Geophysical Research Letters

RESEARCH LETTER
10.1029/2020GL092147

Long-Range Forecasting as a Past Value Problem: Untangling Correlations and Causality With Scaling

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

[1]Physics, McGill University, Montreal, Que, Canada

Key Points:

- Granger causality shows that spatial correlations in temperature cannot be used to improve on memory-based predictions of infinite series
- Scaling-based long-range stochastic forecasting is a past value problem not an initial value problem
- Real-world statistics and teleconnection patterns can be reproduced with stochastic simulations with a total lack of causal relationships

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

L. Del Rio Amador,
delrio@physics.mcgill.ca

Citation:

Del Rio Amador, L., & Lovejoy, S. (2021). Long-range forecasting as a past value problem: Untangling correlations and causality with scaling. *Geophysical Research Letters*, *48*, e2020GL092147. https://doi.org/10.1029/2020GL092147

Received 15 DEC 2020
Accepted 7 APR 2021

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.

Plain Language Summary For forecasts less than about 10 days, Numerical Weather Prediction (NWP) and General Circulation Models (GCMs) have been highly successful, yet for longer ranges, their skill is disappointing. This has motivated the development of alternatives that are based on either the strong spatial correlations - teleconnection patterns such El Niño events - or on the long memories whereby the atmospheric state at any moment is strongly influenced by its own past. In particular, a model only using the long memory already rivals GCM monthly and seasonal temperature forecasts: The Stochastic Seasonal to Interannual Prediction System (StocSIPS).

In this paper, we answer the question of whether StocSIPS skill can be improved by also using teleconnections. We do this by developing the space-time m-StocSIPS model that is optimally forecast by StocSIPS. In m-StocSIPS, spatial co-predictors do not improve the skill: There is no causal relation between different locations useful for long-range predictions. Although m-StocSIPS has strong spatial correlations and reproduces teleconnection patterns including El Niño events, they cannot be used to improve the long-memory StocSIPS forecasts. The teleconnections " were already used" to build the history at every location, which is enough to produce the optimal forecast.

1. Introduction

For forecasts over the weather regime–below the ≈10 days 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. Physically based 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 (eg., 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. Temporal scaling is associated with the wide range spatial scaling that is respected by the empirical data, the governing equations and hence GCMs (the reviews Lovejoy & Schertzer, 2013; Palmer, 2019).

While GCMs are zero-memory, initial value models based purely on the spatial information at $t = 0$ (and the knowledge of the dynamics), stochastic models based on teleconnections are almost as extreme using only data from a few months–they are short (exponential) memory Markovian models. The scaling (power law) memory Stochastic Seasonal to Interannual Prediction System (StocSIPS) model is at the opposite extreme (Del Rio Amador & Lovejoy, 2019, 2021). For each pixel, it exclusively uses historical past data from that pixel to forecast the future without any 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 if 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 series from other locations that are strongly spatially correlated do not improve the skill. Teleconnection correlations may therefore be seductive, but having no Granger causality, they are misleading.

2. Methods

2.1. Stochastic Modeling of the Temperature Anomalies

Macroweather temperature anomalies at position $\mathbf{x}$ (after removing the annual cycle) can be modeled as a trend-stationary process:

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

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 modeling each position independently using an univariate representation the regional StocSIPS model presented in Del Rio Amador and Lovejoy, 2021, hereafter DRAL. However, to investigate whether forecasts for individual series can be improved using other data, a multivariate framework is needed. An arbitrary 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" indicates 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 $\gamma_i(t) = 0$, $\gamma_i^2(t) = 1$ and cross-correlation matrix:

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

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 thus:

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

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 behavior 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. Other Gaussian scaling models could be used (such as auto regressive fractionally integrated moving average, ARFIMA, models), but they are unnecessarily complex.

The fGn–based StocSIPS model was first developed for monthly and seasonal forecast of globally averaged temperature (Del Rio Amador & Lovejoy, 2019; Lovejoy et al., 2015). 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, σ_{T_i} is the standard deviation, c_{H_i} is a normalization constant, and δ_{ij} is the Kronecker δ. $H_i \in (-1,0)$ is the fluctuation exponent that characterizes the scaling of the fluctuations in time ($H_i + 1$ is the Hurst parameter, see the discussion after Equation S2 in the supporting information). 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 Equation 6 implies that the temperature at grid point "i" is an fGn with H_i and σ_{T_i}.

In DRAL it was shown that the fGn model (Equation 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 modeled as a fractional Brownian motion (fBm) process with fluctuation exponent $H_i \in (0,1)$ (see Figure 1a), although cut-off at multi-annual scales (note that the fBm and fGn fluctuation exponents respect $H_{\mathrm{fBm}} = H_{\mathrm{fGn}} + 1$). Both cases are high-frequency approximations of the more general fractional relaxation noise (fRn) process, introduced in (Lovejoy, 2019; Lovejoy et al., 2021).

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. m-StocSIPS itself has a physical interpretation as a high frequency approximation to the regional Fractional Energy Balance Equation (Lovejoy, 2021a, 2021b; Lovejoy et al., 2021).

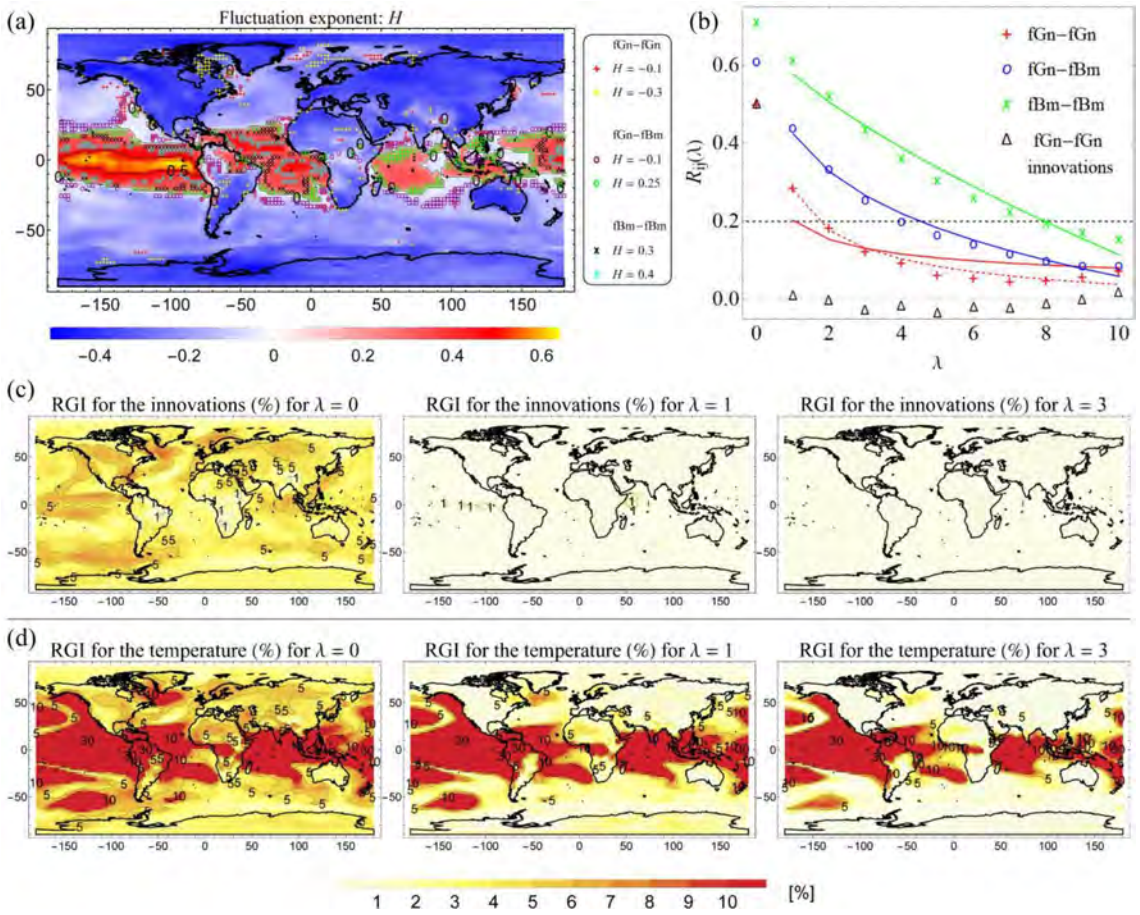

Figure 1. (a) Maximum likelihood estimates of the fluctuation exponent, (h) The grid points forming the pairs used to calculate the average ensemble cross-correlations (shown in (b)) are marked as: "+" for fGn-fGn, "o" for fGn-fBm and "x" for fBm-fBm. The colors indicate the values of (h) (b) Average cross-correlations for $\lambda = 0$–10 for the Cases 1, 2, and 3 (described in the text), with the corresponding fits from Equations 7–8 (which only apply for $\lambda \geq 1$). 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 1 were included as reference ("Δ" 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.

m-StocSIPS is defined by $N(N+3)/2$ parameters; in comparison, a vector autoregressive order m model (VAR(m)) needs mN^2 values (Box et al., 2008; Brockwell & Davis, 1991) and for long-memory processes, m is large. These "black box" type models suffer from opaque physical interpretations, and the large number of VAR parameters makes them unstable and subject to overfitting. The same is true for general vector autoregressive-moving average VARMA (m,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 Equation 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 (Equation 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 Figure S2 in the supporting information) may be enough. Our results indicate that m-StocSIPS is a good approximation, confirmed in Section 3.3 using global simulations that convincingly reproduce the space-time patterns (Figure 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.

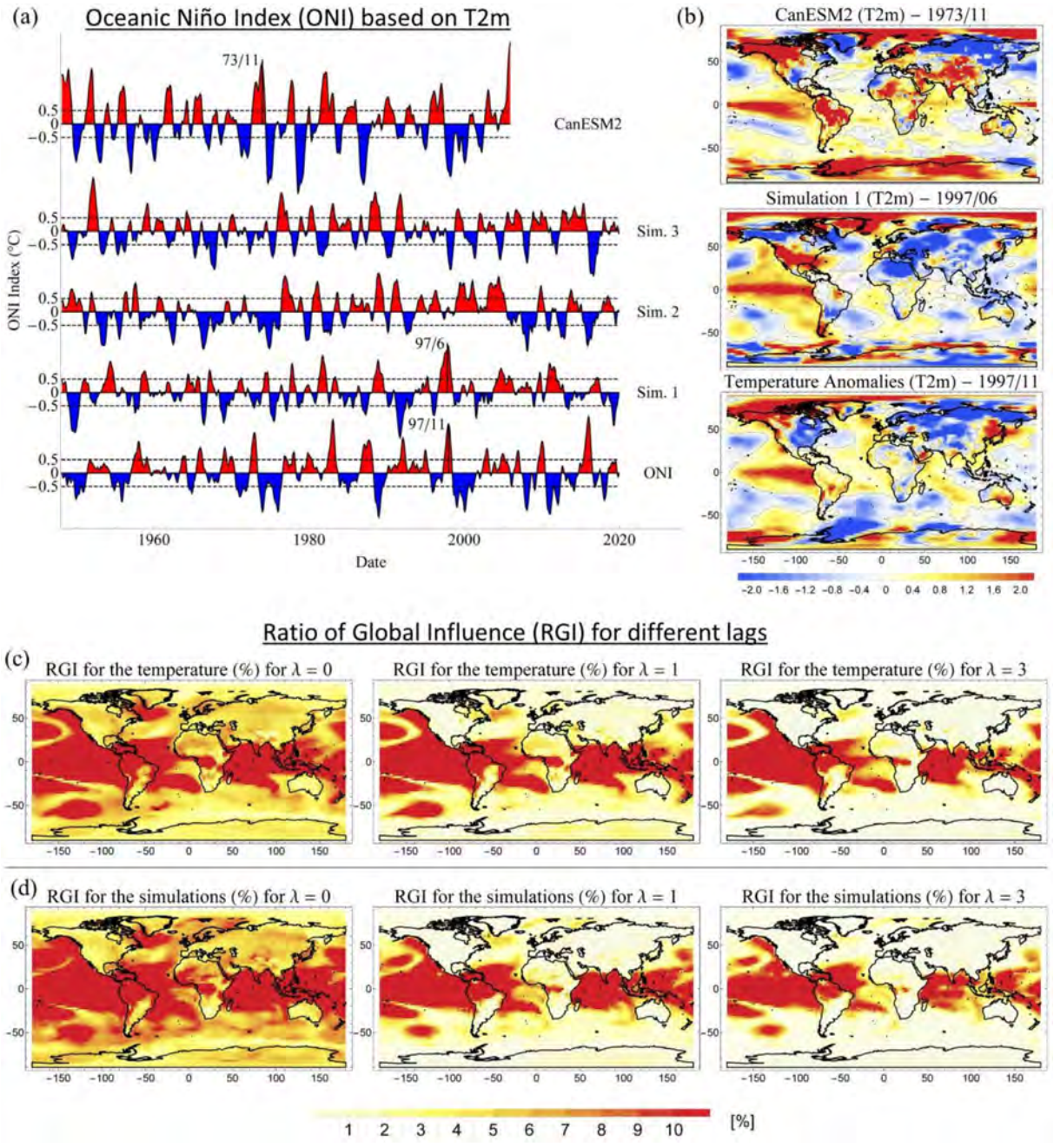

Figure 2. (a) Comparison between series of the Oceanic Niño Index (ONI) derived for surface temperature (T2m) as the 3-months running mean of the average over the region (5°N–5°S, 170°W–120°W). In the bottom, we show the series computed from reanalysis (labeled as ONI); in the middle, samples from three 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 Figure 2(a) for each respective case. (c) Ratio of Global Influence (RGI) for the observational reference data set for $\lambda = 0$, 1, and 3. (d) RGI for the Simulation 1 data set.

2.2. 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 Equation 6 in Equation 4, see supporting information Equation 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, that is, 1 month) are:

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

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

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

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

for $\Delta t \ll \tau_r^{ij}$, where τ_r^{ij} is a characteristic relaxation time (Del Rio Amador & Lovejoy, 2021; Lovejoy, 2019; Lovejoy et al., 2018), and φ_{H_i,H_j} and ϕ_{H_i,H_j} are proportionality constants that depend on H_i and H_j (see Equation 25 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 (Lovejoy et al., 2021).

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 Equation 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, for example, to a wet Central Asia. 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 $T_i\left(t\right)$ based on its own past ($T_i\left(s\right)$ 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. To our knowledge this is the first application of Granger causality to long memory processes.

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; Hipel & McLeod, 1994; Palma, 2007; Wold, 1938), that states that the error of the optimal predictor (in a mean square error sense) is orthogonal to any possible estimator:

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

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 (Equation 2) and given a diagonal kernel $\kappa_{ij}(t)$ as in Equation 6, the optimal predictor is:

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

with error:

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

Optimality follows by first noting that E_i only depends on future innovations $\gamma_i(t'')$ $(t'' > 0)$, while the estimator, $\hat{T}_i(t)$, depends only on past innovations $\gamma_i(t')$ $(t' < 0)$. Then, since the white noise innovations are δ -correlated in time (Equation 3), for any i, j we have:

$$\langle T_j(s) E_i(t) \rangle = 0; \quad s \langle 0, t \rangle 0 \tag{12}$$

Equation 12 implies 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 Equation 10. Hence, the predictor (Equation 10) is optimal given the full field $T(\mathbf{x}, t)$ for $t \leq 0$. This is a precise statement of Granger causality for the m-StocSIPS model. Although there are large cross-correlations inherited from the innovation matrix a_{ij} (Equations 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: $73 \times 144 = 10{,}512$ 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 Equation 2 (see the supporting material). 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 10,512 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 are negligible (Equation 3). For this analysis, we obtained the lagged cross-correlation matrices involving the 10,512 grid points for the innovations, $\rho_{ij}(\Delta t)$, and for the temperature variability, $R_{ij}(\Delta t)$, for Δt from 0 to 12 months (see the supporting information, Figure S2). While temperature correlations decrease with Δt, in contrast to innovation cross-correlations, large values may easily be obtained for relatively large lags (Equations 7 and 8). For $\Delta t = 0$, innovation correlations ($\rho_{ij}(0) = a_{ij}$) can be large yet–following the discrete version of the time-independence condition Equation 3 – at $\Delta t = 1$ month, almost all correlation has been lost.

We can also check that Equations 7 and 8 are good approximations to the empirical $R_{ij}(\Delta t)$. Figure 1a shows the results for ensembles with similar a_{ij}, H_i and H_j values ($a_{ij} = 0.5 \pm 0.025$ gives 10,490 pairs). Comparisons are shown for the three cases (fGn-fGn, fGn-fBm and fBm-fBm):

Case 1: fGn-fGn (marked as "+" in Figure 1a), 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.

Figure 1b shows the average cross-correlations functions of the lag $\lambda = \Delta t / \tau$ ($\tau = 1$ month), with fits from Equations 7 and 8 (which only apply for $\lambda \geq 1$). For case 1, we included the dashed red curve corresponding to higher order corrections for fRn processes (Lovejoy, 2019; Lovejoy et al., 2021). The small values of the cross-correlation innovation pairs ("Δ" in the figure) confirm the independence of these series. Although the expressions (Equations 7 and 8) are only first order 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, H dependent memory component.

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 $R_{ij}(\Delta t)$, Figure 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 $|R_{ij}(\lambda)| > 0.2$, averaged over all j (for innovations $|\rho_{ij}(\lambda)| > 0.2$), for zero lags it is equal to the AWC. Values below 0.2 (dashed line in Figure 1b) 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 based on their empirical distribution. Although the exact value or the threshold is relatively arbitrary, the main idea was to illustrate the difference in connectivity between the temperature anomalies and the innovations (almost zero) for $\lambda \geq 1$.

Figures 1c and 1d 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 Equation 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, Figure 1a).

The orthogonality condition (Equation 12) was derived for infinitely long time series with complete knowledge of the infinite past. However, for finite series, the memory effects 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 Figure 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 Figure 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 Figures 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 be a realistic macroweather model it must also reproduce the observed spatial patterns including teleconnection networks (AWC, RGI), and El Niño events and indices. As with GCMs, m-StocSIPS does not put these features in "by hand," they are emergent model properties that are notoriously difficult to reproduce and their realism provides stringent quality checks. Using m-StocSIPS simulations (detailed in Sections 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 data set (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), Figure 2a. 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 Figure 2a are difficult to distinguish. Both deterministic and stochastic simulations produce realistic-looking ONI anomaly sequences. More impressively, the stochastic simulations reproduce huge regional emergent patterns including El Niño and La Niña events. In Figure 2b, we see canonical El Niño anomaly patterns corresponding to El Niño peaks marked in Figure 2a (see also Figure 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 spontaneously in this region with large innovation cross-correlations and with series sharing high H values (see Figure 1a). These highly persistent series of anomalies spend long periods (from a few months to a few years) before changing sign, so eventually they share the same phases and produce El Niño or La Niña patterns.

StocSIPS also produces realistic and emergent teleconnections patterns: RGI maps, see Figures 2c and 2d 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 strong correlations do not imply any Granger causality.

4. Conclusions

GCM 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 approximate-

ly 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 stochastic 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.

Data Availability Statement

Datasets for this research are available in these in-text data citation references (CCCma, 2020; NCEP/NCAR, 2020).

Acknowledgments
The authors thank Hydro-Québec for a bourse de doctorat Hydro- Québec en science (F213013R02). There are no conflicts of interest. The authors thank R. Procyk, R. Hébert, and D. Clarke for useful discussions.

References

Ali, M. M. (1989). Tests for autocorrelation and randomness in multiple time series. *Journal of the American Statistical Association*, *84*(406), 533–540. https://doi.org/10.1080/01621459.1989.10478800

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 from https://www.climate.gov/news-features/blogs/enso/how-enso-leads-cascade-global-impacts

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*. 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. 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. https://doi.org/10.4324/9781315081656

CCCma. (2020). The second generation Canadian earth system model (CanESM2). Retrieved 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. *2010 18th European Signal Processing Conference*, 1567–1571.

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. (2021). Using regional scaling for temperature forecasts with the Stochastic Seasonal to Interannual Prediction System (StocSIPS). *Climate Dynamics*. https://doi.org/10.1007/s00382-021-05737-5

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. *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
Hipel, K. W., & McLeod, A. I. (1994). Time series modeling of water resources and environmental systems. In K. W. Hipel, & A. I. McLeod, (Eds.), *Time series modeling of water resources and environmental systems* (Vol. 45, pp. 1–1013). Elsevier. https://doi.org/10.1016/S0167-5648(08)70651-8
Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, *75*(371), 602–608. https://doi.org/10.1080/01621459.1980.10477520
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*, *43*(2), 231–239. https://doi.org/10.1111/j.2517-6161.1981.tb01175.x
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. *Nonlinear Processes in Geophysics Discussion*. *18*, 675–695. https://doi.org/10.5194/npg-2019-39
Lovejoy, S. (2021a). The half-order energy balance equation, Part 1: The homogeneous HEBE and long memories. *Earth System Dynamics*, (in press). https://doi.org/10.5194/esd-2020-12
Lovejoy, S. (2021b). The half-order energy balance equation, Part 2: The inhomogeneous HEBE and 2D energy balance models. *Earth System Dynamics Discussion*, (in press). https://doi.org/10.5194/esd-2020-13
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., 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). Springer International Publishing. https://doi.org/10.1007/978-3-319-58895-7_17
Lovejoy, S., Procyk, R., Hébert, R., & Del Rio Amador, L. (2021). The Fractional Energy Balance Equation. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.4005
Lovejoy, S., & Schertzer, D. (2013). *The weather and climate: Emergent laws and multifractal cascades. The weather and climate: Emergent laws and multifractal cascades*. Cambridge University Press. https://doi.org/10.1017/cbo9781139093811
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 of the United States of America*, 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
NCEP/NCAR. (2020). NCEP/NCAR reanalysis 1. Retrieved from https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html
Palma, W. (2007). *Long-memory time series*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470131466
Palmer, T. N. (2019). Stochastic weather and climate models. *Nature Review 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. *Annals of Statistics*, *10*(1), 114–120. https://doi.org/10.1214/aos/1176345694
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 - D: 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
Sugihara, G., May, R., Ye, H., Hsieh, C., 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). 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

References From the Supporting Information

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
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).

Grassberger, P., & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, *9*(1–2), 189–208. https://doi.org/10.1016/0167-2789(83)90298-1

Graves, T., Gramacy, R., Watkins, N., & Franzke, C. (2017). A Brief History of Long Memory: Hurst, Mandelbrot and the Road to ARFIMA, 1951-1980. *Entropy*, *19*(9), 437. https://doi.org/10.3390/e19090437

Gripenberg, G., & Norros, I. (1996). On the prediction of fractional Brownian motion. *J. Appl. Probab.*, *33*(2), 400–410. https://doi.org/10.1017/s0021900200099812

Hentschel, H. G. E., & Procaccia, I. (1983). The infinite number of generalized dimensions of fractals and strange attractors. *Physica D: Nonlinear Phenomena*, *8*(3), 435–444. https://doi.org/10.1016/0167-2789(83)90235-x

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

Hurst, H. E. (1951). Long-Term Storage Capacity of Reservoirs. *T. Am. Soc. Civ. Eng.*, *116*(1), 770–799. https://doi.org/10.1061/taceat.0006518

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

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

Lovejoy, S., & Schertzer, D. (2012). Haar wavelets, fluctuations and structure functions: Convenient choices for geophysics. *Nonlin. Processes Geophys.*, *19*(5), 513–527. https://doi.org/10.5194/npg-19-513-2012

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 Syst. Dynam.*, *6*(2), 637–658. https://doi.org/10.5194/esd-6-637-2015

Lovejoy, S., Procyk, R., Hébert, R., & Del Rio Amador, L. (2021). The Fractional Energy Balance Equation. *Quarterly Journal of the Royal Meteorological Society*, qj.4005. https://doi.org/10.1002/qj.4005

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

Del Rio Amador, L., & Lovejoy, S. (2019). Predicting the global temperature with the Stochastic Seasonal to Interannual Prediction System (StocSIPS). *Clim Dyn*, *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*, (under review).

Schertzer, D., & Lovejoy, S. (1983). On the Dimension of Atmospheric Motions. In *IUTAM Symp. on Turbulence and Chaotic Phenomena in Fluids*. Japan: Kyoto.

Schertzer, D., & Lovejoy, S. (1987). Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res.*, *92*(D8), 9693. https://doi.org/10.1029/jd092id08p09693

1-25

correct