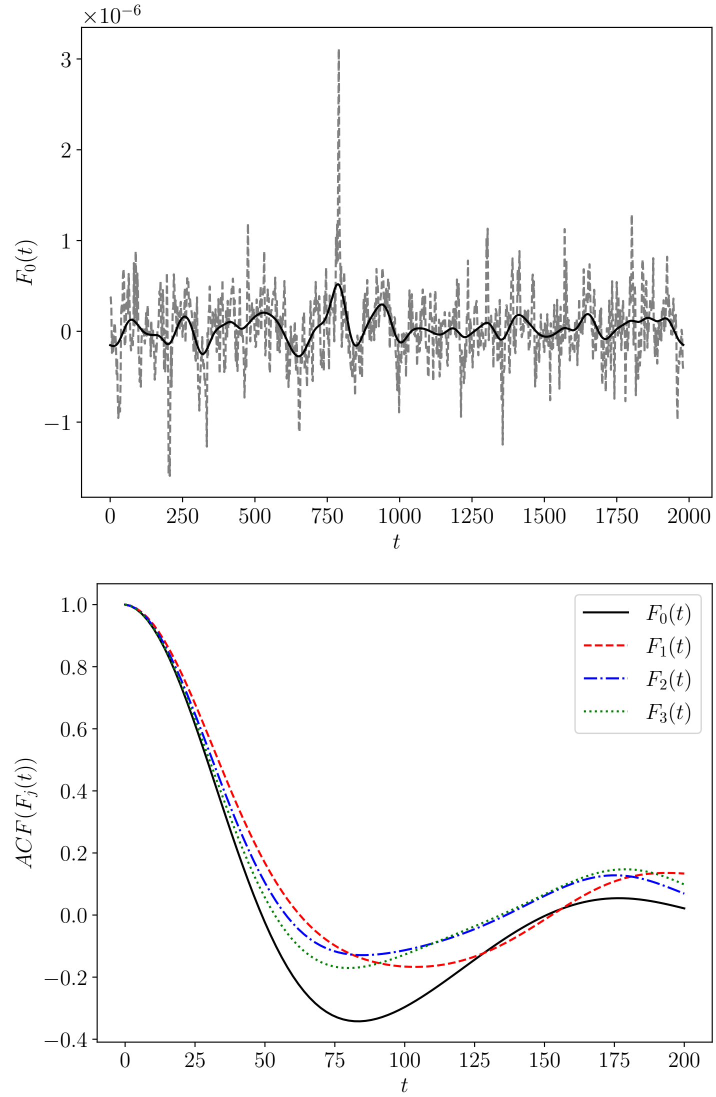
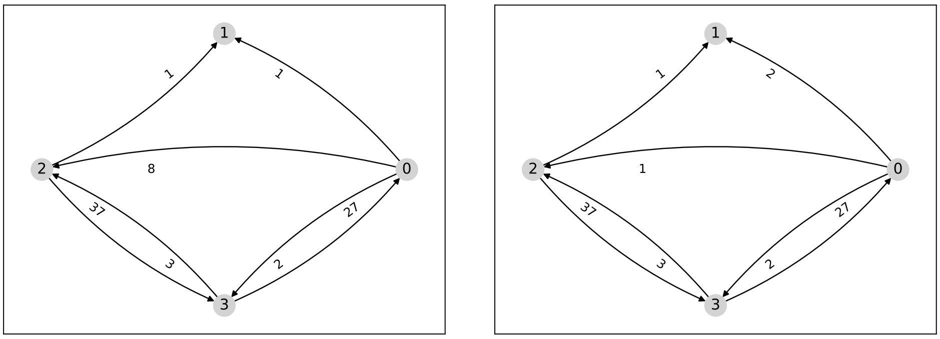


Figure 5: In the top figure, a plot of $F_0(t)$ (dashed) and the result of applying a low pass filter to $F_0(t)$ (solid). In the bottom figure, the autocorrelation of each low pass filtered function $F_j(t)$ is plotted.

with hypothesis testing done at $p = .025$ and $p = .05$. We note that because of the greed optimization strategy of IDTXL, the lags at $p = .025$ are not necessarily subsets of those at $p = .05$. We denote the relative transfer entropy contributed by a target $F_j(t)$ relative to source $F_i(t)$ with lag d as $T^M_{j\to i}(d)$.

As can be seen, both levels of hypothesis testing find a great deal of coupling across scales. In particular, we see a dichotomy in the time scales



$p = .025$ $p = .05$

Figure 6: Lags corresponding to the maximum transfer entropy between low-pass filtered energy functions $F_j(t)$. The maximum allowed lag is 50 deprecated time steps, corresponding to 200 units of non-dimensional time. The left figure is computed with $p = .025$ while the right figure is computed with $p = .05$.

between forward and inverse cascades. For example,for $p = .025$, we find that $T^M_{03}(2) = .00195$ while $T^M_{30}(27) = .00197$, showing a relative equivalence in importance between forward and inverse energy transfer while being distinguished in time scales with the forward cascade progressing much faster than the inverse. Likewise, for $p = .025$, we find that $T^M_{02}(8) = .00133$ while $T^M_{32}(37) = .00177$. The argument we are making is complicated somewhat by the transfer from F_2 to F_1, though we find that $T^M_{21}(1) = .0008$, making the connection markedly more tenuous than the others. A higher hypothesis testing threshold, or more samples in the hypothesis testing would be expected to eliminate the selection of this link. Moreover, as we know from [15], the wave mixing driving energy transfer prevents the formation of as straightforward cascades of information as seen for example in [8]. Further testing would need to be done to determine if this is in fact accurate.

3 Empirical Wavelet Transforms

We now look at developing a method which generates efficient multiscale representations of scalar time series. This is done with an eye towards ultimately detecting energy transfer across said scales using transfer entropy, thereby allowing for the detection of potential cascades from limited measurements.

Our method starts from the empirical wavelet transform of [16]. Given real-valued time signal $x(t)$, we define its Fourier-transform to be $\hat{x}(\omega)$ and corresponding inverse $x^{\vee}(t)$ to be

$$\hat{x}(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt, \; x(t) = \hat{x}^{\vee}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{x}(\omega)e^{i\omega t}d\omega.$$

Throughout, we suppose that $\hat{x}(\omega)$ has support in the interval $[-\omega_M, \omega_M]$, and given that $x(t)$ is assumed real, we then immediately have the symmetry $\hat{x}(-\omega) = \hat{x}^*(\omega)$, where $*$ denotes complex conjugation. Thus we need only study the positive interval $[0, \omega_M]$. In [16], a powerful method for generating a decomposition of $s(t)$ was developed which identifies break points in $[0, \omega_M]$, say ω_j, so that if we identify N_B breaks then we have

$$[0, \omega_M] = \bigcup_{j=0}^{N_B} [\omega_j, \omega_{j+1}], \; \omega_0 = 0, \; \omega_{N+1} = \omega_M,$$

and then, for $j = 1, \cdots, N$ constructs wavelet functions $\psi_j(\omega)$ with support on $[\omega_j - \tau_j, \omega_{j+1} + \tau_{j+1}]$ and approximation function $\phi_0(\omega$ with support on $[0, \omega_1 + \tau_1]$ (see [16] for details on how to choose τ_j) so that

$$\begin{aligned} x(t) = &\left(|\phi_0(\omega)|^2 \hat{x}(\omega) + \sum_{j=1}^{N_B} |\psi_j(\omega)|^2 \hat{x}(\omega) \right)^{\vee}, \\ =& x_0(t) + \sum_{j=1}^{N_B} x_j(t), \end{aligned}$$

with the further restriction that

$$|\phi_0(\omega)|^2 + \sum_{j=1}^{N_B} |\psi_j(\omega)|^2 = 1,$$

so that no energy is lost in the decomposition. This method allows for the identification of multi-scale features without any a-priori assumption of a wavelet basis. Likewise, the method produces far more interpretable results than equivalent approaches such as the Empirical-Mode Decomposition [25]. Finally, the choices for τ_j we use in this work minimize the degree of correlation between scales, making the EWT approach similar to principal-component analysis in so far as the generated modes are only weakly correlated and thus represent nearly orthogonal directions of multi-scale dynamics.

However, in [16], the breaks are discovered through a peak detection algorithm which selects those peaks which persist through a sequential process of convolutional smoothing. While helping to mitigate the effects of noise, this can also unintentionally erase features. Moreover, the user has no control over the number of resulting modes. Thus while a promising approach, we found that the method did not reliably provide us with meaningful modal decompositions that allowed for ready interpretation within our information theoretic computations.

To address these issues, we instead adapt Otsu's partition method [17], so that we specify the number of modes that we want and then use an optimization routine to determine where best to put the break points in the signal spectrum. We call this method the Otsu EWT (OEWT) method. We begin our method by supposing that we are given a scalar data set $\{\hat{x}_j\}_{j=1}^{N_s}$ where $\hat{x}_j \sim \mathbb{P}_u$ is sampled from an absolutely continuous distribution $\mathbb{P}_u$ with affiliated density $p_u(x)dx$. We further imagine that the data is well described as a collection of (N_c+1)-segments with affiliated thresholds $\{\omega_l\}_{l=0}^{N_c+1}$ with $\omega_0 = -\infty$ and $\omega_{N_c+1} = \infty$ such that the l^{th} segment has $N_{c,l}$ members which satisfy one of the inequality:

$$\omega_l < \hat{x}_j < \omega_{l+1}.$$

To determine how best to choose the thresholds $\{\omega_1\}_{l=1}^{N_c}$, following [17], we define the segment probabilities p_l where

$$p_l = \int_{\omega_l}^{\omega_{l+1}} p_u(x)dx, \ l = 0, \cdots, N_c$$

and the conditional averages μ_l such that

$$\mu_l = \frac{1}{p_l}\int_{\omega_l}^{\omega_{l+1}} x p_u(x)dx.$$

We then have the identities/constraints

$$\sum_{l=0}^{N_c} p_l = 1, \ \sum_{l=0}^{N_c} \mu_l p_l = \mu_u, \ \mu_u = \int_{\mathbb{R}} x p_u(x)dx.$$

Likewise, we can define the conditional variances σ_l^2 so that

$$\sigma_l^2 = \frac{1}{p_l}\int_{\omega_l}^{\omega_{l+1}} (x-\mu_l)^2 p_u(x)dx,$$

which has the corresponding constraint that

$$\sum_{l=0}^{N_c} p_l\left(\sigma_l^2 + \mu_l^2\right) = \sigma_u^2 + \mu_u^2, \ \sigma_u^2 = \int_{\mathbb{R}} (x-\mu_u)^2 p_u(x)dx.$$

We then seek to maximize the *between-group* variance σ_B^2 defined as

$$\begin{aligned}\sigma_B^2 =& \sum_{l=0}^{N_c} p_l\left(\mu_l - \mu_u\right)^2 \\ =& \sigma_u^2 - \sigma_W^2,\end{aligned}$$

where the *in-group* variance σ_W^2 is defined to be

$$\sigma_W^2 = \sum_{l=0}^{N_c} p_l \sigma_l^2.$$

Thus we can see the optimization problem as either one in which we want each segment's average maximally separated from the total distribution average, or we want to minimize the conditionally weighted segment variances, thereby generating well defined clusters or segments.

To find the critical points of σ_W^2 with respect to ω_m, we need to solve the equation

$$\partial_{\omega_m}\left(\int_{\omega_{m-1}}^{\omega_m} (x-\mu_{m-1})^2 p_u(x)dx + \int_{\omega_m}^{\omega_{m+1}} (x-\mu_m)^2 p_u(x)dx\right) = 0.$$

We find that, assuming that $p_u(\omega_m) \neq 0$ and that $\mu_m \neq \mu_{m-1}$ that we have critical points when

$$G_m(\omega_{m-1}, \omega_m, \omega_{m+1}) - \omega_m = 0,$$

where

$$G_m(\omega_{m-1}, \omega_m, \omega_{m+1}) = \frac{1}{2}\left(\frac{\int_{\omega_{m-1}}^{\omega_m} x p_u(x)dx}{\int_{\omega_{m-1}}^{\omega_m} p_u(x)dx} + \frac{\int_{\omega_m}^{\omega_{m+1}} x p_u(x)dx}{\int_{\omega_m}^{\omega_{m+1}} p_u(x)dx}\right).$$

From this, we see that Jacobian is necessarily a tri-diagonal matrix with, for $1 \leq m \leq N_c$, the entries

$$\partial_{\omega_{m-1}} G_m = \frac{p_u(\omega_{m-1})}{2p_{m-1}}\left(\mu_{m-1} - \omega_{m-1}\right),$$

$$\partial_{\omega_m} G_m = \frac{p_u(\omega_m)}{2}\left(\frac{\omega_m - \mu_{m-1}}{p_{m-1}} + \frac{\mu_m - \omega_m}{p_m}\right) - 1,$$

and

$$\partial_{\omega_{m+1}} G_m = \frac{p_u(\omega_{m+1})}{2p_m}\left(\omega_{m+1} - \mu_m\right),$$

where we keep in mind that $\omega_0 = -\infty$ and $k_{N_c+1} = \infty$. Numerical quadrature schemes and root-finding routines found in standard numerical libraries can now be used to solve for the relevant fixed points.

3.1 Using OEWT and Information Theory to Detect Multiscale Cascades

Having established a baseline understanding of how energy is moved across scales in the MMT equation, we now look at using OEWT to find multiscale transfer using only a scalar time series measurement. Specifically, we use a subset of samples from $\{|\psi(0,t_k)|\}_{k=1}^{N_{tot}}$ and then perform OEWT on this subset of our original time series. Note, measuring at $x=0$ is arbitrary and has no bearing on the final results. We plot $|\psi(0,t)|$ and its autocorrelation in Figure 7. As can be seen, the time series appears to be all but white noise, akin to the results seen in Figure 5. Any longer-time- correlative structure is relatively buried, so as we will see, the OEWT method at a minimum helps discover meaningful substructure in quickly varying time series.

To generate our results, the original time series is deprecated so that the final sampling rate is $\delta_s = 8$ units of non-dimensional time. As can be seen, it would generally be recognized as a chaotic, perhaps even noisy, signal, and any structure within it is not readily apparent.

Using initial break choices $\omega_b = \{.15, .6, 1.2\}$ so that we get $N_B = 4$, using our OEWT algorithm produces the following decomposition of our time series as seen in Figure 8. Likewise, we see the autocorrelation of the separated components $s_j(t)$ in 9. Similar to what we saw in the prior section, aside from the OEWT helping to identify otherwise difficult to detect substructure in the data, the presence of longer time correlations is indicative of potential scale coupling and information transfer.

To wit then, using the IDTXL library we find the lag corresponding to maximum transfer entropy among the separated scale functions $s_j(t)$. The maximum allowed lag is $d = 100$ corresponding to 800 units of non-dimensional time. Again, we denote the relative transfer entropy contributed by a target $s_j(t)$ relative to source $s_i(t)$ with lag d as $T^M_{j\rightarrow i}(d)$. As can be seen, we successfully find information flow across the scales represented by $s_j(t)$, though the picture of transfer is markedly simpler than what we found for the multiscale energy transfer functions $F_j(t)$. Looking at the $p = .025$ case, the fast transfer from $s_1(t)$ to $s_0(t)$ is again somewhat surprising compared to the longer time transfer from $s_0(t)$ to $s_2(t)$. However, comparing actual transfer entropy values, we find the TE from $s_1(t)$ to $s_0(t)$ is $T^M_{1\rightarrow 0}(9) = .009$ while $T^M_{0\rightarrow 2}(51) = .08$, representing an order of magnitude difference. Therefore, we can say the dominant mechanism of information flow is from the slowest timescale to the second fastest, with a relatively weak inverse transfer from $s_1(t)$ to $s_0(t)$. Further, we also see, looking past the maximum entropy contribution, that there is a transfer $T^M_{1\rightarrow 0}(83) = .003$, which, while a third of $T^M_{1\rightarrow 0}(9)$, is still on the same order of magnitude. Thus longer term contributions are also present and roughly of similar significance. Given the disappearance of linkage to $s_3(t)$ upon lowering p, we might argue that the fastest scale represents essentially noise in the system that is otherwise decoupled from the more meaningful dynamics encoded in $|\psi(0,t)|$.

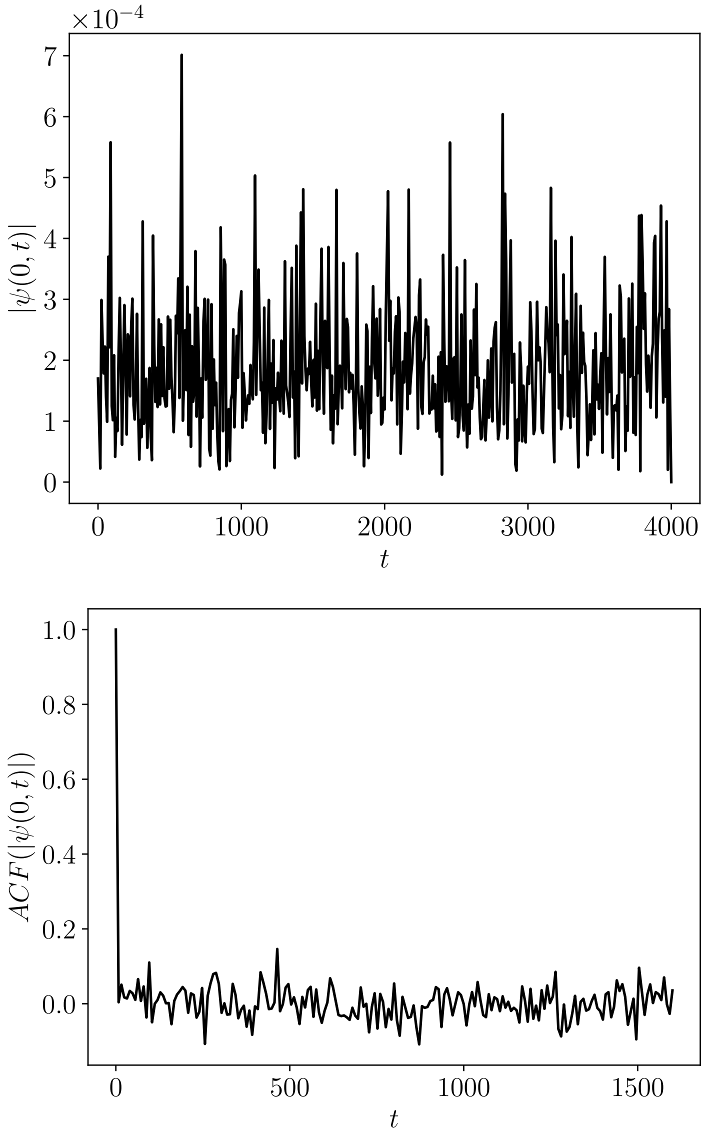


Figure 7: Plot of $|\psi(0,t)|$ and its autocorrelation.

4 Discussion and Future Work

With our perspective on how spatial structures interact across disparate time scales, we are equipped with a new tool for data-driven analysis. We anticipate that this technique will enable us to uncover relationships between variables at different time scales, helping us analyze how systems evolve as parameters change. In particular, we expect that certain systems, including possibly the MMT, may undergo critical events—such as the onset of ex-

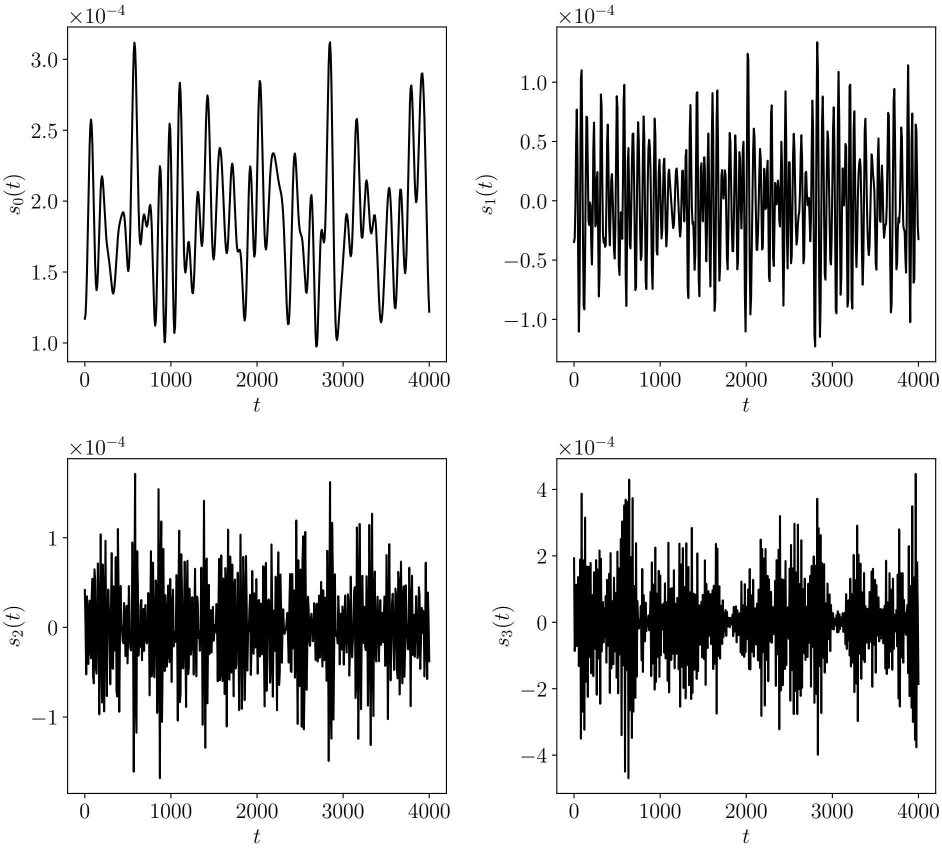


Figure 8: OEWT Decomposition of $|\psi(0,t)|$.

treme behavior—that exhibit observable precursors. These precursors may manifest as shifts in interaction time scales, either becoming critical or obstructed. We hope that this approach will provide a bifurcation analysis of criticality across time scales, offering insights that can be connected to more traditional methods.

Acknowledgments

We are grateful to the support from Office of Naval Research grant N00014-23-1-2106. Also, E.M.B. is supported by the NIH-CRCNS, DARPA RSDN, the ARO, the AFSOR and the ONR. We are likewise grateful to Daniel Jay Alford-Lago and Stefan Cline for many fruitful discussions and suggestions that have helped improve this work.

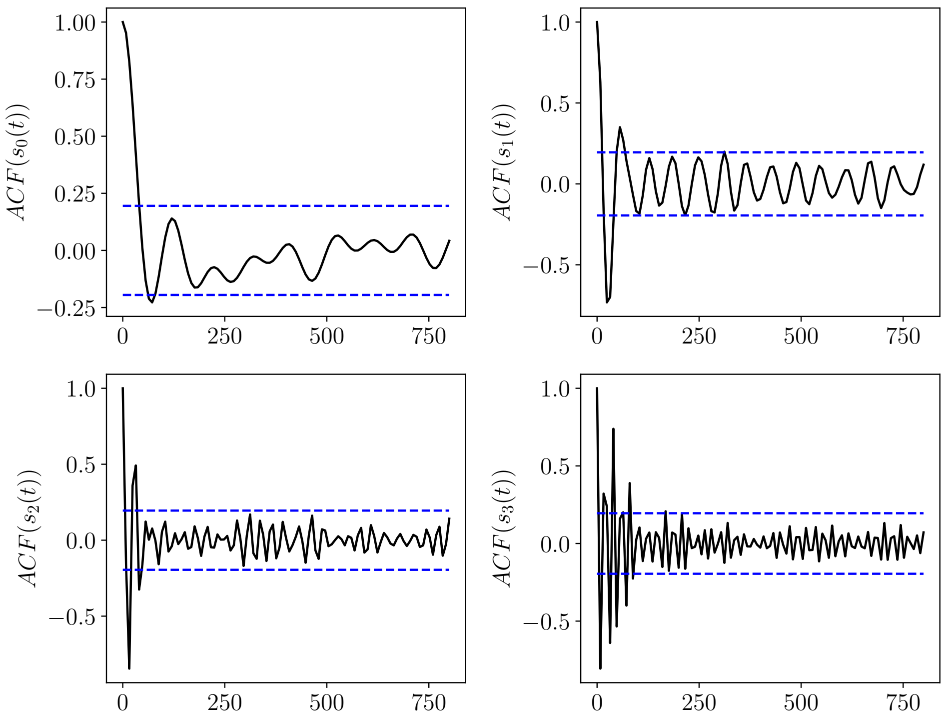
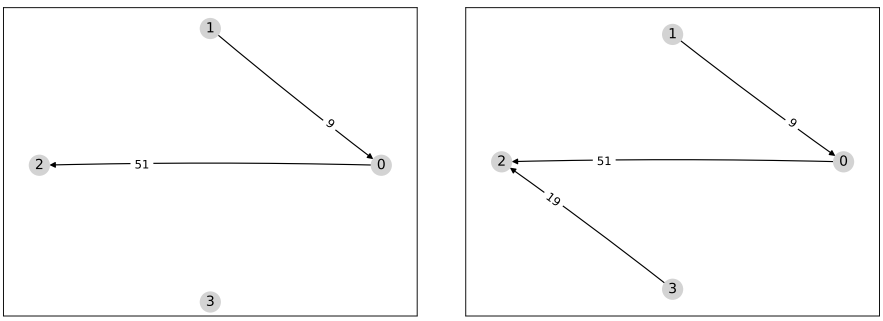


Figure 9: Autocorrelations of the functions $s_j(t)$ generated by the OEWT decomposition. Horizontal bars represent confidence intervals.



$p = .025$ $p = .05$

Figure 10: Lags corresponding to the maximum transfer entropy between the OEWT component functions $s_j(t)$. The maximum allowed lag is $d = 100$. The left figure is computed with $p = .025$ while the right figure is computed with $p = .05$.

References

[1] Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000.

[2] Jie Sun, Dane Taylor, and Erik M. Bollt. Causal network inference by optimal causation entropy. *SIAM Journal on Applied Dynamical Systems*, 14(1):73–106, 2015.

[3] A. A. R. AlMomani, Jie Sun, and Erik M. Bollt. How entropic regression beats the outliers problem in nonlinear system identification. *Chaos*, 30:013107, 2020.

[4] Jie Sun and Erik M Bollt. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. *Physica D: Nonlinear Phenomena*, 267:49–57, 2014.

[5] Jie Sun, Carlo Cafaro, and Erik M Bollt. Identifying the coupling structure in complex systems through the optimal causation entropy principle. *Entropy*, 16(6):3416–3433, 2014.

[6] Patricia Wollstadt, Joseph T. Lizier, Raul Vicente, Conor Finn, Mario Martinez-Zarzuela, Pedro Mediano, Leonardo Novelli, and Michael Wibral. Idtxl: The information dynamics toolkit xl: a python package for the efficient analysis of multivariate information dynamics in networks. *Journal of Open Source Software*, 4(34):1081, 2019.

[7] Leonardo Novelli, Patricia Wollstadt, Pedro Mediano, Michael Wibral, and Joseph T. Lizier. Large-scale directed network inference with multivariate transfer entropy and hierarchical statistical testing. *Network Neuroscience*, 3(3):827–847, 2019.

[8] Adrián Lozáno-Duran and Gonzalo Arranz. Information-theoretic formulation of dynamical systems: Causality, modeling, and control. *Phys. Rev. Research*, 4:023195, 2022.

[9] Nan Chen, Xiao Hou, Qin Li, and Yingda Li. Understanding and predicting nonlinear turbulent dynamical systems with information theory. *Atmosphere*, 10(5), 2019.

[10] Giuseppe Consolini and Massimo Materassi. Chapter 16 - advanced statistical tools in the near-earth space science. In Massimo Materassi, Biagio Forte, Anthea J. Coster, and Susan Skone, editors, *The Dynamical Ionosphere*, pages 243–256. Elsevier, 2020.

[11] Massimo Materassi, Tommaso Alberti, Yenca Migoya-Orué, Sandro Maria Radicella, and Giuseppe Consolini. Chaos and predictability in ionospheric time series. *Entropy*, 25(2), 2023.

[12] A.J. Majda, D.W. McLaughlin, and E.G. Tabak. A one-dimensional model for dispersive wave turbulence. *J. Nonlin. Sci.*, 6:9–44, 1997.

[13] V.E. Zakharov, F. Dias, and A.N. Pushkarev. One-dimensional wave turbulence. *Phys. Rep.*, 398:1–65, 2004.

[14] A. Hrabski and Y. Pan. On the properties of energy flux in wave turbulence. *J. Fluid Mech.*, 936:A47, 2022.

[15] G. Dematteis and Y. V. Lvov. The structure of energy fluxes in wave turbulence. *J. Fluid Mech.*, 954:A30, 2023.

[16] Jérôme Gilles. Empirical wavelet transform. *IEEE Transactions on Signal Processing*, 61(16):3999–4010, 2013.

[17] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[18] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[19] Luca Faes, Giandomenico Nollo, and Alberto Porta. Information-based detection of nonlinear granger causality in multivariate processes via a nonuniform embedding technique. *Phys. Rev. E*, 83:051112, May 2011.

[20] F. Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, pages 366–381. Springer Berlin Heidelberg, Berlin, Heidelberg, 1981.

[21] Ioannis Vlachos and Dimitris Kugiumtzis. Nonuniform state-space reconstruction and coupling detection. *Phys. Rev. E*, 82:016207, 2010.

[22] Christopher W. Curtis, Erik Bollt, and Daniel Jay Alford-Lago. Entropic regression dynamic mode decomposition (erdmd) discovers informative sparse and nonuniformly time delayed models. *International Journal of Bifurcation and Chaos*, 0(0):2450167, 2024.

[23] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys. Rev. E*, 69:066138, Jun 2004.

[24] I.G. Grooms and A.J. Majda. Stochastic superparametrization in a one-dimensional model for wave-turbulence. *Commun. Math. Sci.*, 12:509–525, 2014.

[25] Huang Norden E., Shen Zheng, Long Steven R., Wu Manli C., Shih Hsing H., Zheng Quanan, Yen Nai-Chyuan, Tung Chi Chao, and Liu Henry. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roc. Soc. A*, 45:4903–4995, 1998.