Long-range Forecasting as a Past Value Problem: Using Scaling to Untangle Correlations and Causality

L. Del Rio Amador[1], S. Lovejoy[1]

[1]Physics, McGill University, 3600 University St., Montreal, Que. H3A 2T8, Canada.

Corresponding author: Lenin Del Rio Amador (delrio@physics.mcgill.ca)

Key Points:

- Scaling-based long-range stochastic forecasting is a past value problem not an initial value problem.
- Granger causality shows that while spatial correlations exist in the temperature field, they cannot be used to improve on predictions based on historical data of individual infinite time series.
- The statistics and teleconnection patterns of the real-world can be reproduced with stochastic simulations with a total lack of causal relationships.

Abstract

Conventional long-range weather prediction is an initial value problem that uses the current state of the atmosphere to produce ensemble forecasts. Purely stochastic predictions for long-memory processes are "past value" problems that use historical data to provide conditional forecasts. Teleconnection patterns, defined from cross-correlations, are important for identifying possible dynamical interactions, but they do not necessarily imply causation. Using the precise notion of Granger causality, we show that for long-range stochastic temperature forecasts, the cross-correlations are only relevant at the level of the innovations – not temperatures. This justifies the Stochastic Seasonal to Interannual Prediction System (StocSIPS) that is based on a (long memory) fractional Gaussian noise model. Extended here to the multivariate case, (m-StocSIPS) produces realistic space-time temperature simulations. Although it has no Granger causality, emergent properties include realistic teleconnection networks and El Niño events and indices.

1 Introduction

For forecasts over the weather regime – below the ≈10 day deterministic predictability limit – Numerical Weather Prediction (NWP) and General Circulation Models, (GCMs) have been highly successful, yet for longer term macroweather ("long range") forecasts, their skill is disappointing. This has motivated the development of stochastic alternatives. Successful stochastic forecasts require causal models and the search for causality typically starts with correlations. In the last years, two stochastic strands have emerged each inspired by different sources of strong correlations. A particularly well studied constellation of correlations are associated with large scale spatial structures – teleconnections – as vividly displayed in climate networks [e.g.: (Donges et al., 2009b; Ludescher et al., 2014)]. Teleconnection-inspired forecast models often use climate (especially El Niño) indices [see (Brown & Caldeira, 2020; Eden et al., 2015)]. An alternative source of correlations upon which to base causal models is the system's long range memory (Blender & Fraedrich, 2003; Bunde et al., 2005; Rypdal et al., 2013; Varotsos et al., 2013), a consequence of temporal scaling, itself associated with long range spatial scaling, a basic property of the governing equations that is well respected by both GCMs and the empirical data [(Palmer, 2019), see also the review (Lovejoy & Schertzer, 2013)].

At the moment, these strands are at virtual antipodes. Models based on teleconnections use only data from a few months – they are Markovian, short (exponential) memory models that get their skill largely from spatial information. In this, they are almost as extreme as GCMs that are zero-memory, initial value models based purely on the spatial information at $t = 0$. In contrast, the scaling, long memory Stochastic Seasonal to Interannual Prediction System (StocSIPS) model is at the opposite extreme (Del Rio Amador & Lovejoy, 2019, 2020). For each pixel, it uses historical past data to forecast the future - but uses no other data as co-predictors: it is a purely "past value" model. In spite of this apparent deficiency, for monthly, seasonal, and annual temperature forecasts StocSIPS' skill already rivals – or exceeds – those of GCMs.

This paper attempts to answer the obvious question: is it possible to make a model that combines strong spatial correlations and long memory to produce even more skillful forecasts? While it is well known that correlations and causality are not synonymous, the precise relationship between the two is often unclear and there are no general tools for untangling them. However, the present case is an important exception: the problem of improving StocSIPS using spatial co-predictors can be precisely answered by using the theoretical framework of Granger causality (Granger, 1969).

Two series are Granger causally related iff one can be used as a skillful co-predictor of the other. Therefore, it suffices to enquire as to the Granger causality of the space-time StocSIPS model. If the temperature teleconnections have no Granger causality, then they will not improve StocSIPS forecasts. In the first part of the paper we propose a multivariate surface temperature model (m-StocSIPS) for which the uncoupled regional StocSIPS model gives the optimal forecast. m-StocSIPS also reproduces the empirical cross-correlation structure over a wide range of time lags. This is made more convincing by making simulations that display numerous realistic but emergent model properties including spatial teleconnection networks, realistic El Niño patterns and indices. The optimal m-StocSIPS predictor at a given location is obtained from its own past if the series is long enough. Even strongly spatially correlated series from other locations do not help improve the skill, teleconnection correlations may be seductive, but without Granger causality, they are misleading.

2 Methods

2.1 Stochastic modeling of the temperature anomalies

In macroweather temperature anomalies at position **x** (after removing the annual cycle) can be modeled as a trend-stationary process:

$$T_{\text{anom}}\left(\mathbf{x},t\right)=T_{\text{anth}}\left(\mathbf{x},t\right)+T\left(\mathbf{x},t\right), \tag{1}$$

where $T(\mathbf{x},t)$ is a stochastic stationary component and $T_{\text{anth}}(\mathbf{x},t)$ is a deterministic low-frequency response to anthropogenic forcings as in (Del Rio Amador & Lovejoy, 2019).

The stationary stochastic $T(\mathbf{x},t)$, is the zero-mean residual natural variability that includes "internal" variability and the response of the system to other natural forcings (e.g.: volcanic and solar). These anomalies can be predicted by modelling each position independently using an univariate representation [the regional StocSIPS model presented in (Del Rio Amador

& Lovejoy, 2020), hereafter DRAL]. However, to investigate whether forecasts for individual series can be improved using other data, a multivariate framework is needed. A quasi-Gaussian process, stationary in time, but inhomogeneous in space has a multivariate continuous-in-time Wold representation (moving average of infinite order MA(∞)) (Box et al., 2008; Brockwell & Davis, 1991; Wold, 1938):

$$T_i(t) = \sum_j \int_{-\infty}^{t} \kappa_{ij}(t-t')\gamma_j(t')dt' . \tag{2}$$

The index "i" is a subscript indicating the spatially discrete position ("pixel"), the matrix $\kappa_{ij}(t)$ is a kernel specifying the MA process and the innovations, $\gamma_i(t)$, are normalized Gaussian white noise processes with $\langle\gamma_i(t)\rangle = 0$, $\langle{\gamma_i}^2(t)\rangle = 1$ and cross-correlation matrix:

$$\rho_{ij}(t-t') = \left\langle \gamma_i(t)\gamma_j(t')\right\rangle = a_{ij}\delta(t-t'), \tag{3}$$

where $\delta(t)$ is the Dirac function, $\langle\cdot\rangle$ denotes ensemble averaging and $-1 < a_{ij} < 1$. This "delta-correlated" innovation temporal structure implies that the latter are totally unpredictable and is the key property below.

The cross-covariance for time lag $\Delta t > 0$ for the temperature is:

$$C_{ij}(\Delta t) = \left\langle T_i(t)T_j(t+\Delta t)\right\rangle = \sum_m \sum_n \int_0^{\infty} \kappa_{im}(t')\kappa_{jn}(t'+\Delta t)a_{mn}dt' , \tag{4}$$

hence the cross-correlation is:

$$R_{ij}(\Delta t) = \frac{C_{ij}(\Delta t)}{\sqrt{C_{ii}(0)C_{jj}(0)}} . \tag{5}$$

Since the process is Gaussian with zero mean, it is completely determined by the correlation structure. In the macroweather regime – with the possible exception of extremes – $T_i(t)$ is nearly Gaussian in time, but multifractal in space and the statistics of its fluctuations are scale-invariant over wide ranges (Lovejoy, 2018; Lovejoy et al., 2018; Lovejoy & Schertzer, 2013). The scaling behaviour in time implies that there are power-law correlations and hence potentially a large memory that can be exploited. The simplest relevant scaling process is the statistically stationary fractional Gaussian noise (fGn) process.

The fGn based StocSIPS model was first developed for monthly and seasonal forecast of globally averaged temperature (Lovejoy et al., 2015; Del Rio Amador & Lovejoy, 2019). Recently, DRAL extended StocSIPS to the regional prediction of $T_i(t)$, where each grid point was considered as an independent time series. This univariate representation using a resolution τ fGn process (see the supporting information) can be extended to the multivariate case with the kernel:

$$\kappa_{ij}(t) = \delta_{ij}\frac{1}{\tau}\frac{c_{H_i}\sigma_{T_i}}{\Gamma\left[H_i+3/2\right]}\left[t^{H_i+1/2} - (t-\tau)^{H_i+1/2}\theta(t-\tau)\right], \tag{6}$$

where $\theta(t)$ is the Heaviside (step) function, Γ is the Gamma function, $H_i \in (-1,0)$ is the fluctuation exponent characterizing the scaling of the fluctuations in time, σ_{T_i} is the standard deviation, c_{H_i} is a normalization constant and δ_{ij} is the Kronecker δ. The different temperature series, $T_i(t)$, are correlated, and the spatial correlation structure is inherited from the innovation cross-correlations, a_{ij}. The presence of the Kronecker δ in Eq. (6) implies that the temperature at grid point "i" is an fGn with parameters H_i and σ_{T_i}.

In DRAL it was shown that the fGn model (Eq. (6)) is an accurate univariate representation of the natural temperature variability for most of the globe. However, in the tropical ocean, the fGn model approximates the temperature increments, meaning that the actual temperature variability is modelled as a fractional Brownian motion (fBm) process with fluctuation exponent $H_i \in (0,1)$ (see Fig. 1(a)), although cut-off at multi-annual scales. The fluctuation exponents of fBm and fGn are related as $H_{\mathrm{fBm}} = H_{\mathrm{fGn}} + 1$. Both cases are high-frequency approximations of the more general fractional relaxation noise (fRn) process, introduced in (Lovejoy, 2019; Lovejoy et al., 2020).

The use of a parametric model considerably reduces the number of parameters and clarifies their interpretation. m-StocSIPS is fully determined by the symmetric innovation cross-covariance matrix a_{ij}, the amplitudes of the temperature fluctuations σ_{T_i}, and the memory exponents H_i. These characterize the internal dynamics; for example low values of σ_{Ti} over the oceans are a consequence of the greater heat capacity and thermal inertia and H_i characterizes the memory associated with the multiscale energy storage mechanisms (Lovejoy, 2020; Lovejoy et al., 2020).

m-StocSIPS is defined by $N(N+3)/2$ parameters; in comparison, a vector autoregressive order p model (VAR(p)) needs pN^2 values (Box et al., 2008; Brockwell & Davis, 1991) and for long-memory processes, p is large. These "black box" type models suffer from opaque physical interpretations, and their large number of VAR parameters makes them unstable and subject to overfitting. The same is true for general vector autoregressive-moving average VARMA(p, q) models.

Ultimately, the adequacy of a model must be checked. In this case, the diagnostics are primarily based on the examination of the whiteness and time-independence of the residual vectors $\gamma_i(t)$, which are obtained by inverting Eq. (2) with the estimated parameters. The whiteness was verified in DRAL using the theory in Appendix 1 of (Del Rio Amador & Lovejoy, 2019). To verify the time-independence of the innovations (Eq. (3)), there exist many "goodness-of-fit" tests based on the residual cross-covariance matrices at several lags (Ali, 1989; Hosking, 1980; Li & McLeod, 1981; Poskitt & Tremayne, 1982). In our case, they are either impractical – the matrices have more than $1.1 \cdot 10^8$ elements – or impossible since there is only one realization of our planet. Nevertheless, a visual inspection of the residual cross-correlation matrices for different lags (shown in Fig. S2 in the supporting information) may be enough. Our results indicate that m-StocSIPS is a good approximation, confirmed in Sect. 3.3 using global simulations that convincingly reproduce the space-time patterns (Fig. 2). Aside from minor numerical approximations, StocSIPS predictions presented in DRAL are optimal m-StocSIPS predictions in the minimum mean square error framework, explaining the high StocSIPS forecast skill.

2.3 Correlation, causality and Granger causality

m-StocSIPS uses an fGn model for most of the globe (where $H_i < 0$) and a (truncated) fBm model for the tropical ocean (where $H_i > 0$). The cross-correlation structure for the temperature anomalies is thus determined by three kinds of interaction: 1) fGn-fGn, 2) fGn-fBm and 3) fBm-fBm. The fGn-fGn cross-correlation can be obtained directly by using Eq. (6) in Eq. (4). The exact result is given in the supporting information (Eq. S22). Similar expressions can be obtained for the other two cases (Coeurjolly et al., 2010).

While fGn is a stationary process and fGn-fGn cross-correlations only depend on the lag Δt, this is not the case for fBm. Nevertheless, under some approximations for long enough finite time series, it is possible to obtain expressions that only depend on Δt [see (Delignières, 2015)]. The cross-correlations for $\Delta t \gg \tau$ (τ is the temporal resolution of the time series, i.e. 1 month) are:

<u>Case 1</u>: fGn-fGn ($H_i < 0$ and $H_j < 0$),

$$R_{ij}(\Delta t) \sim \varphi_{H_i,H_j} a_{ij} (\Delta t/\tau)^{H_i+H_j} . \tag{7}$$

<u>Cases 2 and 3</u>: fGn-fBm and fBm-fBm ($H_i > 0$ or/and $H_j > 0$),

$$R_{ij}(\Delta t) \sim \phi_{H_i,H_j} a_{ij} \left[1-\left(\Delta t/\tau_r^{ij}\right)^{H_i+H_j}\right] , \tag{8}$$

for $\Delta t \ll \tau_r^{ij}$, where τ_r^{ij} is a characteristic relaxation time related to the ocean weather-ocean macroweather transition (Lovejoy, 2019; Lovejoy et al., 2018; Del Rio Amador & Lovejoy, 2020), and φ_{H_i,H_j} and ϕ_{H_i,H_j} are proportionality constants that depend on H_i and H_j (see Eq. S25 in the supporting information). As expected, these expressions coincide with the high-frequency approximations of the stationary fRn cross-correlations for H_i and H_j.

Equations (7) and (8) imply that the cross-correlation structure of the temperature field has a spatial correlation component given by the matrix a_{ij}, and a temporal component determined by the memory dependence of the individual series (H_i and H_j). In this sense, they are similar, but more general than the average Statistical Space-Time Factorization (SSTF) proposed earlier by (Lovejoy & de Lima, 2015). For a given location i and lag Δt, the cross-correlation with any other location j will be higher for series whose past is important (large H_i) as compared to series with short memories (small H_i).

Now consider the prediction problem for the general process given by Eq. (2). Since the process is Gaussian, we use the minimum mean square error framework. Although correlations play an important role in the statistical description and in pattern identification, it is wrong to infer causality based on the lagged cross-correlation structure alone. In the words of (Buchanan, 2012): "Not only does correlation not imply causality, but lack of correlation needn't imply a lack of causality either". A classic example is two correlated systems without any dynamic interaction between them but with a common dependence on a third variable. Conversely, there are coupled chaotic systems, that exhibit a complete lack of long-term statistical correlation, despite sharing a clear cause-effect link (Sugihara et al., 2012).

An example from (Barnston, 2014; Lyon & Barnston, 2005) may clarify the discussion. They argue that El Niño events *lead* to a cascade of global impacts, e.g.: wet Central Asias. However, in GCM terms, a given set of initial conditions is the ultimate cause of both an El Niño and a wet season in Central Asia. The chain of events starting from those initial conditions explains the mutual correlations without mutual causation. In traditional mechanistic terms, the best that can be done to reconcile the two viewpoints is the notion of causal chain (e.g. Bunge, 2017). In this fairly qualitative view, the ultimate cause – the initial conditions – triggers a causal chain of events in which El Niño is a "proximate" link leading to a wet season in Central Asia.

From a stochastic point of view, (Andree, 2019) argues that a time series (e.g. the temperature at a given location) has a memory part depending on its own past and a causal part from the past at other locations. For short-memory processes, this causal contribution may be important, explaining how some empirical models obtain their skill by effectively borrowing memory from co-predictors. However, the longer the memory – the more autoregressive steps that are needed – the lower the influence of the causal component. In the limit, all the causal chain for a given time series may be embedded in its own past, so that GCMs and StocSIPS exploit a whole chain of causation, not only the last links in the chain so that their skill is higher than models that only exploit proximate causes.

The precise tool needed to clarify stochastic causality issues is Granger causality (Granger, 1969). We say that the temperature T_j at location j fails to Granger-cause the temperature T_i, if for all future times $t > 0$, the mean square error (MSE) of a forecast of $T_i(t)$ based on its own past ($T_i(s)$ for $s \leq 0$) is the same as the MSE of a forecast of $T_i(t)$ based on both $T_i(s)$ and $T_j(s)$. The notion of Granger causality is intuitive and provides a much more rigorous criterion for causation than simple lagged cross-correlations. While other notions of causality exist, Granger causality does imply forecasting ability, which is our only concern here.

We now investigate the Granger causality of m-StocSIPS. A necessary and sufficient condition for the optimality of an estimator is given by the orthogonality principle (Box et al., 2008; Brockwell & Davis, 1991; Palma, 2007; Straškraba, 2007; Wold, 1938), that states that the error of the optimal predictor (in a mean square error sense) is orthogonal to any possible estimator:

$$\left\langle \hat{T}_i(t) E_i(t) \right\rangle = 0, \tag{9}$$

where $\hat{T}_i(t)$ is the temperature predictor for position i at a future time $t > 0$ and $E_i(t) = T_i(t) - \hat{T}_i(t)$ is the error.

From the integral representation (Eq. (2)) and given a diagonal kernel $\kappa_{ij}(t)$ as in Eq. (6), we find that the optimal predictor satisfying this principle is:

$$\hat{T}_i(t) = \int_{-\infty}^{0} \kappa_{ii}(t-t') \gamma_i(t') dt', \tag{10}$$

with error:

$$E_i(t) = \int_{0}^{t} \kappa_{ii}(t-t'') \gamma_i(t'') dt''. \tag{11}$$

E_i only depends on future innovations $\gamma_i(t'')$ ($t'' > 0$), while the estimator, $\hat{T}_i(t)$, depends only on past innovations $\gamma_i(t')$ ($t' < 0$). Since the white noise innovations are δ-correlated in time (Eq. (3)), for any i, j we have:

$$\left\langle T_j(s) E_i(t) \right\rangle = 0; \qquad s < 0,\ t > 0. \tag{12}$$

This means that any predictor that is a linear combination of past temperature values from any position j, is orthogonal to the error of the predictor obtained from the past at location i, given by Eq. (10). Hence, the predictor (Eq. (10)) is optimal given the full field $T(\mathbf{x},t)$ for $t \leq 0$. This is a precise statement of Granger causality. Although there are large cross-correlations inherited from the innovation matrix a_{ij} (Eqs. (7) and (8)), the information of past temperatures from other locations does not help improve the forecast. For StocSIPS predictions, it is the lack of innovation connectivity at non-zero lags that implies that the optimal predictor for any given location is obtained from its past. In effect, these occasionally strong spatial correlations "were already used" for building the past of any given time series, whose past is therefore enough to yield the optimal predictor for that specific series.

3 Results

3.1 Empirical cross-correlations

Our analysis were based on monthly, 2.5° resolution surface temperatures (T2m: 73 × 144 = 10512 points) from 1948 to 2019 (864 months in total) from the National Centers for Environmental Prediction/National Center for Atmospheric Research Reanalysis 1 (Kalnay et al., 1996; NCEP/NCAR, 2020).

The validity of the univariate fGn (StocSIPS) model was confirmed in DRAL by testing the whiteness of the innovations $\gamma_i(t)$ for every grid point i, which were obtained by inverting the discrete version of Eq. (2) (see the supporting material for the theoretical details). We used the fact that a white noise process is a particular case of fGn with fluctuation exponent $H_\gamma = -1/2$. Maximum likelihood estimates for the residuals at 10512 grid points give $H_\gamma = -0.498 \pm 0.003$ and standard deviations $\sigma_\gamma = 1.000 \pm 0.002$, which confirms that the innovations are unit variance δ-correlated white noise and hence the adequacy of the fGn model for the natural temperature variability in the univariate case.

To show that the multivariate model is also realistic, we must check that the lagged cross-correlations between the innovations at different locations (Eq. (3)) are negligible. For this analysis, we obtained the lagged cross-correlation matrices involving the 10512 grid points for the innovations, $\rho_{ij}(\Delta t)$, and for the temperature variability, $R_{ij}(\Delta t)$, for Δt from 1 to 12 months. These matrices are shown in the supporting information (Fig. S2) for $\Delta t = 0$, 1, 2 and 3 months. For the temperature, the correlations decrease with Δt, but large values are often obtained for relatively large lags, following Eqs. (7) and (8). For the innovation cross-correlations, the values decrease much faster. For $\Delta t = 0$, the elements $\rho_{ij}(0) = a_{ij}$ are relatively large, but even for $\Delta t = 1$ month, almost all the correlation is lost. This indicates that the innovations closely satisfy the discrete version of the time-independence condition Eq. (3).

Another way of testing the model is by checking that Eqs. (7) and (8) are good approximations of the empirical $R_{ij}(\Delta t)$. Fig. 1(a) shows the results for ensembles with similar a_{ij}, H_i and H_j values ($a_{ij} = 0.5 \pm 0.025$ gives 10490 pairs). Comparisons are shown for the three cases (fGn-fGn, fGn-fBm and fBm-fBm):

Case 1: fGn-fGn (marked as "+" in Fig.1(a)), we chose the series with $H_i = -0.1 \pm 0.025$ (red symbol) and $H_j = -0.3 \pm 0.025$ (yellow), 380 pairs.

Case 2: fGn-fBm (marked as “o”), the series with $H_i = -0.1 \pm 0.025$ (purple) and $H_j = 0.25 \pm 0.025$ (green), 569 pairs.

Case 3: fBm-fBm (marked as “x”), the series with $H_i = 0.3 \pm 0.025$ (black) and $H_j = 0.4 \pm 0.025$ (cyan), 323 pairs.

Fig. 1(b) shows the average cross-correlations functions of the lag $\lambda = \Delta t/\tau$ ($\tau = 1$ month), with fits from Eqs. (7) and (8). For case 1, we included the dashed red curve corresponding to higher order corrections for fRn processes (Lovejoy, 2019; Lovejoy et al., 2020). The small values of the cross-correlation innovation pairs (“*” in the figure) confirm the independence of these series. Although the expressions (Eqs. (7) and (8)) are only first order

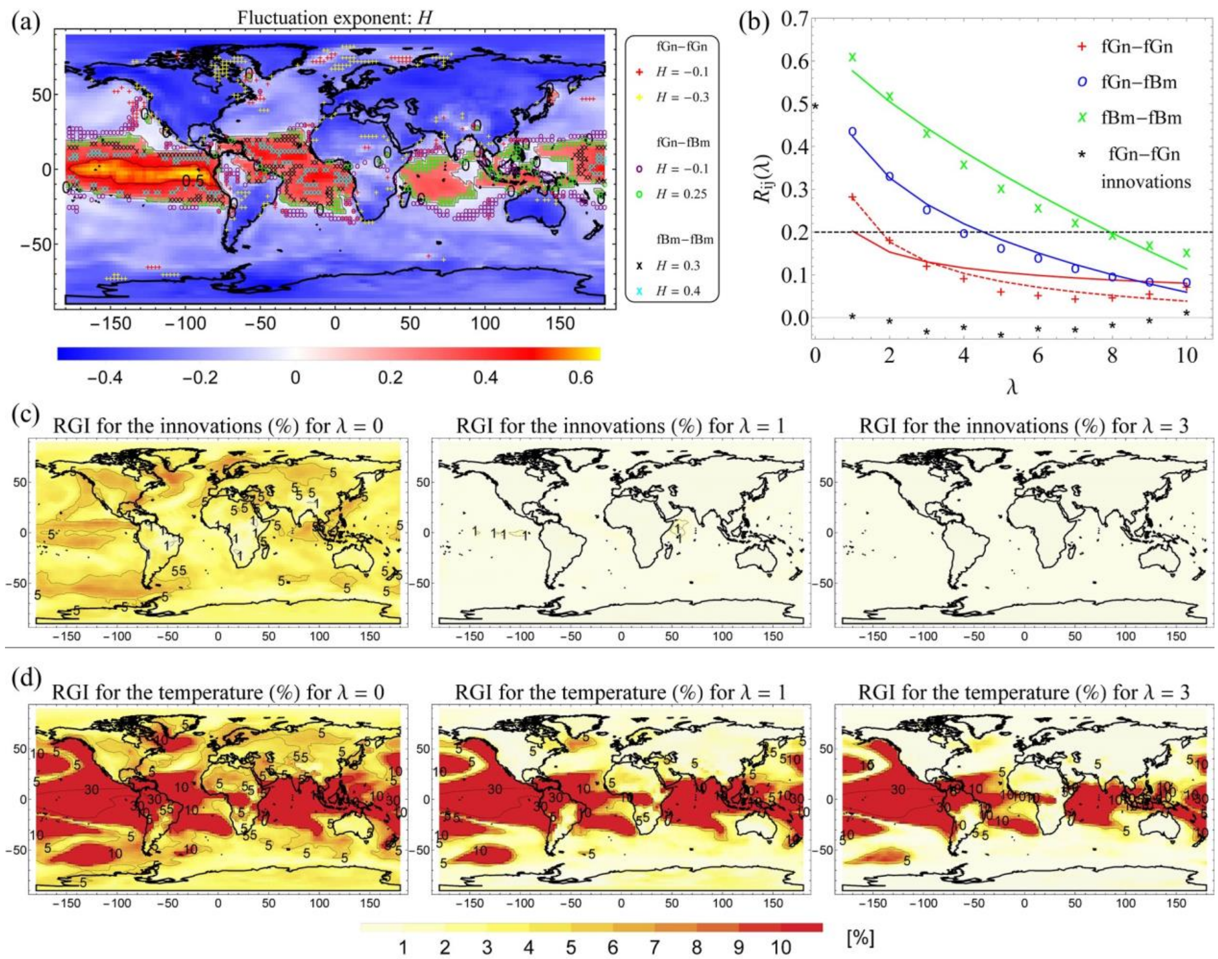

Fig. 1 (a) Maximum likelihood estimates of the fluctuation exponent, H. The grid points forming the pairs used to calculate the average ensemble cross-correlations (shown in (b)) are marked as: “+” for fGn-fGn, “o” for fGn-fBm and “x” for fBm-fBm. The colours indicate the values of H. (b) Average cross-correlations for $\lambda = 1 - 10$ for the cases 1, 2 and 3 (described in the text), with the corresponding fits from Eqs. (7-8) (we also included in dashed red the curve corresponding to higher order corrections for fRn processes). The average cross-correlations for the pairs of innovations corresponding to the series selected in Case 1were included as reference (“*” symbol). (c) Ratio of Global Influence (RGI) for innovations for $\lambda = 0$, 1 and 3. (d) RGI for temperature anomalies. The RGI for pixel i was defined as the fraction of the area of the planet for which the cross-correlation $|R_{ij}(\lambda)| > 0.2$ for all j.

approximations, there is good agreement with the empirical values. This supports the model and shows that the correlation structure has an intrinsic spatial component proportional to a_{ij}, and a temporal, memory-dependent component that depends on H_i and H_j.

3.2 Ratio of global influence

Empirical Orthogonal Functions (EOF) or Principal Component Analysis (PCA) decomposition techniques are often used to interpret the lagged cross-correlations (the matrices $R_{ij}(\Delta t)$, Fig. S2). This includes temperature teleconnection patterns, even though – if our model is valid – these have no Granger causality. An alternative to EOF teleconnection analysis is provided by network analysis (Donges et al., 2009a; Steinhaeuser et al., 2012; Tsonis, 2018; Tsonis et al., 2006; Yamasaki et al., 2008) based on the zero lag cross-correlations that define the area weighted connectivity (AWC).

Since the zero-lag statistics have no causal information, we generalized the AWC to nonzero lags by defining the ratio of global influence (RGI). The RGI for pixel i is the fraction of the area of the planet for which $\left|R_{ij}(\lambda)\right| > 0.2$, averaged over all j (for innovations $\left|\rho_{ij}(\lambda)\right| > 0.2$), for zero lags it is equal to the AWC. Values below 0.2 (dashed line in Fig. 1(b)) are considered to be of low influence. In climate networks, a threshold of 0.5 is typically used for defining connectivity, but innovation correlations – relevant to Granger causality – are much weaker, hence 0.2 was chosen

Figure 1(c) and (d) shows RGI maps for innovations and temperatures, respectively, for $\lambda = 0$, 1 and 3. For the innovations, almost all the correlation is lost for $\lambda > 0$, in agreement with Eq. (3): there is no significant influence on future values for any pixel. For $\lambda = 0$, we see that the region of largest innovation influence is the tropical Pacific where RGI $\approx 5\%$. For temperature anomalies (panel (d)), much larger correlations and RGIs are obtained. For $\lambda > 0$, almost all the influence from land disappears, but the ocean's influence is preserved up to around 1 year (not shown). Unsurprisingly, the tropical ocean has the largest correlations. As we mentioned earlier, this is a consequence of the long memory (large H, Fig. 1(a)).

Rigorously, the orthogonality condition (Eq. (12)) was derived for infinitely long time series with complete knowledge of the infinite past. In practice, we only have finite series and for each pixel, the memory effects of the unknown past will depend on the H values. For a fixed, finite length of past data, series with H closer to zero have more past information that can be borrowed. In the supporting information, we confirm that there is a small improvement in skill using a co-predictor series from different locations, but this improvement decreases with the memory, m, and is very small when sufficient past data points are used to build the predictor (see Fig. S3). For 20 months of past data, forecast skill improves by a maximum of 2%, which is roughly the noise level of the skill estimates (see Fig. S6). If only a few memory-steps are used, then the improvement in skill from borrowing memory from co-predictors is larger, but in all cases the combined predictor / co-predictor skill is lower than for the single long-memory predictor (see Figs. S4 and S5).

3.3 Simulations and emergent properties

At each pixel, m-StocSIPS has the same statistics as StocSIPS, which DRAL showed to be quite accurate. However in addition, m-StocSIPS takes into account the spatial correlations: to

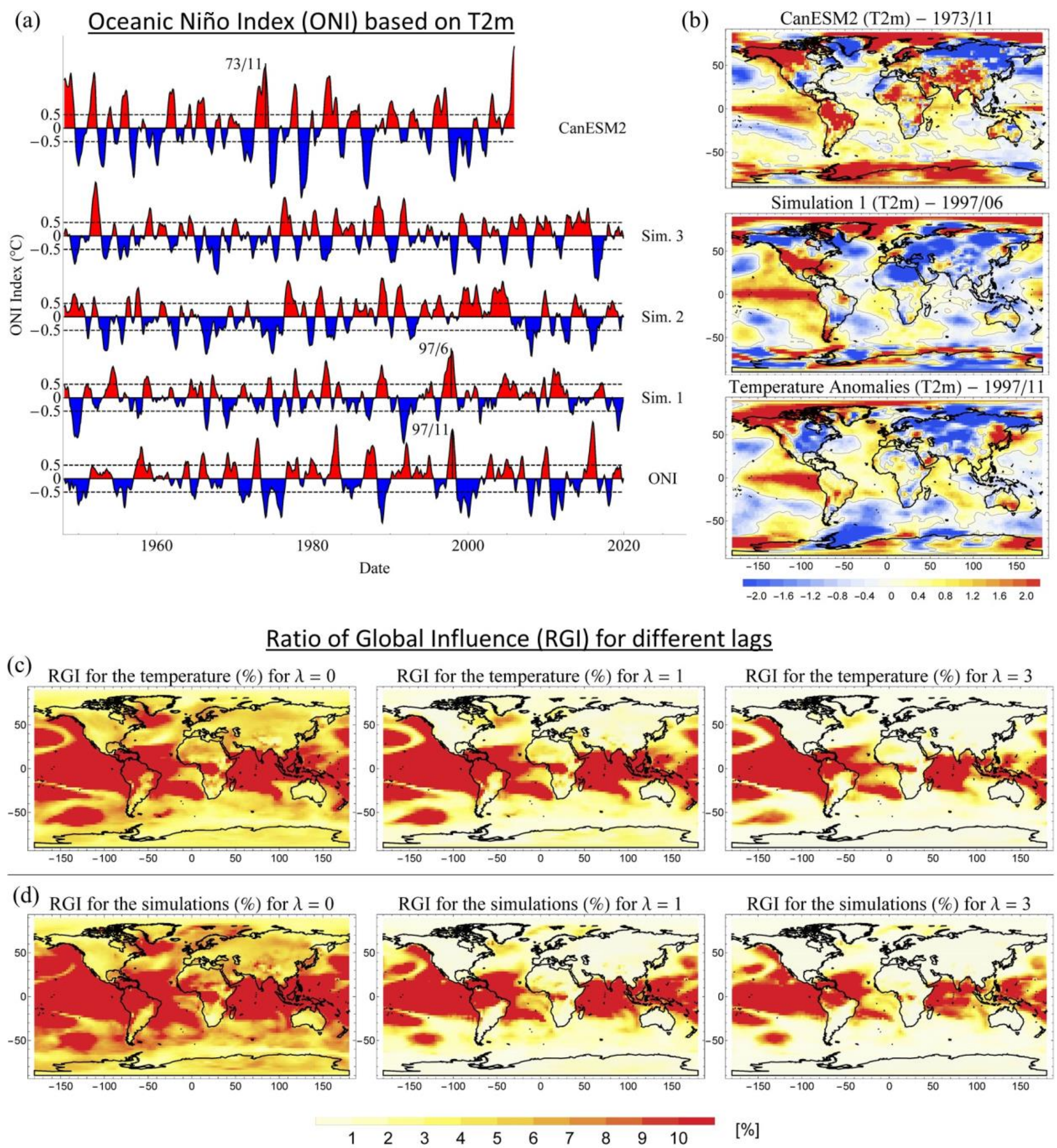

Fig. 2 (a) Comparison between series of the Oceanic Niño Index (ONI) derived for surface temperature (T2m) as the 3-month running mean of the average over the region (5°N-5°S, 170°W-120°W). In the bottom, we show the series computed from reanalysis (labelled as ONI); in the middle, samples from three different simulations (marked as Sim.1-3) and in the top, the index computed from one of the historical runs of the second generation Canadian Earth System Model (CanESM2) for the period 1948-2005. (b) Canonical anomaly pattern associated with the El Niño peaks marked in the series in Fig. 2(a) for each respective case. (c) Ratio of Global Influence (RGI) for the observational reference dataset for $\lambda = 0$, 1 and 3. (d) RGI for the Simulation 1 dataset.

be a realistic macroweather model it must also reproduce the observed spatial patterns including teleconnection networks (AWC, RGI), and El Niño events and indices. Since – just as in GCMs –

m-StocSIPS does not put these features in “by hand”, they are emergent model properties that are notoriously difficult to reproduce so that their realism (or lack thereof) provide stringent quality checks. Using m-StocSIPS simulations, (detailed in Sects. S2 and S7 of the supporting information) we now show that indeed, these emergent properties are well reproduced.

In order to compare StocSIPS space-time statistical structures to reanalysis and to GCM outputs, we produced simulations with the same resolutions and overall length as our reference NCEP/NCAR Reanalysis 1 dataset (864 months, 2.5° resolution). Although full movies of the model outputs are available (Movie S1), here we focus on El Niño events that are particularly difficult to simulate. First consider the Oceanic Niño Index (ONI) derived for surface temperature (T2m) as the 3-month running mean of the average over the region (5°N-5°S, 170°W-120°W), Fig 2(a). The bottom (“ONI”) is a reanalysis series above which are samples from three different m-StocSIPS realizations (“Sim.1-3”, middle). The top series is from a historical run of the CanESM2 GCM (CCCma, 2020), the ONI was estimated after standard detrending (but without variance adjustments).

Except for the larger GCM amplitude, the time series in Fig. 2(a) are difficult to distinguish. Both deterministic and stochastic simulations produce realistic-looking ONI anomalies sequences. More impressively, the stochastic simulations reproduce huge regional emergent patterns including El Niño and La Niña events. In Fig. 2(b), we see canonical El Niño anomaly patterns corresponding to El Niño peaks marked in Fig. 2(a) (see also Fig. S11 for map sequences). While the deterministic models explain these events as an expression of the dynamics implicit in the governing equations, in the stochastic model they emerge from random synchronizations from places sharing high H values (see Fig. 1(a)) and long ocean weather-macroweather transition times.

StocSIPS also produces realistic and emergent teleconnections patterns: RGI maps, see Fig. 2(c) and (d) for lags $\lambda = 0$, 1 and 3. Despite these striking spatial patterns, there is no Granger causality connecting any two points: the optimal predictor is obtained from the past of each individual series without any contribution from the teleconnection patterns. These correlations do not imply any Granger causality.

4 Conclusions

GCMs long range forecasting skill is low, and this has stimulated the development of stochastic alternatives often inspired by correlations. Two competing approaches have developed, one that primarily exploits teleconnections (space) with only a short memory in time (Markovian), the other – StocSIPS – that only exploits the long memory in time without using any spatial information. While Markovian models are approximately initial value problems GCMs are strictly so. In comparison, StocSIPS exploits the system’s (scaling) long range memory; it is a “past value” model. Although it is tempting to try to improve StocSIPS skill by using spatially correlated co-predictors, to be useful the correlations must also be causal.

Untangling correlations and causality is possible thanks to the precise notion of Granger causality. To apply this, we first extended StocSIPS to the full space-time process, m-StocSIPS, that has identical single pixel statistics but that includes pixel-pixel cross-correlations. Although m-StocSIPS’s time-lagged temperature cross-correlations are strong, they are generated by temporally uncorrelated innovations and it has no Granger causality. For a given position, past information from other locations cannot be used to improve on the forecast obtained as an optimal linear combination of past data: those correlations “were already used”. Whereas the ultimate causation in deterministic models is their initial conditions, the ultimate cause in StocSIPS is its white noise innovations.

To make this convincing, we provided a full space-time macroweather model, producing global space-time stochastic simulations at one month and 2.5° resolution over 864 months (Movie S1). Emergent model properties include realistic teleconnection networks and El Niño and La Niña events that have both realistic spatial warming patterns as well as Oceanic El Niño indices. For real data, only a finite length of the past series is known, but even in this case, we showed that by exploiting the correlations in the temperature series, maximum improvements in skill of only 1-2% are possible (and this is in the noise).

What then is the status of causal mechanisms such as those linking El Niño events to a wet central Asia (Barnston, 2014)? GCMs and StocSIPS provide ultimate causes that eschew such mechanisms. At best, it may be argued that ultimate causes initiate a causal chain in which an El Niño could be regarded as a proximate cause, and this proximate cause could presumably be captured in short memory empirical models. However, thanks to Granger causality we can now affirm that at best, at a given pixel i, the short memory models (partially) compensate for their under-exploitation of the memory by effectively “borrowing” the memory of particularly strong memory pixels j such as those in the El Niño region. StocSIPS obviates the need to borrow memory from pixel j by fully exploiting the memory at pixel i.

Acknowledgments

The authors thank Hydro-Québec for a bourse de doctorat Hydro- Québec en science (F213013R02). There are no conflicts of interest. We also thank R. Procyk, R. Hébert, and D. Clarke for useful discussions. NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their Web site at https://psl.noaa.gov/.

References

Ali, M. M. (1989). Tests for Autocorrelation and Randomness in Multiple Time Series. *Journal of the American Statistical Association*, *84*(406), 533. https://doi.org/10.2307/2289939

Andree, B. P. J. (2019). Probability, Causality and Stochastic Formulations of Economic Theory. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3422430

Barnston, A. (2014). How ENSO leads to a cascade of global impacts. Retrieved November 12, 2020, from https://www.climate.gov/news-features/blogs/enso/how-enso-leads-cascade-global-impacts

Biagini, F., Hu, Y., Øksendal, B., & Zhang, T. (2008). Stochastic Calculus for Fractional Brownian Motion and Applications. London: Springer London. https://doi.org/10.1007/978-1-84628-797-8

Blender, R., & Fraedrich, K. (2003). Long time memory in global warming simulations. *Geophysical Research Letters*, *30*(14). https://doi.org/10.1029/2003GL017666

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis*. Wiley. https://doi.org/10.1002/9781118619193

Brockwell, P. J., & Davis, R. A. (1991). *Time Series: Theory and Methods*. New York, NY:

Springer New York. https://doi.org/10.1007/978-1-4419-0320-4

Brown, P. T., & Caldeira, K. (2020). Empirical Prediction of Short-Term Annual Global Temperature Variability. *Earth and Space Science*, *7*(6). https://doi.org/10.1029/2020EA001116

Buchanan, M. (2012). Cause and correlation. *Nature Physics*, *8*(12), 852–852. https://doi.org/10.1038/nphys2497

Bunde, A., Eichner, J. F., Kantelhardt, J. W., & Havlin, S. (2005). Long-Term Memory: A Natural Mechanism for the Clustering of Extreme Events and Anomalous Residual Times in Climate Records. *Physical Review Letters*, *94*(4), 048701. https://doi.org/10.1103/PhysRevLett.94.048701

Bunge, M. (2017). *Causality and Modern Science* (4th ed.). Routledge. Retrieved from https://books.google.ca/books?id=xCIxDwAAQBAJ

CCCma. (2020). The second generation Canadian earth system model (CanESM2). Retrieved October 29, 2020, from https://climate-modelling.canada.ca/climatemodeldata/cgcm4/CanESM2/index.shtml

Coeurjolly, J., Amblard, P., & Achard, S. (2010). On multivariate fractional brownian motion and multivariate fractional Gaussian noise. In *2010 18th European Signal Processing Conference* (pp. 1567–1571).

Delignières, D. (2015). Correlation Properties of (Discrete) Fractional Gaussian Noise and Fractional Brownian Motion. *Mathematical Problems in Engineering*, *2015*, 1–7. https://doi.org/10.1155/2015/485623

Donges, J. F., Zou, Y., Marwan, N., & Kurths, J. (2009a). Complex networks in climate dynamics. *The European Physical Journal Special Topics*, *174*(1), 157–179. https://doi.org/10.1140/epjst/e2009-01098-2

Donges, J. F., Zou, Y., Marwan, N., & Kurths, J. (2009b). The backbone of the climate network. *EPL (Europhysics Letters)*, *87*(4), 48007. https://doi.org/10.1209/0295-5075/87/48007

Eden, J. M., van Oldenborgh, G. J., Hawkins, E., & Suckling, E. B. (2015). A global empirical system for probabilistic seasonal climate prediction. *Geoscientific Model Development*, *8*(12), 3947–3973. https://doi.org/10.5194/gmd-8-3947-2015

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424. https://doi.org/10.2307/1912791

Gripenberg, G., & Norros, I. (1996). On the prediction of fractional Brownian motion. Journal of Applied Probability, 33(2), 400–410. https://doi.org/10.1017/S0021900200099812

Hébert, R., & Lovejoy, S. (2018). Regional Climate Sensitivity- and Historical-Based Projections to 2100. *Geophysical Research Letters*, *45*(9), 4248–4254. https://doi.org/10.1002/2017GL076649

Hirchoren, G. A., & Arantes, D. S. (1998). Predictors for the discrete time fractional Gaussian processes. In ITS'98 Proceedings. SBT/IEEE International Telecommunications Symposium (Cat. No.98EX202) (pp. 49–53). IEEE. https://doi.org/10.1109/ITS.1998.713090

Hosking, J. R. M. (1980). The Multivariate Portmanteau Statistic. *Journal of the American Statistical Association*, *75*(371), 602. https://doi.org/10.2307/2287656

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Li, W. K., & McLeod, A. I. (1981). Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, *43*(2), 231–239. Retrieved from http://www.jstor.org/stable/2984853

Lovejoy, S. (2018). Spectra, intermittency, and extremes of weather, macroweather and climate. *Scientific Reports*, *8*(1), 12697. https://doi.org/10.1038/s41598-018-30829-4

Lovejoy, S. (2019). Fractional relaxation noises, motions and the fractional energy balance equation. *Nonlin. Processes Geophys. Discuss., in review*. https://doi.org/https://doi.org/10.5194/npg-2019-39

Lovejoy, S. (2020). The fractional heat equation. *Geophysical Research Letters*, (in review).

Lovejoy, S., & de Lima, M. I. P. (2015). The joint space-time statistics of macroweather precipitation, space-time statistical factorization and macroweather models. *Chaos*, *25*(7), 075410. https://doi.org/10.1063/1.4927223

Lovejoy, S., & Schertzer, D. (2013). *The Weather and Climate: Emergent Laws and Multifractal Cascades*. *The Weather and Climate: Emergent Laws and Multifractal Cascades*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139093811

Lovejoy, S., Del Rio Amador, L., & Hébert, R. (2015). The ScaLIng Macroweather Model (SLIMM): Using scaling to forecast global-scale macroweather from months to decades. *Earth System Dynamics*, *6*(2), 637–658. https://doi.org/10.5194/esd-6-637-2015

Lovejoy, S., Del Rio Amador, L., & Hébert, R. (2018). Harnessing Butterflies: Theory and Practice of the Stochastic Seasonal to Interannual Prediction System (StocSIPS). In *Advances in Nonlinear Geosciences* (pp. 305–355). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-58895-7_17

Lovejoy, S., Procyk, R., Hébert, R., & Del Rio Amador, L. (2020). The Fractional Energy Balance Equation. *Quarterly Journal of the Royal Meteorological Society*, (in review).

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., & Schellnhuber, H. J. (2014). Very early warning of next El Niño. *Proceedings of the National Academy of Sciences*, 201323058. https://doi.org/10.1073/pnas.1323058111

Lyon, B., & Barnston, A. G. (2005). ENSO and the Spatial Extent of Interannual Precipitation Extremes in Tropical Land Areas. *Journal of Climate*, *18*(23), 5095–5109. https://doi.org/10.1175/JCLI3598.1

Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian Motions, Fractional Noises and Applications. SIAM Review, 10(4), 422–437. https://doi.org/10.1137/1010093

NCEP/NCAR. (2020). NCEP/NCAR Reanalysis 1. Retrieved January 3, 2020, from https://www.esrl.noaa.gov/psd/

Palma, W. (2007). *Long-Memory Time Series*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Palmer, T. N. (2019). Stochastic weather and climate models. *Nature Reviews Physics*, *1*(7), 463–471. https://doi.org/10.1038/s42254-019-0062-2

Poskitt, D. S., & Tremayne, A. R. (1982). Diagnostic Tests for Multiple Time Series Models. *The Annals of Statistics*, *10*(1), 114–120. Retrieved from http://www.jstor.org/stable/2240503

Del Rio Amador, L., & Lovejoy, S. (2019). Predicting the global temperature with the Stochastic Seasonal to Interannual Prediction System (StocSIPS). *Climate Dynamics*, *53*(7–8), 4373–4411. https://doi.org/10.1007/s00382-019-04791-4

Del Rio Amador, L., & Lovejoy, S. (2020). Using regional scaling for temperature forecasts with the Stochastic Seasonal to Interannual Prediction System (StocSIPS). *Climate Dynamics*, (submitted).

Rypdal, K., Østvand, L., & Rypdal, M. (2013). Long-range memory in Earth's surface temperature on time scales from months to centuries. *Journal of Geophysical Research: Atmospheres*, *118*(13), 7046–7062. https://doi.org/10.1002/jgrd.50399

Steinhaeuser, K., Ganguly, A. R., & Chawla, N. V. (2012). Multivariate and multiscale dependence in the global climate system revealed through complex networks. *Climate Dynamics*, *39*(3–4), 889–895. https://doi.org/10.1007/s00382-011-1135-9

Straškraba, M. (2007). K. W. Hipel and A. I. McLeod: Time Series Modelling of Water Resources and Environmental System. 1013 pp. Amsterdam, Elsevier, 1994, XXXII. ISBN 0–444–89270–2. Available from Elsevier Science Publishers B. V., Order Fulfilment Department. P.O. Box 211, 10. *Internationale Revue Der Gesamten Hydrobiologie Und Hydrographie*, *80*(1), 60–60. https://doi.org/10.1002/iroh.19950800107

Sugihara, G., May, R., Ye, H., Hsieh, C. -h., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting Causality in Complex Ecosystems. *Science*, *338*(6106), 496–500. https://doi.org/10.1126/science.1227079

Tsonis, A. A. (2018). Insights in Climate Dynamics from Climate Networks. In *Advances in Nonlinear Geosciences* (pp. 631–649). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-58895-7_29

Tsonis, A. A., Swanson, K. L., & Roebber, P. J. (2006). What Do Networks Have to Do with Climate? *Bulletin of the American Meteorological Society*, *87*(5), 585–596. https://doi.org/10.1175/BAMS-87-5-585

Varotsos, C. A., Efstathiou, M. N., & Cracknell, A. P. (2013). On the scaling effect in global surface air temperature anomalies. *Atmospheric Chemistry and Physics*, *13*(10), 5243–5253. https://doi.org/10.5194/acp-13-5243-2013

Wold, H. (1938). A Study in Analysis of Stationary Time Series. *Journal of the Royal Statistical Society*.

Yamasaki, K., Gozolchiani, A., & Havlin, S. (2008). Climate Networks around the Globe are Significantly Affected by El Niño. *Physical Review Letters*, *100*(22), 228501. https://doi.org/10.1103/PhysRevLett.100.228501