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

Lenin Del Rio Amador[1] and Shaun M. Lovejoy[1]

[1]McGill University

November 23, 2022

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.

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 spatial correlations of temperature cannot improve memory-based predictions of individual infinite time series.
- The statistics and teleconnection patterns of the real-world can be reproduced with stochastic simulations without 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:

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

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 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):

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

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:

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

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:

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

hence the cross-correlation is:

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

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:

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

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:

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 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,

196 2017). In this fairly qualitative view, the ultimate cause – the initial conditions – triggers a causal
197 chain of events in which El Niño is a "proximate" link leading to a wet season in Central Asia.

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

207 The precise tool needed to clarify stochastic causality issues is Granger causality
208 (Granger, 1969). We say that the temperature T_j at location j fails to Granger-cause the
209 temperature T_i, if for all future times $t > 0$, the mean square error (MSE) of a forecast of $T_i(t)$
210 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
211 both $T_i(s)$ and $T_j(s)$. The notion of Granger causality is intuitive and provides a much more
212 rigorous criterion for causation than simple lagged cross-correlations. While other notions of
213 causality exist, Granger causality does imply forecasting ability, which is our only concern here.

214 We now investigate the Granger causality of m-StocSIPS. A necessary and sufficient
215 condition for the optimality of an estimator is given by the orthogonality principle (Box et al.,
216 2008; Brockwell & Davis, 1991; Palma, 2007; Straškraba, 2007; Wold, 1938), that states that the
217 error of the optimal predictor (in a mean square error sense) is orthogonal to any possible
218 estimator:

219 $$\left\langle \hat{T}_i(t) E_i(t) \right\rangle = 0, \tag{9}$$

220 where $\hat{T}_i(t)$ is the temperature predictor for position i at a future time $t > 0$ and $E_i(t) = T_i(t) -$
221 $\hat{T}_i(t)$ is the error.

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

224 $$\hat{T}_i(t) = \int_{-\infty}^{0} \kappa_{ii}(t - t') \gamma_i(t') dt', \tag{10}$$

225 with error:

226 $$E_i(t) = \int_{0}^{t} \kappa_{ii}(t - t'') \gamma_i(t'') dt''. \tag{11}$$

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

230 $$\left\langle T_j(s) E_i(t) \right\rangle = 0; \qquad s < 0,\ t > 0. \tag{12}$$

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

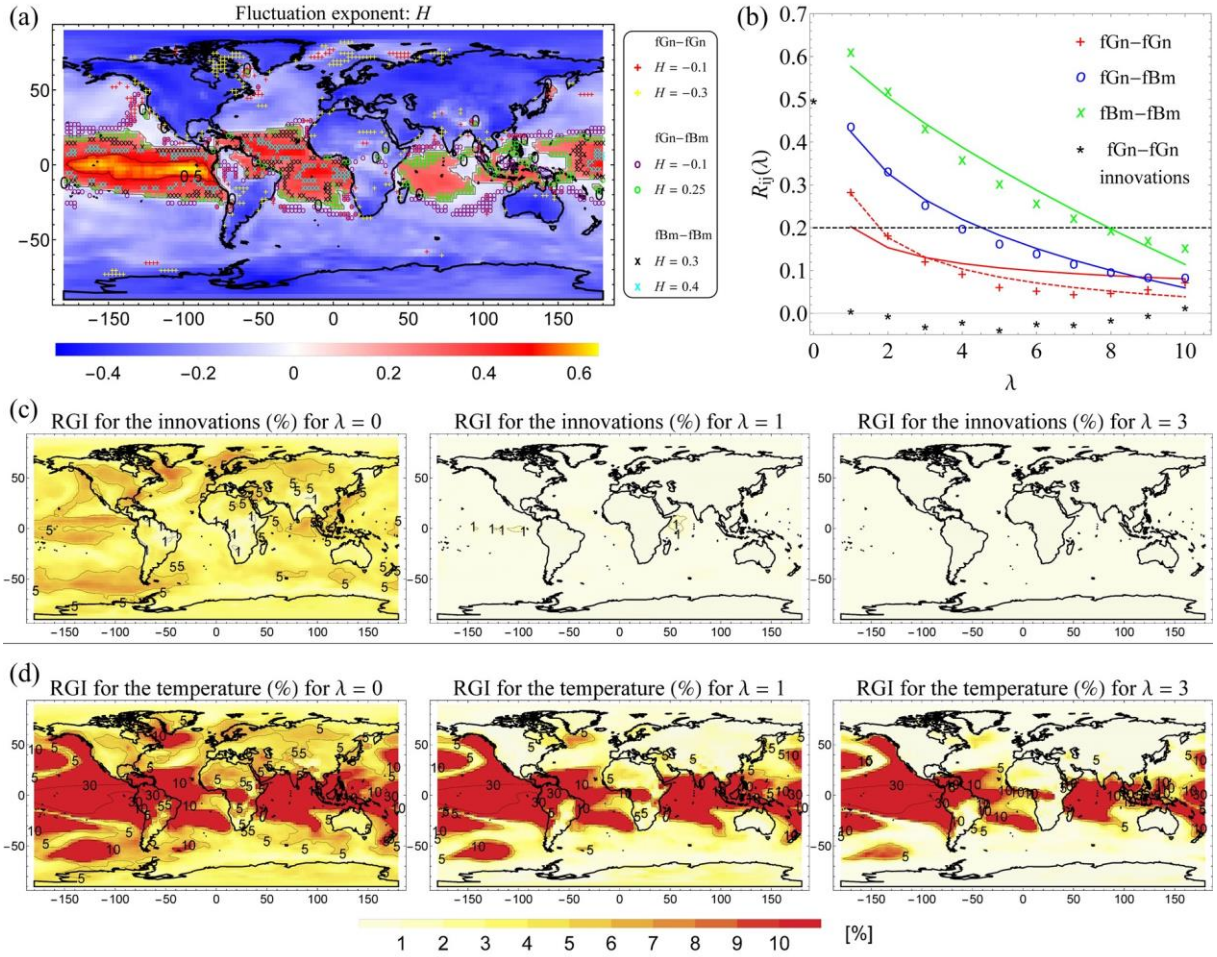

Fig. 1 (a) Maximum likelihood estimates of the fluctuation exponent, H. The grid points forming the pairs used to calculate the average ensemble cross-correlations (shown in (b)) are marked as: "+" for fGn-fGn, "o" for fGn-fBm and "x" for fBm-fBm. The colours indicate the values of H. (b) Average cross-correlations 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 $|R_{ij}(\lambda)| > 0.2$, 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 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

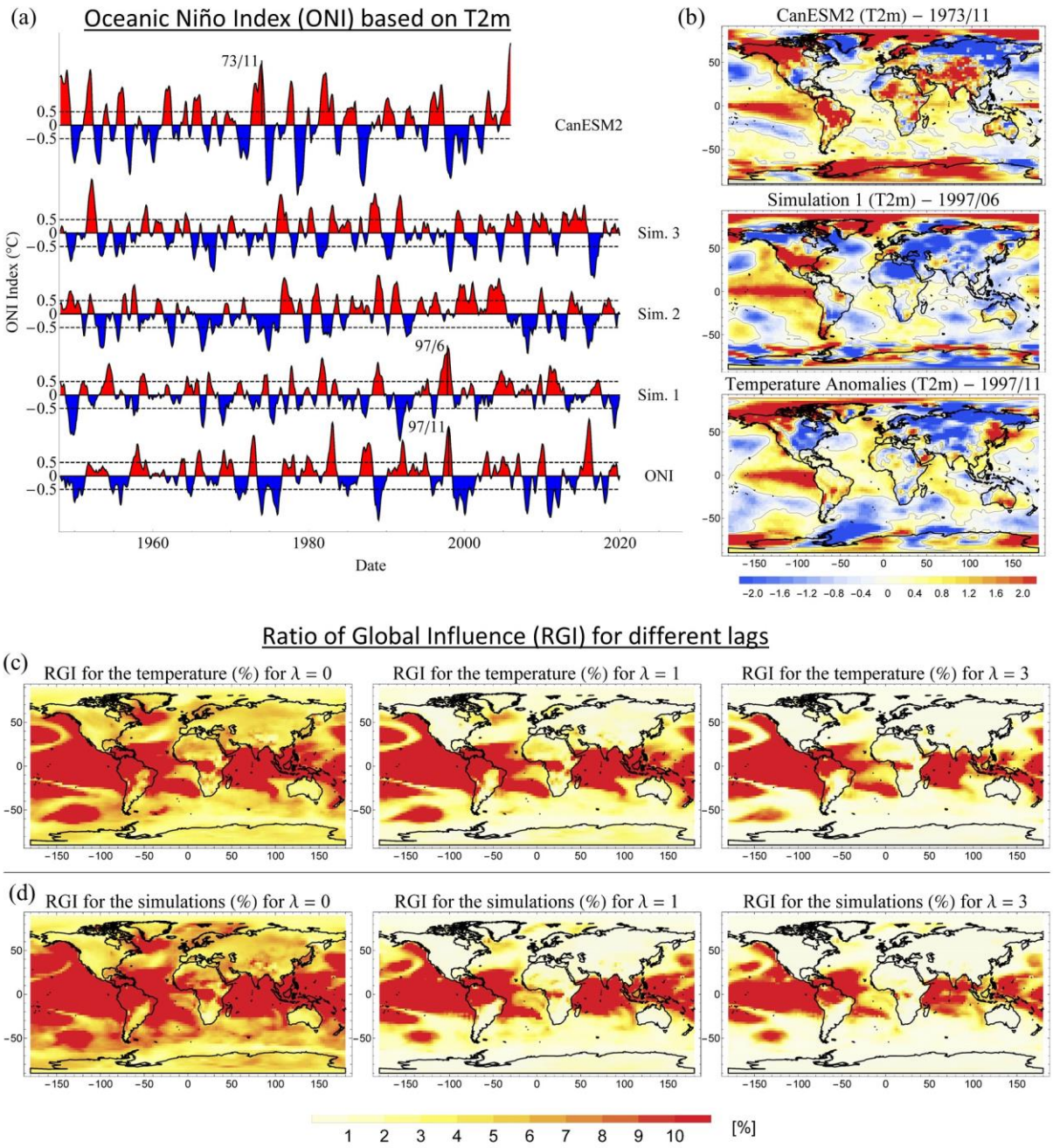

Fig. 2 (a) Comparison between series of the Oceanic Niño Index (ONI) derived for surface temperature (T2m) as the 3-month running mean of the average over the region (5°N-5°S, 170°W-120°W). In the bottom, we show the series computed from reanalysis (labelled as ONI); in the middle, samples from three 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^o 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. Datasets for this research are available in these in-text data citation references: (CCCma, 2020; NCEP/NCAR, 2020)

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

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

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

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

NCEP/NCAR. (2020). NCEP/NCAR Reanalysis 1. Retrieved January 3, 2020, from https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html

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*, (under review).

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

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).

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

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

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

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

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, (under review).

Geophysical Research Letters

Supporting Information for

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)

Contents of this file

Additional Supporting Information (Files uploaded separately)

Introduction

In (Del Rio Amador & Lovejoy, 2019, 2020), the stochastic natural variability component of globally averaged and regional temperature was represented as a fractional Gaussian noise (fGn) process. In order to determine Granger causality, we need to construct a space-time multivariate fGn process that reproduces the cross-correlations. This appendix describes the main technical details.

The main properties of fGn relevant for the present paper are summarized in the following Text S1-S3. We derive expressions for the cross-correlation function for two fGn processes in Text S4. The empirical cross-correlation matrices for the natural temperature variability and for the corresponding innovations for all the 10512 datapoints are presented in Text S5, showing that the innovations have negligible cross-correlations for nonzero lags. In Text S6 we analyze the influence of co-predictors in the forecasts under this condition. Finally, Text 7 gives some specific details for the

simulations of the tropical ocean and compares the autocorrelations and spatial structures of the simulations with global data.

Basic Theory for fGn Processes.

Text S1. Continuous-in-time fGn

An fGn process at resolution τ (the scale at which the series is averaged) has the following integral representation:

$$T_\tau(t) = \frac{1}{\tau}\frac{c_H \sigma_T}{\Gamma(H+3/2)}\left[\int_{-\infty}^{t}(t-t')^{H+1/2}\gamma(t')dt' - \int_{-\infty}^{t-\tau}(t-\tau-t')^{H+1/2}\gamma(t')dt'\right], \quad \text{(S1)}$$

where $\gamma(t)$ is a unit Gaussian δ-correlated white noise process with $\langle\gamma(t)\rangle = 0$ and $\langle\gamma(t)\gamma(t')\rangle = \delta(t-t')$ [δ is the Dirac function], Γ is the Euler gamma function, σ_T is the ensemble standard deviation (for $\tau = 1$) and

$${c_H}^2 = \frac{\pi}{2\cos(\pi H)\Gamma(-2-2H)}. \quad \text{(S2)}$$

This is the canonical value for the constant c_H that was chosen to make the expression for the statistics particularly simple. In particular, the variance is $\langle T_\tau(t)^2\rangle = {\sigma_T}^2\tau^{2H}$ for all t, where $\langle\cdot\rangle$ denotes ensemble averaging. The parameter H, with $-1 < H < 0$, is the fluctuation exponent of the corresponding fractional Gaussian noise process, the Hurst exponent, $H' = H + 1$. Fluctuation exponents are used due to their wider generality; they are well defined even for strongly intermittent non-Gaussian multifractal processes.

Equation (S1) can be interpreted as the smoothing by the fractional integral of a white noise process or as the power-law weighted average of past innovations, $\gamma(t)$. This power-law weighting accounts for the memory effects in the temperature series. The closer the fluctuation exponent is to zero, the larger is the influence of past values on the current temperature. This is evidenced by the behaviour of the autocorrelation function:

$$R_H(\Delta t) = \frac{\langle T_\tau(t)T_\tau(t+\Delta t)\rangle}{\langle T_\tau(t)^2\rangle} = \frac{1}{2}\left(\left|\frac{\Delta t}{\tau}+1\right|^{2H+2} + \left|\frac{\Delta t}{\tau}-1\right|^{2H+2} - 2\left|\frac{\Delta t}{\tau}\right|^{2H+2}\right), \quad \text{(S3)}$$

for $|\Delta t| \geq \tau$. In particular, for $\Delta t \gg \tau$ we obtain:

$$R_H(\Delta t) \approx (H+1)(2H+1)\left(\frac{\Delta t}{\tau}\right)^{2H}, \quad \text{(S4)}$$

which has the power-law behaviour mentioned earlier with the same exponent as the average squared fluctuation and due to the Wiener–Khinchin theorem, it implies a spectrum $E(\omega)\sim\omega^{-\beta}$ with exponent $\beta = 1 + 2H$. For more details on fGn processes see (Biagini et al., 2008; Gripenberg & Norros, 1996; Mandelbrot & Van Ness, 1968).

Text S2. Discrete-in-time fGn

A detailed explanation of the theory for modeling and predicting using the discrete version of fGn processes was presented in (Del Rio Amador & Lovejoy, 2019); the main results are summarized next. The analogue of Eq. (S1) in the discrete case for a finite series, $\{T_t\}_{t=1,\dots,N}$, with length N and zero mean is:

$$T_t = \sum_{j=1}^{t} m_{tj}\gamma_{t+1-j} = m_{t1}\gamma_t + \dots + m_{tt}\gamma_1, \tag{S5}$$

for $t = 1, \dots, N$, where $\{\gamma_t\}_{t=1,\dots,N}$ is a discrete white noise process and the coefficients m_{ij} are the elements of the lower triangular matrix $\mathbf{M}^N_{H,\sigma_T}$ given by the Cholesky decomposition of the autocovariance matrix, $\mathbf{C}^N_{H,\sigma_T} = {\sigma_T}^2[R_H(i-j)]_{i,j=1,\dots,N}$:

$$\mathbf{C}^N_{H,\sigma_T} = \mathbf{M}^N_{H,\sigma_T}\left(\mathbf{M}^N_{H,\sigma_T}\right)^T, \tag{S6}$$

with $m_{ij} = 0$ for $j > i$ (we assume $\tau = 1$ is the smallest scale in our system). The superscript T denotes transpose operation. In vector form, Eq. (S5) can be written as:

$$\mathbf{T}_N = \mathbf{M}^N_{H,\sigma_T}\boldsymbol{\gamma}_N. \tag{S7}$$

Equations (S5-S7) can be used to create synthetic samples of fGn with a given length N, autocorrelation function given by Eq. (S3) and set of parameters $\sigma_T > 0$ and $-1 < H < 0$ (the mean of the series is always assumed equal to zero). Conversely, given an actual temperature series with vector $\mathbf{T}_N = [T_1, \dots, T_N]^T$, we can estimate the parameters σ_T and H using the maximum likelihood method [details are given in Appendix A of (Del Rio Amador & Lovejoy, 2019)] and we can verify that it could be well approximated by an fGn model by inverting Eq. (S7) and obtaining the residual vector of innovations:

$$\boldsymbol{\gamma}_N = \left(\mathbf{M}^N_{H,\sigma_T}\right)^{-1}\mathbf{T}_N. \tag{S8}$$

If the model provides a good description of the data, the residual vector $\boldsymbol{\gamma}_N = [\gamma_1, \dots, \gamma_N]^T$ is a white noise, i.e. the elements should be $\mathrm{NID}(0,1)$ with autocorrelation function $\langle\gamma_i\gamma_j\rangle = \delta_{ij}$ (δ_{ij} is the Kronecker delta and $\mathrm{NID}(0,1)$ stands for Normally and Independently Distributed with mean 0 and variance 1). It is worth mentioning that a white noise process is a particular case of fGn with $H = -1/2$.

Text S3. Prediction

If $\{T_t\}_{t<0}$ is an fGn process, the optimal k-step predictor for T_k ($k > 0$), based on a finite number, m (memory), of past values, is given by:

$$\hat{T}_k = \sum_{j=-m}^{0} \phi_j(k)T_j = \phi_{-m}(k)T_{-m} + \dots + \phi_0(k)T_0, \tag{S9}$$

where the vector, $\boldsymbol{\phi}(k) = [\phi_{-m}(k), \dots, \phi_0(k)]^T$, satisfies the Yule-Walker equations:

$$\mathbf{R}_H\boldsymbol{\phi}(k) = \mathbf{r}_H(k), \tag{S10}$$

with the vector $\mathbf{r}_H(k) = [R_H(k-i)\,]_{i=-m,\ldots,0}^T = [R_H(m+k), \ldots, R_H(k)\,]^T$ and $\mathbf{R}_H = [R_H(i-j)]_{i,j=-t,\ldots,0}$ being the autocorrelation matrix (see Eq. (S3)) (Hirchoren & Arantes, 1998).

The root mean square error (RMSE) for the predictor at a future time k, using a memory of m values, is defined as:

$$\mathrm{RMSE}(k,m) = \sqrt{\left\langle \left[T_k - \hat{T}_k(m) \right]^2 \right\rangle}\,. \tag{S11}$$

The following analytical expression can be obtained:

$$\mathrm{RMSE}(k,m,\sigma_T,H) = \sigma_T \sqrt{1 - \mathbf{r}_H(k)^T \left(\mathbf{R}_H\right)^{-1} \mathbf{r}_H(k)}\,. \tag{S12}$$

For a given forecast horizon, k, the RMSE only depends on the parameters σ_T and H, and the memory used, m.

The theoretical mean square skill score (MSSS), is defined as:

$$\mathrm{MSSS}(k) = 1 - \frac{\left\langle \left[T(k) - \hat{T}(k) \right]^2 \right\rangle}{\left\langle T(k)^2 \right\rangle}\,. \tag{S13}$$

(the reference forecast is the mean of the series, assumed equal to zero here).

From the definition of the RMSE, Eq. (S11), we obtain the theoretical value:

$$\mathrm{MSSS}(k,m,H) = 1 - \frac{\mathrm{RMSE}(k,m,\sigma_T,H)^2}{{\sigma_T}^2}, \tag{S14}$$

or, replacing Eq. (S12):

$$\mathrm{MSSS}(k,m,H) = \mathbf{r}_H(k)^T \left(\mathbf{R}_H\right)^{-1} \mathbf{r}_H(k) = \boldsymbol{\phi}(k) \cdot \mathbf{r}_H(k)\,. \tag{S15}$$

For $H = -1/2$, the fGn process is a white noise process and $\mathrm{MSSS} = 0$. The skill increases with H and the process becomes perfectly predictable when $H \to 0$.

Text S4. Cross-correlation function

Let $T_i(t)$ and $T_j(t)$ be two fGn processes with zero mean and respective parameters σ_{Ti}, H_i and σ_{Tj}, H_j, which could represent, for example, the natural temperature variability at locations "i" and "j", respectively. The cross-covariance function:

$$C_{ij}(t_1,t_2) = \left\langle T_i(t_1) T_j(t_2) \right\rangle, \tag{S16}$$

can be found using the integral representation (Eq. (S1)) for each process:

$$C_{ij}(t_1,t_2) = \frac{1}{\tau^2} \frac{c_{Hi} c_{Hj} \sigma_{Ti} \sigma_{Tj}}{\Gamma\left[H_i + 3/2\right] \Gamma\left[H_j + 3/2\right]} \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} (t_1 - t')^{H_i + 1/2} (t_2 - t'')^{Hj + 1/2} \Big[\left\langle \gamma_i(t') \gamma_j(t'') \right\rangle + \\ + \left\langle \gamma_i(t' - \tau) \gamma_j(t'' - \tau) \right\rangle - \left\langle \gamma_i(t') \gamma_j(t'' - \tau) \right\rangle - \left\langle \gamma_i(t' - \tau) \gamma_j(t'') \right\rangle \Big] dt' dt'' \tag{S17}$$

where we have changed variables in each of the respective second integrals in Eq. (S1).

Let us assume that the white-noise innovations satisfy:

$$\left\langle \gamma_i(t')\gamma_j(t'')\right\rangle = a_{ij}\delta(t'-t''), \tag{S18}$$

where $-1 < a_{ij} < 1$. If we assume that $T_i(t)$ and $T_j(t)$ are jointly wide-sense stationary, then, without loss of generality, we can replace $t_1 = 0$, $t_2 = \Delta t \geq \tau$, in Eq. (S17) to obtain:

$$C_{ij}(\lambda) = c_{H_i}c_{H_j}\sigma_{T_i}\sigma_{T_j}a_{ij}\tau^{H_i+H_j}\frac{\cos(\pi H_j)\Gamma(-2-H_i-H_j)}{\pi}, \\ \left[(\lambda+1)^{2+H_i+H_j}+(\lambda-1)^{2+H_i+H_j}-2\lambda^{2+H_i+H_j}\right] \tag{S19}$$

where $H_i + H_j \neq -1$, $\lambda = \Delta t/\tau$ and c_{H_i} and c_{H_j} are given by Eq. (S2).

For the cross-correlation function:

$$R_{ij}(\Delta t) = \frac{\left\langle T_i(t)T_j(t+\Delta t)\right\rangle}{\sqrt{\left\langle T_i(t)^2\right\rangle\left\langle T_j(t+\Delta t)^2\right\rangle}} = \frac{C_{ij}(\Delta t)}{\sigma_{Ti}\tau^{H_i}\sigma_{Tj}\tau^{H_j}}, \tag{S20}$$

where we replaced Eq. (S16), $\langle T_i(t)^2\rangle = {\sigma_{Ti}}^2\tau^{2H_i}$ and $\langle T_j(t+\Delta t)^2\rangle = {\sigma_{Tj}}^2\tau^{2H_j}$, we obtain:

$$R_{ij}(\lambda) = \frac{a_{ij}}{2}\frac{\Gamma(-2-H_i-H_j)}{\sqrt{\Gamma(-2-2H_i)\Gamma(-2-2H_j)}}\sqrt{\frac{\cos(\pi H_j)}{\cos(\pi H_i)}}. \\ \left[(\lambda+1)^{2+H_i+H_j}+(\lambda-1)^{2+H_i+H_j}-2\lambda^{2+H_i+H_j}\right] \tag{S21}$$

The integral in Eq. (S17) can also be evaluated for $t_1 = t_2 = \Delta t = 0$. The final result for all cases is:

$$R_{ij}(a_{ij},H_i,H_j,\lambda) = \frac{a_{ij}}{2}\frac{\Gamma(-2-H_i-H_j)}{\sqrt{\Gamma(-2-2H_i)\Gamma(-2-2H_j)}}F_{ij}(\lambda,H_i,H_j) \\ \left[|\lambda+1|^{2+H_i+H_j}+|\lambda-1|^{2+H_i+H_j}-2|\lambda|^{2+H_i+H_j}\right] \tag{S22}$$

where $\lambda = (t_2 - t_1)/\tau$ is the lag of process "i" with respect to "j" in units of τ and

$$F_{ij}(\lambda,H_i,H_j) = \begin{cases} \sqrt{\dfrac{\cos(\pi H_j)}{\cos(\pi H_i)}}; & \text{if} \quad \lambda \geq 1 \\ \sqrt{\dfrac{\cos(\pi H_i)}{\cos(\pi H_j)}}; & \text{if} \quad \lambda \leq -1 \\ \dfrac{1}{2}\left[\sqrt{\dfrac{\cos(\pi H_i)}{\cos(\pi H_j)}}+\sqrt{\dfrac{\cos(\pi H_j)}{\cos(\pi H_i)}}\right]; & \text{if} \quad \lambda = 0 \end{cases}. \tag{S23}$$

This expression is equivalent to the one obtained by (Coeurjolly et al., 2010), except that they use a different normalization. For $\lambda \gg 1$, we obtain the asymptotic approximation:

$$R_{ij}\left(a_{ij}, H_i, H_j, \lambda\right) \approx \varphi_{H_i,H_j} a_{ij} \lambda^{H_i+H_j} . \tag{S24}$$

where

$$\varphi_{H_i,H_j} = \frac{\Gamma\left(-2-H_i-H_j\right)\left(2+H_i+H_j\right)\left(1+H_i+H_j\right)}{2\sqrt{\Gamma\left(-2-2H_i\right)\Gamma\left(-2-2H_j\right)}} \sqrt{\frac{\cos\left(\pi H_j\right)}{\cos\left(\pi H_i\right)}} . \tag{S25}$$

Note that the function $F_{ij}\left(\lambda, H_i, H_j\right)$ satisfies $F_{ij}(\lambda) = F_{ji}(-\lambda)$. The other factors in Eq. (S22) are symmetric with respect to a permutation of the indexes or to the sign of the lag. Hence, the cross-correlation function satisfies the symmetry property for jointly wide sense stationary-processes: $R_{ij}(\lambda) = R_{ji}(-\lambda)$. If $a_{ij} = 1$ and $H_i = H_j$, we recover the autocorrelation function Eq. (S3).

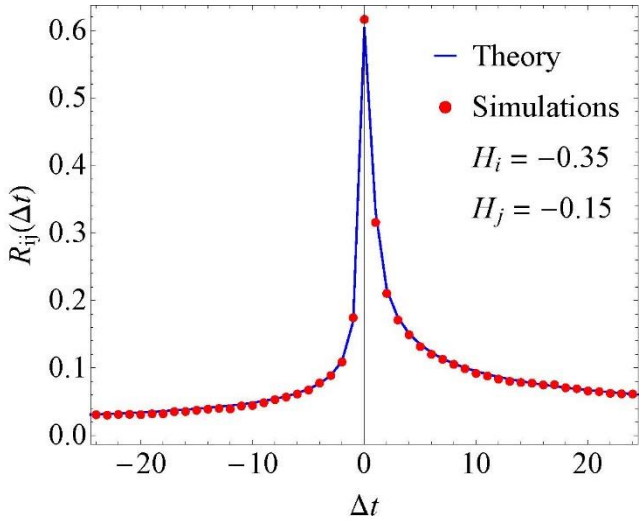

Figure S1. Cross-correlation function for $a_{ij} = 0.7$ and fluctuations exponents $H_i = -0.35$ and $H_j = -0.15$. The blue line is the graph of the theoretical expression Eq. (S22) and the red dots were obtained from a set of 500 pairs of fGn series each $N = 1000$ timesteps long.

Figure S1 shows an example of the cross-correlation function for $a_{ij} = 0.7$ and fluctuations exponents $H_i = -0.35$ and $H_j = -0.15$. The blue line is the graph of the theoretical expression Eq. (S22) and the red dots were obtained from a set of 500 pairs of fGn series each $N = 1000$ timesteps long. The simulations were produced using Eqs. (S5-S7) where the white noise series of innovations, $\{\gamma_i\}$ and $\{\gamma_j\}$, were generated from a multivariate Gaussian distribution in such a way that: $\langle\gamma_i(t')\gamma_i(t'')\rangle = \langle\gamma_j(t')\gamma_j(t'')\rangle = \delta_{t't''}$ and $\langle\gamma_i(t')\gamma_j(t'')\rangle = \langle\gamma_j(t')\gamma_i(t'')\rangle = a_{ij}\delta_{t't''}$ (Notice that we replaced the Dirac by the Kronecker δ for discrete-in-time series). The estimated cross-correlation function for the simulations (red dots) was computed as:

$$R_{ij}(\Delta t) = \frac{\sum_{t=1}^{N-\Delta t} T_i(t) T_j(t+\Delta t)}{\sqrt{\sum_{t=1}^{N} T_i(t)^2 \sum_{t=1}^{N} T_j(t)^2}} . \quad \text{(S26)}$$

Text S5. Empirical cross-correlation matrices

For the dataset described in Sect. 2.1, we estimated the natural temperature variability, $T_i(t)$, and, using the theory presented in Text S2, we obtained the series of innovations for each location, $\gamma_i(t)$. Using Eq. (S26), we estimated the lagged cross-correlation matrices (shown in Fig. S2) for the innovations and for the natural temperature variability: $[\rho_{ij}(\lambda)]$ and $[R_{ij}(\lambda)]$, respectively ($\lambda = \Delta t/\tau$ is the lag in units of $\tau = 1$ month). In panel (a) we show the full cross-correlation matrix for lag $\lambda = 0$ including the 10512 grid points (73 latitudes × 144 longitudes) for the innovations (left) and for the temperature (right). The pixels were indexed as: 1→{90°S, 0°E}, 2→{90°S, 2.5°E}, …, 10511→{90°N, 5°W}, 10512→{90°N, 2.5°W}. These 10512×10512-pixel images are too big to appreciate the detailed structure of the teleconnections. The large autocorrelation values are visible only along the main diagonal, as well as in the top-left and bottom-right corners corresponding to the poles where the grid points are very close to each other and where they share the same temperature values. Large correlations are also observed in the tropical region for the temperature anomalies.

To discern the details of the cross-correlations, we blew-up the regions shown as black squares in panel (a). In panel (b), we show the lagged cross-correlation matrices for the innovations for $\lambda = 0$, 1, 2 and 3 months (left to right), only for these 576 grid points between latitudes 42.5°N and 52.5°N. In panel (c), we show similar cross-correlation matrices as in (b), but now for the natural temperature variability. In the figure captions in panels (b) and (c), we show the values of the average cross-correlation $\pm$ one standard deviation and the maximum absolute value for each matrix (i.e. out of more than $3 \cdot 10^5$ values).

For $\lambda = 0$, the elements $\rho_{ij}(0) = a_{ij}$ shown in Fig. S2(b), are relatively large. The maximum values are evidently 1 along the main diagonal, but very large values are also obtained along the diagonals separated by 144 pixels because they represent places only 2.5° away in latitude. For the temperature (panel (c)), the correlations decrease with the lag, but some of the structure is preserved and relatively large values are obtained even for $\lambda = 3$ months, mainly along the diagonals. The temperature cross-correlation, $R_{ij}(\lambda)$ is proportional to a_{ij}, but it also depends on the fluctuation exponents H_i and H_j for every λ, following Eqs. (7) and (8). For the cross-correlation of the innovations (panel (b)), the values decrease much faster. Even for $\lambda = 1$, we can see that almost all the correlation is lost (see the distribution values in the figure captions). This indicates that the innovation series closely satisfy the time-independence condition given by Eq. (S18) (actually, its discrete version where we replace the Dirac by the Kronecker delta). The same analysis exemplified here for the small sample square region of 576×576 pixels,

was performed in the full 10512×10512 correlation matrices obtaining similar distributions for each respective lag.

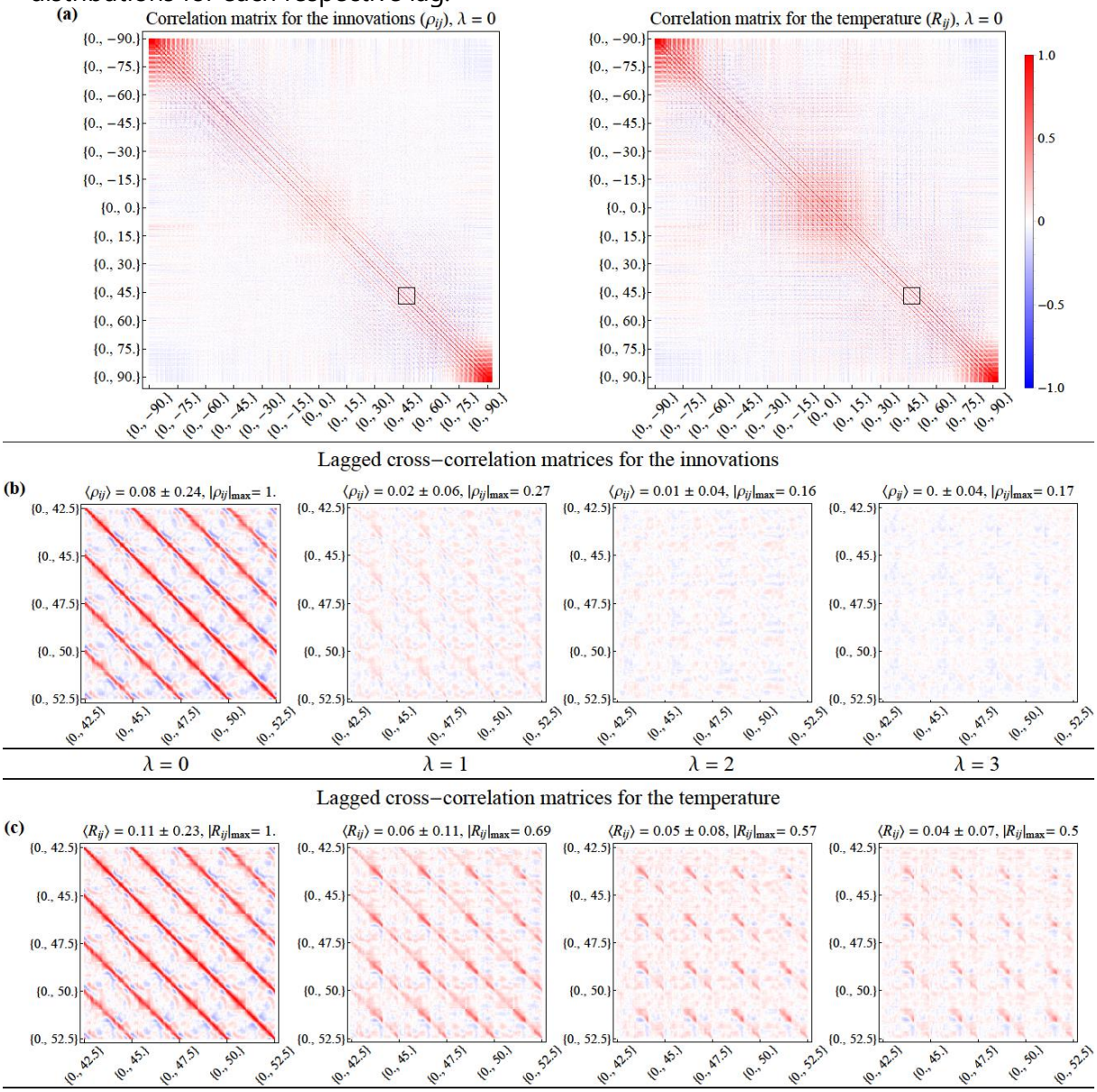

Figure S2. Lagged cross-correlation matrices for the innovations and for the natural temperature variability: $[\rho_{ij}(\lambda)]$ and $[R_{ij}(\lambda)]$, respectively. (a) Full cross-correlation matrices for lag $\lambda = 0$ including the 10512 grid points for the innovations (left) and for the temperature (right). (b) Lagged cross-correlation matrices for the innovations for $\lambda =$ 0, 1, 2 and 3 (left to right), only for the 576 grid points between latitudes 42.5°N and 52.5°N. They correspond to the small square region shown in the respective matrix in panel (a). (c) Same as in (b), but now for the natural temperature variability. In the figure captions in panels (a) and (b), we show the values of the average cross-correlation $\pm$ one standard deviation and the maximum absolute value for each matrix.

Text S6. Co-predictors

In Text S3 we mentioned that, given m datapoints, the *k*-step predictor for the temperature at location "*i*", with $T_i(t)$ an fGn process with parameters σ_i and H_i, is in vector form:

$$\hat{T}_i^0(k) = \boldsymbol{\phi}_i^0(k) \cdot \mathbf{T}_i, \tag{S27}$$

where $\mathbf{T}_i = [T_i(-m), \dots, T_i(0)]^T$ and the vector of coefficients $\boldsymbol{\phi}_i^0(k) = \left[\phi_{-m}^{i0}(k), \dots, \phi_0^{i0}(k)\right]^T$, satisfies the Yule-Walker equations (Eq. (S10)). The superscript "0" indicates that only the location "*i*" is considered.

Let us assume that we have another time series at location "*j*", $T_j(t)$ (fGn process with parameters σ_j and H_j) and we want to add this information to improve the predictor for the temperature at location "*i*". The optimal *k*-step predictor is now given by the sum of $2m + 2$ terms ($m + 1$ in the scalar product for each location):

$$\hat{T}_i(k) = \boldsymbol{\phi}_i(k) \cdot \mathbf{T}_i + \frac{\sigma_{Ti}}{\sigma_{Tj}} \boldsymbol{\phi}_j(k) \cdot \mathbf{T}_j, \tag{S28}$$

where the vectors of coefficients, $\boldsymbol{\phi_i}(k)$ and $\boldsymbol{\phi_j}(k)$ satisfy the Yule-Walker equations:

$$\begin{pmatrix} \mathbf{R}_{ii} & \mathbf{R}_{ij} \\ \mathbf{R}_{ji} & \mathbf{R}_{jj} \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi}_i(k) \\ \boldsymbol{\phi}_j(k) \end{pmatrix} = \begin{pmatrix} \mathbf{r}_{ii}(k) \\ \mathbf{r}_{ij}(k) \end{pmatrix}. \tag{S29}$$

The matrices $\mathbf{R}_{ii} = [R_{ii}(1, t_1 - t_2)]_{t_1,t_2=-m,\dots,0}$ and $\mathbf{R}_{jj} = \left[R_{jj}(1, t_1 - t_2)\right]_{t_1,t_2=-m,\dots,0}$ are the autocorrelation matrices for processes "*i*" and "*j*", respectively, $\mathbf{R}_{ij} = \mathbf{R}_{ji}{}^T = \left[R_{ij}(a_{ij}, t_1 - t_2)\right]_{t_1,t_2=-m,\dots,0}$ are the cross-correlation matrices and the vectors $\mathbf{r}_{ii}(k) = [R_{ii}(1, k - s)\,]_{s=-m,\dots,0}^T$ and $\mathbf{r}_{ij}(k) = \left[R_{ij}(a_{ij}, k - \mathrm{s})\,\right]_{s=-m,\dots,0}^T$, (the elements are obtained from Eq. (S22)).

For the case where we only have one time series at location "*i*", Eq. (S15) gives:

$$\mathrm{MSSS}_i^0(k, m, H_i) = \mathbf{r}_{ii}(k)^T \left(\mathbf{R}_{ii}\right)^{-1} \mathbf{r}_{ii}(k) = \boldsymbol{\phi}_i^0(k) \cdot \mathbf{r}_{ii}(k). \tag{S30}$$

The MSSS for the forecast at location "*i*", considering now the information from the two locations "*i*" and "*j*" is:

$$\mathrm{MSSS}_i(k, m, a_{ij}, H_i, H_j) = \boldsymbol{\phi}_i(k) \cdot \mathbf{r}_{ii}(k) + \boldsymbol{\phi}_j(k) \cdot \mathbf{r}_{ij}(k). \tag{S31}$$

The skill score for horizon k is a function of the memory, m, the intrinsic spatial correlation of the innovations, a_{ij}, (independent of the scaling) and the fluctuation exponents, H_i and H_j.

The main question is how much the new location helps to improve the accuracy of the forecast at position "*i*". This can be quantified by the difference $\Delta\mathrm{MSSS}_i = \mathrm{MSSS}_i - \mathrm{MSSS}_i^0$:

$$\Delta\mathrm{MSSS}_i\left(k,m,a_{ij},H_i,H_j\right)=\left[\boldsymbol{\phi}_i\left(k\right)-\boldsymbol{\phi}_i^0\left(k\right)\right]\cdot\mathbf{r}_{ii}\left(k\right)+\boldsymbol{\phi}_j\left(k\right)\cdot\mathbf{r}_{ij}\left(k\right) \tag{S32}$$

which can also be written as:

$$\Delta\mathrm{MSSS}_i\left(k,m,a_{ij},H_i,H_j\right)=\boldsymbol{\phi}_j\left(k\right)\cdot\left[\mathbf{r}_{ij}\left(k\right)-\mathbf{R}_{ij}{}^T\boldsymbol{\phi}_i^0\left(k\right)\right]. \tag{S33}$$

This is the normalized projection of the predictor from the new location,

$$\hat{T}_j\left(k\right)=\boldsymbol{\phi}_j\left(k\right)\cdot\mathbf{T}_j, \tag{S34}$$

in the direction of the error from the first predictor only:

$$\Delta\mathrm{MSSS}_i\left(k\right)=\left\langle\hat{T}_j\left(k\right)\left[T_i\left(k\right)-\hat{T}_i^0\left(k\right)\right]\right\rangle. \tag{S35}$$

This is in agreement with the orthogonality principle that states that the error of the predictor at location "i" is orthogonal to the data used to build that predictor. Location "j" can only contribute with new information that gives some component along this orthogonal direction.

For continuous-in-time infinitely long series (infinite memory), the predictor $\hat{T}_j(k)$ for $k>0$ is a linear combination of past innovations, $\{\gamma_j(t)\}_{t<0}$, while the error, $T_i(k)-\hat{T}_i^0(k)$, is a linear combination of future innovations $\{\gamma_i(t)\}_{t>0}$ (see the development in Sect. 2.5 of (Lovejoy et al., 2015)). That means that, if the condition Eq. (S18) is satisfied (future and past innovations are independent), then $\Delta\mathrm{MSSS}_i=0$ in Eq. (S35) and the new location does not help to improve the forecast. That does not mean that the two series at location "i" and "j" are independent; they are still correlated with the correlation function given by Eq. (S22). It is just that this correlation is already included in the information obtained from the past at location "i", which is enough for obtaining the optimal prediction for that location.

For discrete-in-time finite series, there is some improvement in the prediction from using a co-predictor, but this improvement decreases with the memory, m, and is very small if enough past data points are used to build the predictor. In Fig. S3 we show contour plots of the relative difference $\Delta\mathrm{MSSS}_i/\mathrm{MSSS}_i^0$ (in %) as a function of H_i and H_j for $k=1$, $m=50$ and values of $a_{ij}=0.6$, 0.7 0.8 and 0.9. Notice that for a wide range of values, when H_i and H_j are relatively close (dark blue region in the plots), the second location brings almost no new information to the forecasts. The relative gain increases with a_{ij}, but even for $a_{ij}=0.6$ (which is a fairly high correlation already) it remains lower than 1% for all values of H_i and H_j (see top-left panel of Fig. S3). That is why we did not include plots for $a_{ij}<0.6$. In fact, even for highly correlated locations at the level of innovations with $a_{ij}=0.9$ (bottom-right panel), the maximum relative improvement is lower than 4%. This maximum improvement is obtained when $H_i\approx-0.5$, for which the original MSSS_i^0 is very low, so the difference is actually at the noise level and is not statistically significant. This means that, in practice, for any set of fluctuation exponents H_i and H_j and values of a_{ij} as large as 0.9, we only gain less than 2% of the original MSSS by using a co-predictor.

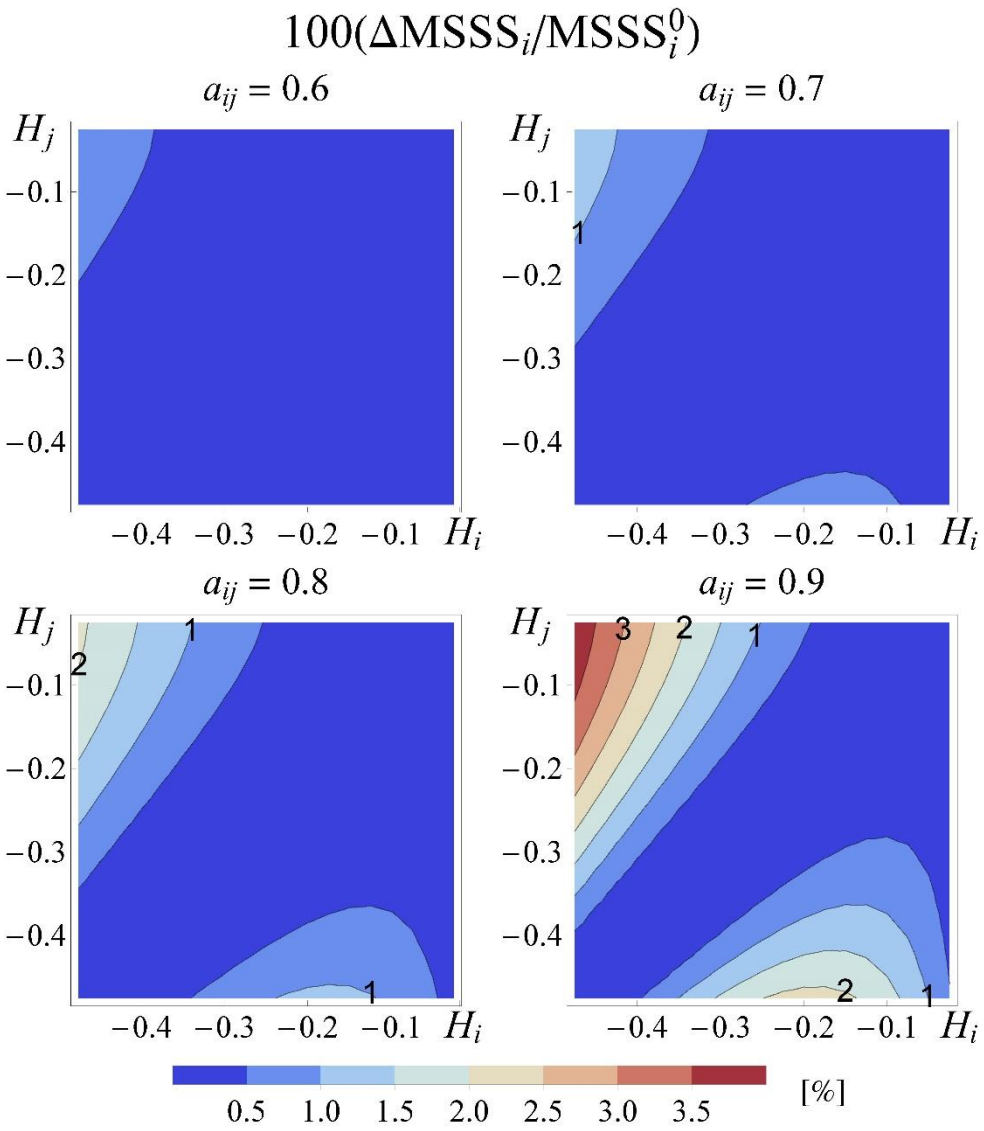

Figure S3. Contour plots of the relative improvement $\Delta\mathrm{MSSS}_i/\mathrm{MSSS}_i^0$ (in %) as a function of H_i and H_j for one step ($k = 1$), with 50 past values ($m = 50$) and correlations $a_{ij} = 0.6$, 0.7 0.8 and 0.9.

If MSSS_i^0 is the score obtained by predicting the temperature series at location i independently and MSSS_i is the score considering also the information from another location j, then the improvement $\Delta\mathrm{MSSS}_i = \mathrm{MSSS}_i - \mathrm{MSSS}_i^0$ is, by definition, a measure of the Granger causality between series i and j. It can be proven that $\Delta\mathrm{MSSS}_i$ is the projection of the predictor at time t from the new location, $\hat{T}_j(t)$, in the direction of the error from the first predictor only:

$$\Delta\mathrm{MSSS}_i(t) = \left\langle \hat{T}_j(t)\left[T_i(t) - \hat{T}_i^0(t)\right]\right\rangle. \tag{S36}$$

The orthogonality condition Eq. 12 implies that, for infinite series, if the new predictor is a linear combination of past data, then $\Delta\mathrm{MSSS}_i = 0$. If only a few memory

steps are used, then larger improvements are obtained by borrowing memory from co-predictors. Figure S4 shows the relative improvement $\Delta\text{MSSS}_i/\text{MSSS}_i^0$ (in %) as a function of H_i and H_j for $k = 1$, $a_{ij} = 0.8$ and values of $m = 20$, 5, 3 and 1. As the number of autoregressive steps used (memory) decreases, the larger the relative improvement becomes from using a co-predictor. In all cases, the long-memory predictor skill from a single location, MSSS_i^{∞} (with $m = 50$), is larger than the combined short memory one for $m = 5$, 3, 2 and 1 (see Fig. S5).

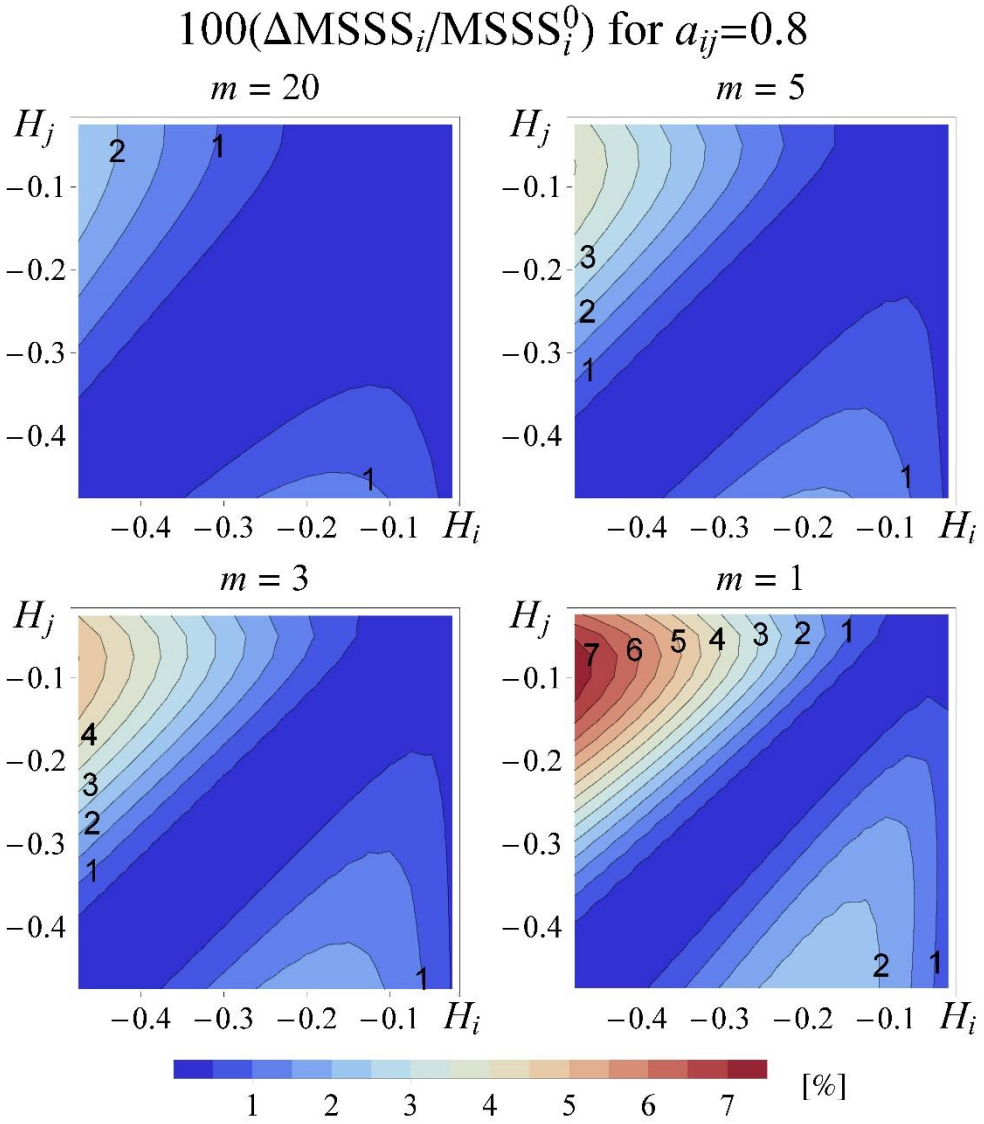

Figure S4. Contour plots of the relative improvement $\Delta\text{MSSS}_i/\text{MSSS}_i^0$ (in %) as a function of H_i and H_j for one step ($k = 1$), with $a_{ij} = 0.8$ and $m = 20$, 5, 3 and 1.

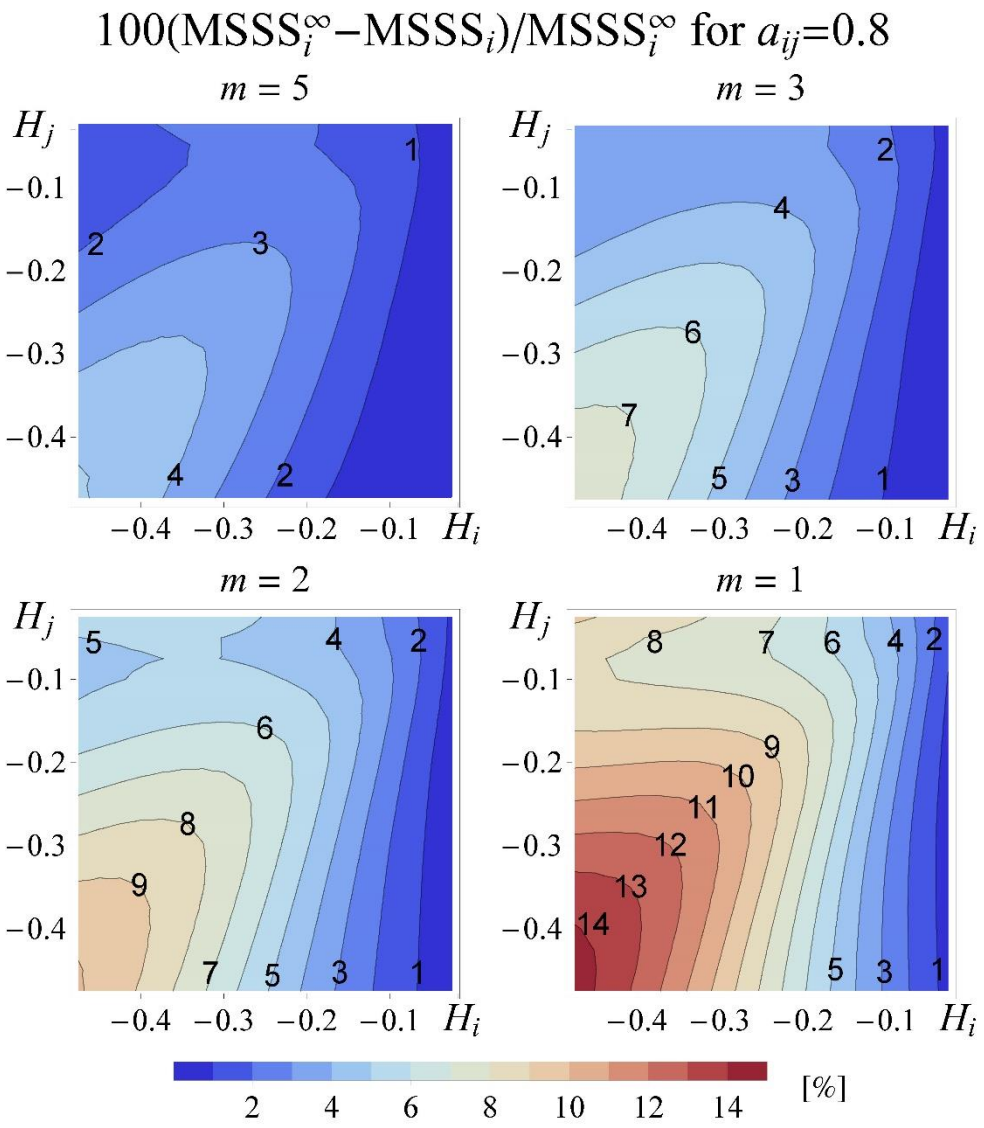

Figure S5. Contour plots of the relative difference $(\mathrm{MSSS}_i^\infty - \mathrm{MSSS}_i)/\mathrm{MSSS}_i^\infty$ as a function of H_i and H_j for one step ($k = 1$), with $a_{ij} = 0.8$. We used $m = 50$ for the single location long-memory MSSS_i^∞ and $m = 5$, 3, 2 and 1 for the combined skill MSSS_i.

Following Eq. (S33), we computed a map of the maximum relative improvement $\Delta\mathrm{MSSS}_i/\mathrm{MSSS}_i^0$ (Fig. S6) based on the empirical parameters. We see that the contribution from any other location is very small, reaching a maximum of 2% only in a few places, which is in the noise level of the skill estimates. This empirically confirms the lack of Granger causality between the series. As we show in Fig. S3, the maximum improvement at location i comes from a place j with large correlation, a_{ij}, and fluctuation exponent H_j very different than H_i. This gives the largest component of the co-predictor orthogonal to the m-dimensional space defined by the m past values of temperature at location i used to build the predictor. To produce the forecasts used in Fig. S6, we used $m = 20$. The relative contribution from co-predictors can be decreased by just increasing the

number m of past temperature values used to build the original predictor, in the limit $m \to \infty$, the relative improvement vanishes.

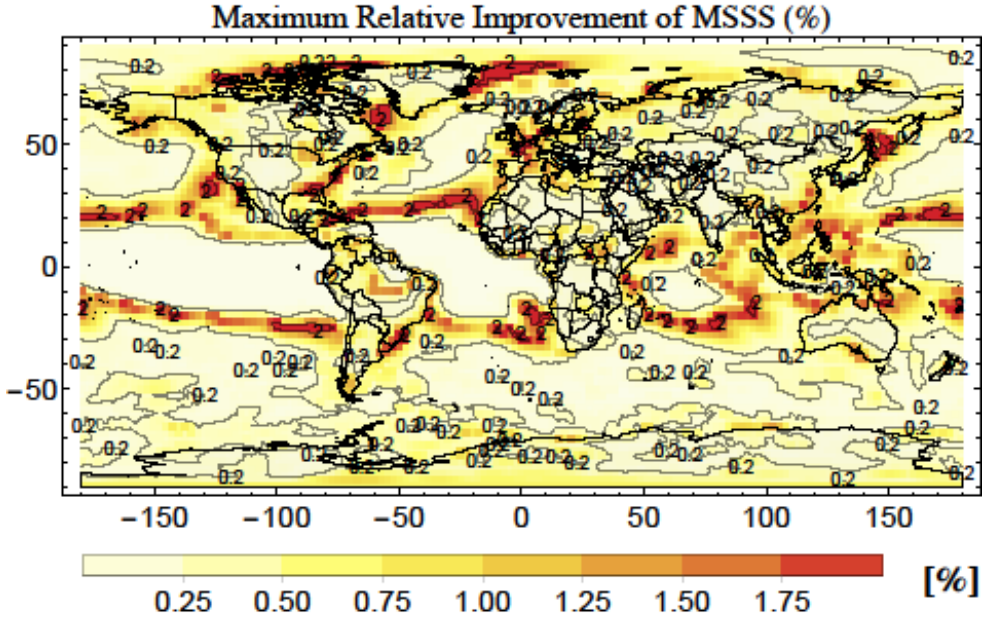

Figure S6. Maximum relative improvement, $\Delta\mathrm{MSSS}_i/\mathrm{MSSS}_i^0$, that could be obtained from using any other location as co-predictor, using $m = 20$ past values for the forecasts.

Text S7. Simulations

For most of the globe, the natural temperature variability has a transition from the weather (characterized by fluctuations increasing with the time scale) to the macroweather (with decreasing fluctuations) at a transition time τ_w lower than one month. This means that, for monthly averaged temperature, only the macroweather regime is present with the corresponding $H < 0$ (see Fig. 1(a)) and the temperature can be modeled and simulated using the theory presented in sections S1 and S2.

As presented in (Del Rio Amador & Lovejoy, 2020), for the places shown in red and yellow in Fig. 1(b) (generally over the tropical ocean), the weather-macroweather transition occurs at τ_w generally between 1 and 2 years, corresponding to the longer predictability limit of the ocean. For monthly averages, there is a biscaling behaviour: for time scales lower than τ_w, $H > 0$, the temperature can be modelled as a fractional Brownian motion (fBm) process (first differences are fGn) and for time scales larger than τ_w, $H < 0$, the anomalies behave as fGn. Lovejoy (2019, 2020; Lovejoy et al., 2020) shows how this biscaling process – called fractional relaxation noise (fRn) – emerges naturally as the solution to a fractional energy balance equation (FEBE), and τ_w can be identified as a characteristic relaxation time.

In this paper, we follow a more pragmatic approach to reproduce the two scaling regimes by still using the theory for fGn processes. This is achieved by expressing the natural temperature variability as a smoothed fGn process with a window τ_w:

$$T_{\tau_w}(t) = \int_{t-\tau_w}^{t} T_\tau(t')\,dt', \tag{S37}$$

where $T_\tau(t)$ is the resolution τ fGn defined by Eq. (S1), which is the τ increments of the corresponding fBm process, $B_{H'}(t)$, with Hurst exponent $H' = H_{\mathrm{fBm}} > 0$ (the fluctuation exponent for the fGn is $H_{\mathrm{fGn}} = H' - 1 < 0$ as $0 < H' < 1$):

$$T_\tau(t) = \frac{1}{\tau}\left[B_{H'}(t) - B_{H'}(t-\tau)\right]. \tag{S38}$$

For $\tau \ll \tau_w$ (or exactly in the discrete case where we make $\tau = 1$ and we replace the integral by the corresponding sum in Eq. (S37)), it can be shown that $T_{\tau_w}(t)$ is the τ_w increment of the same fBm process:

$$T_{\tau_w}(t) = B_{H'}(t) - B_{H'}(t-\tau_w). \tag{S39}$$

Therefore, it is stationary with variance $\langle T_{\tau_w}(t)^2\rangle = {\sigma_T}^2 {\tau_w}^{2H}$ and correlation function:

$$R_{H,\tau_w}(\Delta t) = \frac{1}{2}\left(\left|\frac{\Delta t}{\tau_w}+1\right|^{2H} + \left|\frac{\Delta t}{\tau_w}-1\right|^{2H} - 2\left|\frac{\Delta t}{\tau_w}\right|^{2H}\right), \tag{S40}$$

with H being the fluctuation exponent for the high frequencies (i.e. $0 < H < 1$). The only difference with Eq. (S3) is that now we do not have the restriction $|\Delta t| \geq \tau_w$, thus the new process has two scaling regimes. For $\Delta t \ll \tau_w$ it can be approximated by:

$$R_{H,\tau_w}(\Delta t) \approx 1 - \left(\frac{\Delta t}{\tau_w}\right)^{2H}, \tag{S41}$$

while for $\Delta t \gg \tau_w$ it follows a power-law equivalent to Eq. (S4):

$$R_{H,\tau_w}(\Delta t) \approx H(2H-1)\left(\frac{\Delta t}{\tau_w}\right)^{2H-2}. \tag{S42}$$

Similar asymptotic behaviours are obtained for the high and the low frequency approximations of the fRn process, solution of the FEBE. Also, notice how the cross-correlations (Eqs. (7) and (8)) satisfy similar equations if we take $2H = H_i + H_j$.

Figure S7 shows the dependence of the autocorrelation function with the lag $\lambda = \Delta t/\tau_w$ (Eq. (S40)) for different values of H. The high and the low frequency approximations for $H = 0.25$, given by Eqs. (S41) and (S42), were included as dashed and dotted lines, respectively.

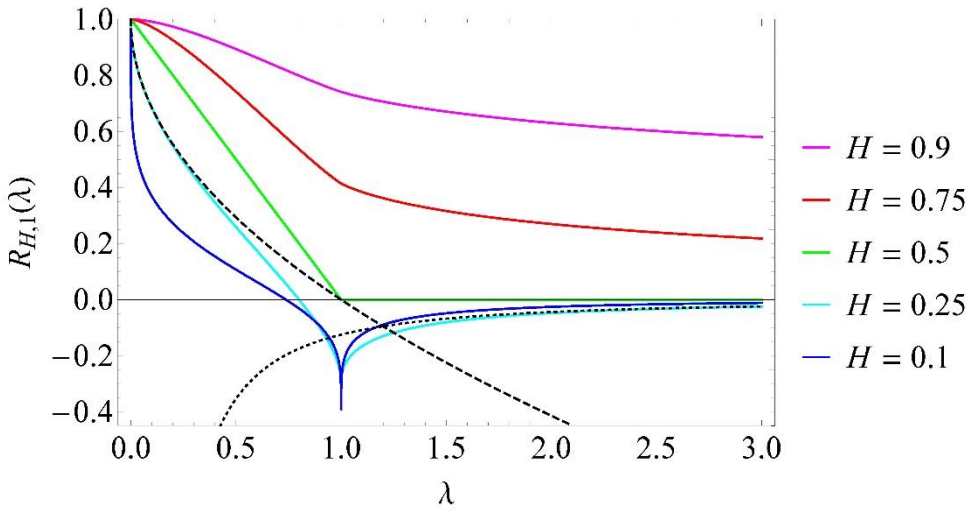

Figure S7. Autocorrelation function vs. $\lambda = \Delta t/\tau_w$ (Eq. (S40)) for different values of H. The high and the low frequency approximations for $H = 0.25$, given by Eqs. (S41) and (S42), were included as dashed and dotted lines, respectively.

To estimate the transition time, τ_w, in the places where $\tau_w > 1$ month ($H > 0$ in Fig. 1(a)), we can fit the theoretical autocorrelation function (ACF) (Eq. (S40)) to the empirical one obtained from the data. Figure S8 shows examples of the ACF for both cases: pure fGn with $H < 0$ and biscaling process with transition at τ_w from the weather, $H > 0$ (fBm-like) to the macroweather, $H < 0$ (fGn-like). The ACF's for the reference dataset (marked as "+"), for one set of simulations (marked as "o") and calculated using Eqs. (S3) or (S40) with the corresponding values of H and τ_w (solid curve) are shown in Fig. S8 for: (a) grid point in the North Atlantic at (55°N, 22.5°W), with estimated $H = -0.2$; (b) same as in (a) but with logarithmic scales in both axes to highlight the scaling; (c) grid point in the Tropical Pacific at (5°S, 177.5°W), with estimated high frequency $H = 0.38$ and $\tau_w = 16$ months and (d) the monthly mean temperature for the Niño 3.4 region (5°N-5°S, 170°W-120°W), with average fluctuation exponent $\langle H \rangle = 0.54$ and $\langle \tau_w \rangle = 10$ months. There is good agreement between the empirical, the simulated and the theoretical ACF's in all cases.

The empirical reproduction of the ACF validates the realism of the stochastic model on a pixel-by-pixel basis, but we must also reproduce the coupling between different pixels, i.e.: the cross-correlation structure. This is verified by comparing the empirical orthogonal functions (EOF's) obtained from the decomposition of the cross-correlation matrix. In Fig. S9 we show the first five EOF's for the reference dataset (left) and for a single realization simulations (right). To avoid strong multiplicative seasonality effects in the higher latitudes, we only considered the anomalies between 60°S and 60°N. In Fig. S10 we show the EOF's using only sea surface temperatures (SST's). The stochastic simulations reproduce very well the main modes of variability of the reference dataset. In the main test, we show how the cross-correlation structure for non-zero lags is also well reproduced by comparing the empirical and simulated ratio of global influence (RGI) – equivalent to the area weighted connectivity (AWC) for zero lags.

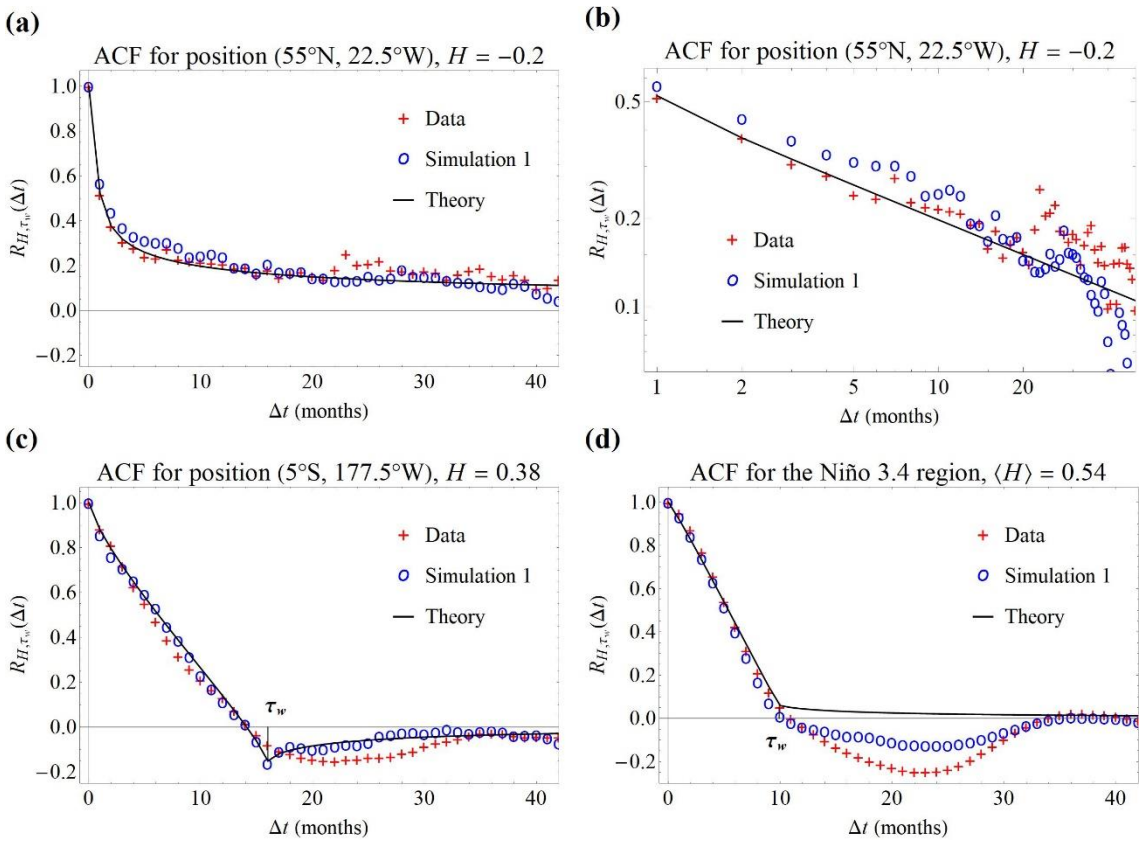

Figure S8. Autocorrelation functions (ACF's) for the reference dataset (marked as "+"), for one set of simulations (marked as "o") and obtained theoretically using Eqs. (S3) or (S40) (solid curve) for the temperature series at different positions: (a) grid point in the North Atlantic at (55°N, 22.5°W), with estimated $H = -0.2$; (b) same as in (a) but with logarithmic scale in both axes to highlight the scaling; (c) grid point in the Tropical Pacific at (5°S, 177.5°W), with $H = 0.38$ and $\tau_w = 16$ months and (d) the monthly mean temperature for the Niño 3.4 region (5°N-5°S, 170°W-120°W), with average $\langle H \rangle = 0.54$ and $\langle \tau_w \rangle = 10$ months.

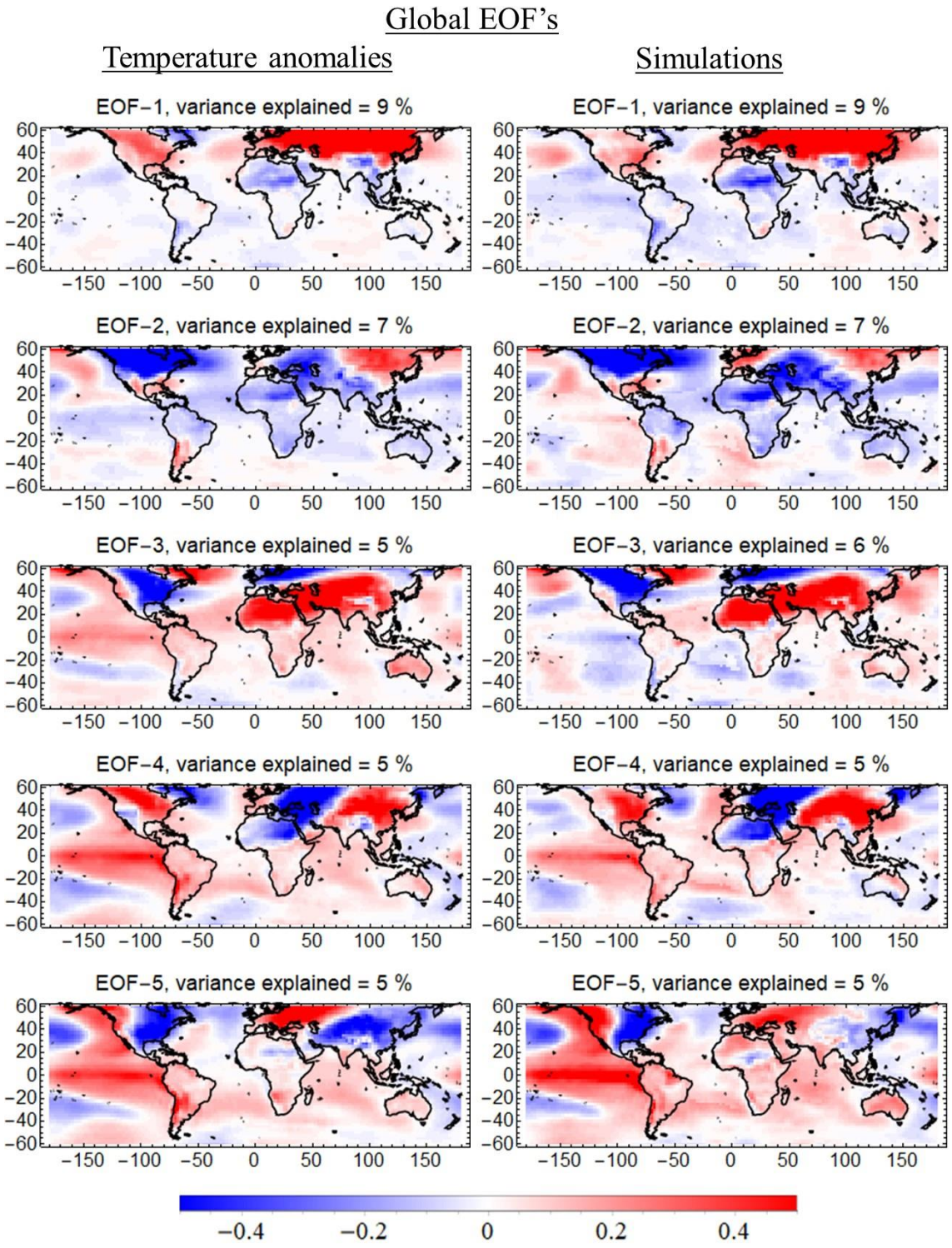

Figure S9. First five EOF's for the reference dataset (left) and for one simulation (right). We only considered the anomalies between 60°S and 60°N to avoid strong multiplicative seasonality effects in the higher latitudes.

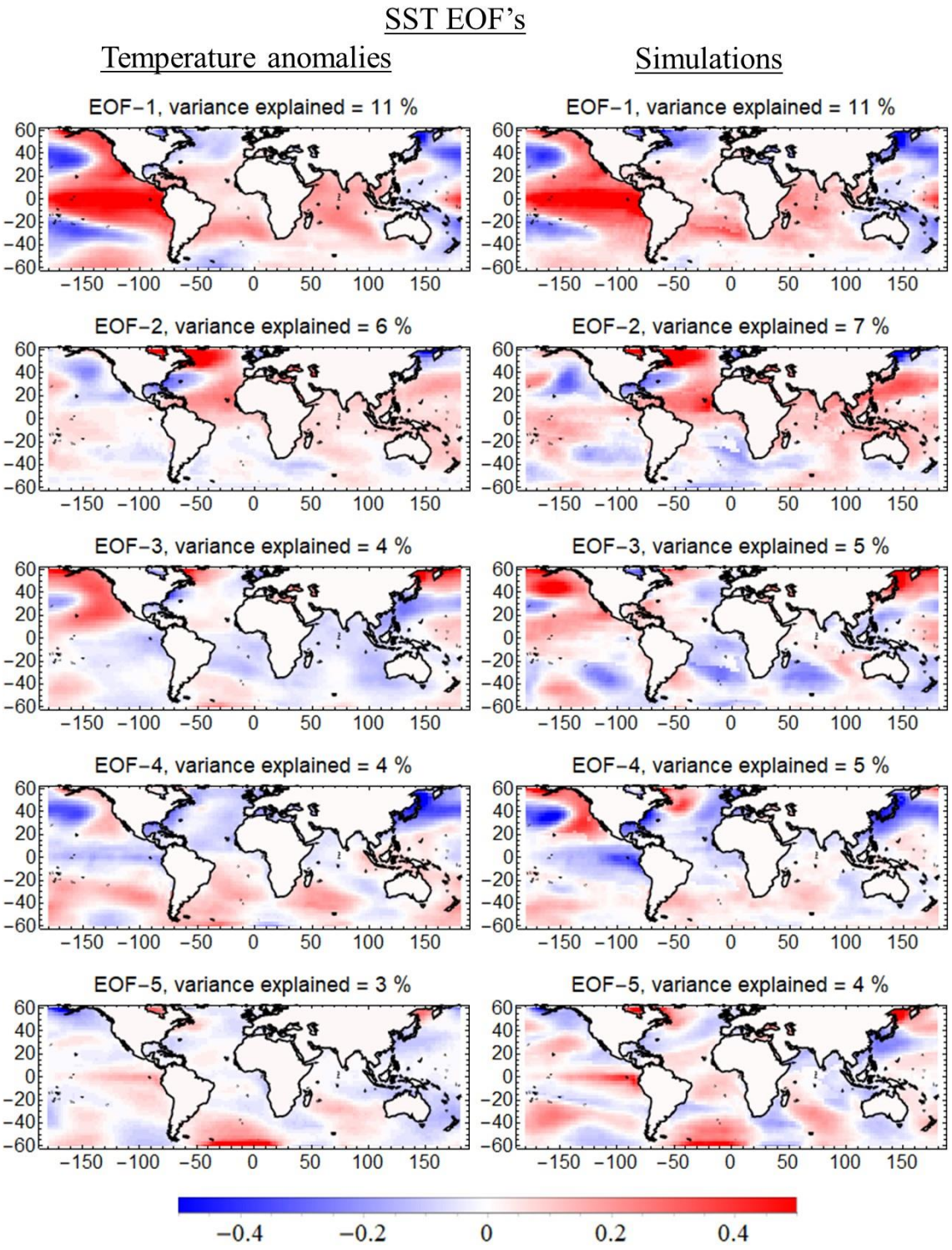

Figure S10. First five EOF's for the reference dataset (left) and for one simulation (right) only considering sea surface temperature anomalies (SST) between 60°S and 60°N.

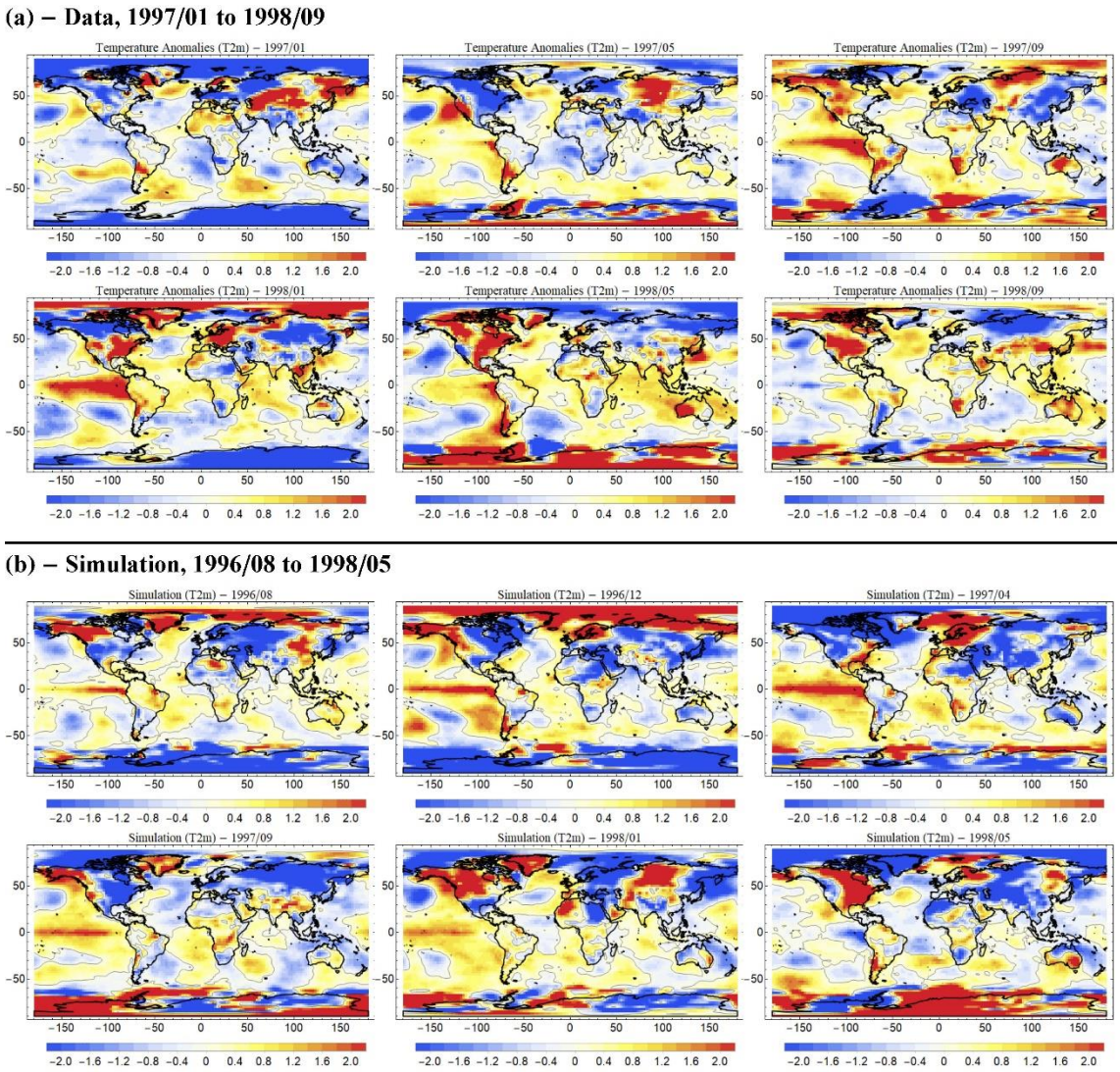

Figure S11. Sequence showing the evolution of one El Niño event (see the central and east-central equatorial Pacific) since January 1997 to September 1998 for the reference dataset (a) and from August 1996 to May 1998 for the simulations (b).

Additional Supporting Information (Files uploaded separately)

The evolution of the detrended temperature anomalies obtained from NCEP/NCAR Reanalysis 1 and one simulation for the period January 1948 – December 2019 is shown in Movie S1. It is important to realize that the simulation (right) is not supposed to be the same as the reality (left), it is only supposed to have the same type of variability (the date above the simulation is totally fictional). Among other realistic features, the simulation reproduces huge regional patterns including El Niño and La Niña events that are emergent properties of the model. One example of the evolution of the El Niño event is

shown in Fig. S11 with a sequence since January 1997 to September 1998 for the reference dataset (a) and from August 1996 to May 1998 for the simulations (b).

Movie S1. Evolution of the detrended temperature anomalies obtained from NCEP/NCAR Reanalysis 1 and one simulation for the period January 1948 – December 2019. It is important to realize that the simulation (right) is not supposed to be the same as the reality (left), it is only supposed to have the same type of variability (the date above the simulation is totally fictional). Among other realistic features, the simulation reproduces huge regional patterns including El Niño and La Niña events that are emergent properties of the model.