The joint space-time statistics of macroweather precipitation, space-time factorization and a space-time macroweather model

S. Lovejoy,¹,a) and M. I. P. de Lima²,³

¹Physics Department, McGill University, Montreal, Que., H3A 2T8, Canada
²Institute of Marine Research (IMAR) and Marine and Environmental Sciences Centre (MARE), Portugal
³Department of Civil Engineering, University of Coimbra, 3030-788 Coimbra, Portugal

Abstract

In addition to the familiar weather and climate regimes, over the range of time scales from about 10 days to 30–100 years, there is an intermediate “macroweather” regime that has unique scaling properties characterized by negative temporal fluctuation exponents; thus, contrary to the weather and climate regimes, in the macroweather regime fluctuations tend to cancel each other out, averages tend to converge. The macroweather regime thus covers seasonal, annual and decadal periods. We focus on the macroweather space-time variability, empirically and theoretically.

First, we systematically study the temporal and spatial scaling properties of three centennial, global scale precipitation products: one instrument based, one reanalysis based, one satellite and gauge based, we use Haar fluctuations, spectra and trace moment analysis. By jointly analyzing space-time fluctuations we obtain a complete space-time statistical description. This empirical description is used to verify the predictions of a stochastic weather-climate model prediction that the joint functions factor into separate spatial and temporal terms with the part being strongly variable (multifractal) due to very long time period (decades, century scale) climate process, but with the temporal part being nearly monofractal. We make an explicit stochastic model in which at each point on the earth, the temporal behavior is a fractional Gaussian noise but in the spatial correlations are due to a spatial multifractal climate cascade (the spatially homogeneous SLIM model; SLIM is the recently introduced Scaling LInear Macroweather model). This model explicitly has a property long assumed implicitly by climatologists, it can be “homogenized” by normalizing it by the standard deviation of the anomalies. Physically, it means that the spatial
macroweather variability corresponds to different climate zones that multiplicatively modulate the local, temporal statistics. This explicit model provides a framework for macroweather models that can make forecasts exploiting the long range memory as well as the spatial correlations.

---

*I)* Corresponding Author. Electronic mail: lovejoy@physics.mcgill.ca.

**Lead paragraph**

Scaling analyses of precipitation and other atmospheric fields have shown the existence of an intermediate regime between the familiar weather and climate regimes: “macroweather”, over the range of time scales from about 10 days to decadal, centennial scales. Although macroweather is important for seasonal, annual and decadal forecasts, and there have been no studies of its spatial variability and few studies of its temporal variability, with no coherent picture emerging. A recent paper (de Lima and Lovejoy 2015) makes a step in this direction by systematically studying the separate temporal and spatial variabilities in three centennial, global scale precipitation products: one instrument based, one reanalysis based, one satellite and gauge based. In this paper, we build on this work to analyze the joint space-time fluctuations using spectra as well as Haar structure functions allowing us to verify the prediction that the joint functions factor into separate spatial and temporal terms. We make an explicit space-time stochastic model with the observed statistics. Physically, factorization means that that the spatial macroweather variability corresponds to different climate zones that multiplicatively modulate the local, temporal statistics. The findings provide a framework for macroweather models that can make forecasts exploiting the long range memory as well as the spatial correlations.

**I. INTRODUCTION**

Ever since at least [Van der Hoven, 1957] it has been recognized that the atmosphere undergoes a drastic transition in its statistical properties at time scales of the order of 2-10 days. At first, this was theorized as “migratory pressure systems of synoptic weather map scale...” and termed the ”synoptic maximum” by [Kolesnikov and Monin, 1965] and
[Panofsky, 1969]); more recently, [Vallis, 2010] attributed it to baroclinic instabilities. However, following [Lovejoy and Schertzer, 1986], it was alternatively theorized as a transition scale $\tau_w$ between two scaling regimes at a scale corresponding to the lifetime of planetary sized structures. This interpretation was bolstered by the demonstration [Lovejoy and Schertzer, 2010] that the scale can be theoretically estimated from first principles from knowledge of the solar output and the efficiency of conversion of solar to mechanical energy ($\approx 4\%$). Further evidence in favor of the theory was the demonstration in [Lovejoy and Schertzer, 2013] that the spectrum of the ocean could be analogously explained (with a transition at $\approx 1$ year) and the theory has now also successfully explained the structure of the Martian atmosphere (with a transition at about 1.8 days; [Lovejoy et al., 2014]).

If we consider the temporal fluctuations in an atmospheric variable $I$ over an interval $\Delta t$ as $\Delta I(\Delta t)$, then in a scaling regime the mean fluctuations is $<\Delta I(\Delta t)>\approx \Delta t^H$, where $H$ is the fluctuation exponent (for Gaussian processes, it is equal to the Hurst exponent). [Lovejoy and Schertzer, 2013] showed for rain as well as the thermodynamic and dynamical variables, that the transition was between an $H>0$ high frequency weather regime and an $H<0$ low frequency “macroweather” regime. In particular, for rain, the transition scale $\tau_w$ varied somewhat with latitude from about 2–5 days, see Fig. 1a (a little less than for the temperature, 5–10 days). The different behaviors ($H>0$, $H<0$) correspond to average fluctuations growing or decreasing with scale; in the macroweather regime, since $H<0$, they tend to cancel so that averages over longer and longer times converge. However – at least for temperature, but probably also for the precipitation and other atmospheric fields – rather than converging to a fixed “climate” as one might expect, after about 10-30 years
(industrial period) and \(\approx100\) years (pre-industrial period), there is a transition to another scaling regime - the climate proper - again with with fluctuations growing with scale \(H>0\); see the rise in Fig. 1a beyond \(\approx30\) years). In the pre-industrial period over the last millennia, the situation is not so clear since the macroweather-climate transition scale \(\tau_c\) apparently has a great deal of spatial variability, see Lovejoy and Schertzer 2013, section 11.1.3.

**FIG. 1a.** The first order structure function (the mean absolute Haar fluctuation) using precipitation data from the Climate Prediction Center (CPC, continental US) gauges (dots) as well as the corresponding structure function of the 20CR reanalysis at 45°N (6 hours, 2° resolution, from 1871-2008, thick green line). \(\Delta R\) is the mean absolute fluctuation in the rain rate over a time interval \(\Delta t\). For the CPC product we also show the corresponding grid point to grid point one-standard-deviation limits (thin) with reference lines slopes \(H=-0.42\) (solid) and \(-0.5\) (dashed, corresponding to a Gaussian white noise process). NOAA's CPC product is unique in its high temporal resolution over a large number of contiguous grid points. The product analyzed was a (near complete) subset of the CPC data for the 29 years 1948-1976 (at this date there is a data gap of several weeks so that we did not extend the analysis to more recent times). The CPC data were
gridded on 2.5°×2.0° boxes by using a modified Cressman Scheme (an interpolation technique); we used its central rectangular 13 × 21 point region from: -122.5° to -72.5° longitude (every 2.5°≈ 210 km at these latitudes), and from 30° to 54° latitude (every 2°≈220 km). Each grid box had a near complete ≈257,000 hourly series. Adapted from [Lovejoy et al., 2012].

![Graph](image)

FIG. 1b. The ratio of the mean $q = 1$ and RMS fluctuations for the CPC data set. Reference lines have slopes $K(2)/2 ≈ C_1$ and show the transition from high intermittency behavior at scales less than a few days (the “weather regime”) to low but not insignificant intermittency behavior at scales of months to years (the “macroweather regime”). Gaussian white noise would be flat ($K(2) = 0$). Reproduced from [Lovejoy et al., 2012].

Although the exact value of the preindustrial $\tau_c$ may still be uncertain, for the temperature field the basic three scaling regime picture is relatively robust notably because a) of the existence of high quality paleotemperature data that allow us to estimate the statistics at decadal and centennial scales for preindustrial epochs, b) the temperatures are not too intermittent so that the statistics are more robust, c) the theoretical and GCM modeling implications of anthropogenic effects on the temperature are much clearer than for precipitation, d) (deterministic) numerical models well reproduce the weather and
macroweather regimes including exponents, as do (stochastic) turbulence based cascade models (see Sec. III).

Although, for precipitation, the basic picture seems to be the same, there is more uncertainty. For example, in the weather regime, the existence of transitions from zero to finite rain rates breaks the scaling ([Lovejoy et al., 2008]; [de Montera et al., 2009]; Mandapaka et al., 2010; Sun and Barros, 2010; [Verrier et al., 2010]; [Verrier et al., 2011]; [Hoang et al., 2012]; [Gires et al., 2013]) but the distinction between real and spurious breaks due to instrumental problems at low rain rates is still under debate (see the review in [Lovejoy and Schertzer, 2013], section 4.4). In contrast, in the macroweather regime there are numerous papers showing scaling analyses with data spanning a range of weather and macroweather scales, but only a few that explicitly attempted to distinguish the two regimes and to estimate macroweather exponents (i.e. from several days to years or decades). Examples of the former include [Pathirana et al., 2003]; [Bunde et al., 2005; Douglas and Barros, 2003]; [Garcia-Marin et al., 2008]; [de Lima and de Lima, 2009]; [Bunde et al., 2013]; [Rysman et al., 2013]; while examples of the latter are [Ladoy et al., 1991]; [De Lima, 1998]; [Tessier et al., 1996]; [Kantelhardt et al., 2006], [Lovejoy et al., 2012]. The exponents from these studies are summarized in table 1 of [de Lima and Lovejoy, 2015] (hereafter dLL) (see also table 10.1 in [Lovejoy and Schertzer, 2013], they generally concur with the results discussed below – i.e. low intermittency and $H \approx 0.4$, table I). It could be noted that some of the papers cited in the former category did quote exponents in the macroweather regime, they suffered from technical issues that led to large inaccuracies: two types of relevant technical problems are discussed in section III.B.
While the temporal statistics still need clarification, as far as we can tell, there have been no studies at all of the spatial macroweather statistics. This is partly due to the diversity of analysis techniques used and partly due to the strong focus on scaling statistics from single stations. We need clarification of a) the (possible) variation of the exponents with latitude, b) the variation over land, over ocean, c) the global scale averaged values, d) the expected anthropogenic (low frequency) effects, e) the degree of agreement/disagreement between different techniques for estimating areal precipitation, f) the joint space-time macroweather statistics. Issues a–e were addressed in dLL; while the joint space-time statistics needed for macroweather modeling with the related theoretical issues are the focus in this paper.

Empirical investigations cannot be divorced from theoretical frameworks and macroweather precipitation is no exception. [Lovejoy and Schertzer, 2010] and [Lovejoy and Schertzer, 2013] showed that space-time turbulent cascade models (the Fractionally Integrated Flux, FIF model) that were developed for weather scales could be extended to the macroweather regime by the simple expedient of allowing the cascades to develop starting from an (outer) time scale much longer than \( \tau_w \) (the Extended FIF, the EFIF model). The argument – summarized in more detail in Sec. II – leads to the conclusion that space-time macroweather statistics should, at least approximately, satisfy a fundamental space-time statistical factorization property. Applied to the spectral density \( P_{xyt}(k_x,k_y,\omega) \) in horizontal space-time \((x,y,t)\) \((k_x, k_y, \omega) \) are the corresponding wavenumbers and frequency), it implies that \( P_{xyt}(k_x,k_y,\omega) = P_{xy}(k_x,k_y)P_t(\omega) \), where \( P_{xy}(k_x,k_y) \) and \( P_t(\omega) \) are the horizontal and temporal spectral densities respectively. This contrasts with the space-time weather statistics that have spectra involving (turbulent) space-time scale functions such as powers
of \( (k_x^2 + k_y^2 + \omega^2)^{1/2} \) raised to various powers and that therefore cannot be factored in this way (see Sec.III C and [Pinel et al., 2014], [Pinel and Lovejoy, 2014], for generalizations including the extensions to waves). Factorization means that different (spatially distributed) climate zones modulate the local temporal statistics without changing their type (e.g. their temporal scaling). The factorization principle is already implicitly used in practical climatology when for example local station statistics are nondimensionalized by local standard deviations or by using (nondimensional) probability distributions so as to “homogenize” the data of to produce various climate indices that may be compared between different stations with different climates.

II. DATA AND FLUCTUATIONS

A. The data

We are interested in the space-time structure of precipitation over time scales from about 1 month to centuries (and longer if possible) and, in space, from global scales to scales of a few degrees (or smaller if possible). The main relevant gauge based data set is the Global Historical Climatology Network product (GHCN, [Lawrimore et al., 2011]) available from the NOAA site, which is monthly data for the period 1900–2012 at 5°×5° resolution. In order to alleviate issues to do with missing data, series consistency etc., only the precipitation anomalies were reported (i.e. with the annual cycle removed and relative to the 1961-1990 reference period). The data are gauge based, they are therefore restricted to land (but virtually all the pixels have significant outages). By excluding the oceans, the GHCN product will likely give a biased view of global scale precipitation. This is true not only because the oceans comprise 70% of the earth’s surface, but also because
oceanic precipitation is likely to be different from precipitation over land, and this includes a potentially much stronger response to anthropogenic warming. The only two relevant ocean precipitation data sets of which we are aware are from the Twentieth Century Reanalysis (20CR, [Compo et al., 2011]) and the [Smith et al., 2008], [Smith et al., 2012] satellite/gauge reconstruction (hereafter abbreviated “Smith”). Both of these data sets are very indirect, for example, the 20CR data (which is available at 2°×2° and 6 hour resolution, much higher than needed here) are derived solely from surface pressure data and monthly Sea Surface Temperature (SST) data, the precipitation is entirely inferred from a numerical model; the result – a “reanalysis” – is a kind of data/model hybrid (the product used here is monthly at 1.875° resolution, see e.g. [Kalnay, 2003] for the data assimilation techniques used in reanalyses). In comparison, the Smith data uses a gauge calibrated Infra Red satellite rain algorithm to infer global scale rain over the satellite observation period (1979-2012). This is then used to calculate Empirical Orthogonal Functions (EOF’s). Finally, in the pre-satellite era, the historic land based gauge data (GHCN) is used to estimate the coefficients of each EOF, yielding global scale estimates at 5°×5°, monthly resolution. dLL discusses the differences in detail as functions of space and time scale. For the means of the absolute rates for the Smith and 20CR data (the GHCN gives only anomalies); they find that there is a disagreement of ≈20% for the land estimates, but only about 5% for the ocean estimates: overall the disagreement is about a 10% for the global values.
B. Quantifying the variability over scales: fluctuations, structure functions

Consider the global scale averages, the anomalies are shown in Fig. 2. Notice that the gauge based product (GHCN) is much more variable than the Smith product, itself more variable than the 20CR product. We can also note that while there is some overall agreement at the lowest frequencies, the higher frequencies are often in disagreement. In order to quantify the high and low frequency variability, we can consider the mean and the root mean square (RMS) fluctuations.

FIG. 2a. The annual precipitation rate anomalies averaged over land only (bottom), ocean (middle) and globally (top). The black (thick) curve is the Global Historical Precipitation network (GHCN) from Jan. 1900 to Dec. 2012. The red (dashed) curves were the 20CR from 1880-2004 and the green (thin) curves were for the Smith product. The GHCN and Smith products were at 5°×5° resolution, the 20CR data were at 1.875° resolution. The data were shifted upwards by 2 mm/month increments as indicated by the dashed horizontal lines.
While many fluctuation definitions are possible, in this paper we use Haar fluctuations.
The Haar fluctuation of the precipitation \( \Delta R(\Delta t) \) at time scale \( \Delta t \) is simply the difference of the mean of \( R \) over the first and second halves of the interval \( \Delta t \):

\[
(\Delta R(\Delta t))_{\text{Haar}} = \frac{2}{\Delta t} \int_{\frac{t-\Delta t/2}{2}}^{t} R(t') dt' - \frac{2}{\Delta t} \int_{\frac{t+\Delta t/2}{2}}^{t} R(t') dt'
\]

where we have added the subscript “Haar” to distinguish it from other common definitions of fluctuation and we have suppressed the \( t \) dependence because we will assume that the fluctuations are statistically stationary. These are simple to understand because with an appropriate “calibration” constant (a factor 2 used throughout this paper), in scale regions where \( H>0 \), the Haar fluctuations are nearly equal to the differences, in scale regions where \( H<0 \), they are nearly equal to the anomalies:

\[
(\Delta R(\Delta t))_{\text{diff}} = R(t+\Delta t) - R(t) \\
(\Delta R(\Delta t))_{\text{anom}} = \frac{1}{\Delta t} \int_{\frac{t-\Delta t/2}{2}}^{t} R'(t') dt'; \quad R' = R - \overline{R}
\]

where \( \overline{R} \) is the mean over the entire series. Mathematically, we have

\[
(\Delta R(\Delta t))_{\text{Haar}} \approx (\Delta R(\Delta t))_{\text{diff}}, \quad (0<H<1) \quad \text{and} \quad (\Delta R(\Delta t))_{\text{Haar}} \approx (\Delta R(\Delta t))_{\text{anom}}, \quad (-1<H<0).
\]

Now that we have defined the fluctuations, we need to characterize them; the simplest way is through (generalized) structure functions (generalized to fluctuations other than the usual differences, and generalized to moments of order other than the usual value 2):

\[
\langle \Delta R(\Delta t)^q \rangle \quad \text{where “<,>” indicates statistical (ensemble) averaging.}
\]

Physically, if the system is scaling, then the fluctuations are related to the driving flux \( \phi \) by:
\[ \Delta R(\Delta t) = \varphi_{\Delta t}\Delta t^H \]  

where we have used the subscript “\(\Delta t\)" on \(\varphi\) to indicate that it is the flux at resolution \(\Delta t\). The structure function is:

\[ \langle (\Delta R(\Delta t))^q \rangle = \langle \varphi_{\Delta t}^q \rangle \Delta t^{qH} \]  

Turbulent fluxes are conserved from scale to scale so that \(\langle \varphi_{\Delta t} \rangle = \text{constant}\) (independent of scale) implying that \(\langle \Delta R \rangle \sim \Delta t^H\) so that \(H\) is the mean fluctuation exponent. Beyond the simplicity of interpretation, the Haar fluctuations give a good characterization of the variability for stochastic processes with \(H\) over the range \(-1 < H < 1\). In contrast, fluctuations defined as differences or as anomalies are only valid over the narrower ranges \(0 < H < 1, \ -1 < H < 0\) respectively (see [Lovejoy and Schertzer, 2012a], [Lovejoy et al., 2013]). Outside these ranges in \(H\) the fluctuation at scale \(\Delta t\) is no longer dominated by frequencies \(\approx \Delta t^{-1}\) so that the fluctuations depend spuriously on details of the finite data sample, specifically either the highest or the lowest frequencies that happen to be present.

The generic scaling process is multifractal so that, in general, \(\varphi\) has statistics:

\[ \langle \varphi_{\Delta t}^q \rangle \sim \Delta t^{-K(q)} \]  

where \(K(q)\) is a convex function. Substituting this into Eq. 4, we obtain:

\[ \langle (\Delta R(\Delta t))^q \rangle \sim \Delta t^{\xi(q)}; \quad \xi(q) = qH - K(q) \]  

where \(\xi(q)\) is the “structure function exponent”. Although we return to this in more detail in Sec. III, for the moment note that the mean \((q=1)\) flux \(\langle \varphi_{\Delta t} \rangle\) is independent of \(\Delta t\), so that
$K(1) = 0$ and hence $\xi(1) = H$. Note also that for quasi Gaussian processes, none of the moments of $\varphi_{\Delta t}$ have any scale dependence so that $K(q) = 0$ and $\xi(q) = qH$ (all the scale dependence is characterized by $H$). A useful characterization of $K(q)$ is provided by $C_1 = K'(1)$ which quantifies the intermittency near the mean ($q = 1$, see below). Finally, the RMS fluctuation ($\left\langle [\Delta R(\Delta t)]^2 \right\rangle^{1/2}$) has exponent $\xi(2)/2$ so that the error in using the quasi-Gaussian approximation to estimate $H$ (i.e. the approximation $\xi(2)/2 = H$) is $\xi(2)/2 - H = K(2)/2$. In the temporal domain the latter is typically small – in the range $0.02$ to $0.04$ so that the approximation $\xi(2)/2 \approx H$ is fairly accurate. Figure 1b shows a direct estimate using the enormous CPC hourly gridded raingauge product over the continental US ($\approx 7 \times 10^7$ gridded quantities were used in the analysis, a product already derived from a much larger set of station series measurements). We see that the intermittency as quantified by $K(2)/2$ is large at weather scales (lower left of the figure, $K(2)/2 \approx 0.35$) but is much smaller at macroweather scales ($\approx 0.035$, upper right part of the figure). The use of the second moment is conventional since it directly determines the exponent $\beta$ of the spectrum $E(\omega) \approx \omega^\beta$, where $\omega$ is the frequency: $\beta = 1 + \xi(2)$; we therefore have used the RMS statistics below. However in dLL and in section III.B (using a different method) we show that spatial macroweather has $K(2)/2 \approx 0.1$ so that in the spatial domain, the approximation $\xi(2)/2 \approx H$ is poor, and a full multifractal characterization is needed (i.e. including $K(q)$). Note that in the weather regime, the spatial intermittency is even stronger: $K(2)/2 \approx 0.4$ (see Tables I and II for global scale weather regime estimates). Finally, even when $K(2)/2$ is small, the probability distribution of the fluctuations may be far from Gaussian. For example, dLL shows the distribution of monthly precipitation rate changes $Pr(\Delta R > s) \approx s^{-q_D}$ with $q_D = 3.6$
(a bit lower than the macroweather temperature value \( q_D \approx 5 \) but close to the weather scale precipitation value \( q_D \approx 3 \), see table 5.1b in [Lovejoy and Schertzer, 2013] for a review over a dozen references with \( q_D \approx 3 \)).

C. Overview of temporal and spatial macroweather precipitation statistics

In dLL the GHCN, 20CR and Smith data sets were systematically analyzed and compared over the ranges 1 month to \( \approx 100 \) years and from one pixel (primarily \( 5^\circ \times 5^\circ \)) to global scales; the overall summary of the statistical properties is shown in Tables I and II. In the time domain, for all products, we found very similar behaviors: for land, ocean, global, and for various latitude bands for pixel and global scales, we found \( H_t \approx -0.4 \) which is a little higher than Gaussian white noise \( (H = -\frac{1}{2}) \), the main differences being the amplitudes of the fluctuations (e.g. the RMS variability at a given scale such as one year) and the outer scale \( \tau_c \), the transition scale to the climate regime (here presumably associated with anthropogenic effects). The global scale analyses had \( \tau_c \approx 20 \) years whereas the pixel scale analyses (including several individual long station series from Portugal) had \( \tau_c \approx 40 \) years.

Fig. 2b shows the spatial distribution of \( H_t \) estimates for the three data sets. There is no obvious organization with possible exception of the 20CR analyses that showed a tendency of \( H_t \) to be low over oceans (especially the Pacific ocean) and most of the values are within 0.1 of the above cited mean. This is physically plausible since lower H values correspond to longer range predictability ([Lovejoy et al., 2015]), however, it could also be an artifact since the 20CR temperatures have lower \( H \) over the oceans and the 20CR calculates the precipitation rather indirectly. More precisely, on a pixel by pixel basis, for \( H_t \) for GHCN, 20CR and Smith data we found respectively: \(-0.41 \pm 0.07, -0.38 \pm 0.09, = -\)
0.43±0.10 (the corresponding estimates of the intermittency parameter $C_1$ were: 0.026±0.02, 0.020±0.025, 0.00±0.01). We therefore at least provisionally conclude that $H_t$ is globally constant with a value $\approx -0.4$ and the spatial variations are mostly due to estimation error.

TABLE I. Macroweather precipitation anomaly fluctuation exponents ($H$). The exponents were generally not estimated with high accuracy partly because of the high intermittency (the scaling was noisy) but also because more exact values are not warranted since the exact limits of the scaling ranges are not clear. Note that while the data agree very well on the temporal exponent, they disagree strongly on the spatial exponent including the sign (the EW and NS values of $H$ were judged to be quite close so only a single spatial value was given). For the spatial 20CR and Smith fields where the absolute precipitation rates were known, the spatial $H$'s for the raw and anomaly fields were quite close. CPC refers to the Climate Prediction Center, GHCN the Global Historical Climate Network, 20CR the Twentieth Century Reanalysis and “Smith” refers to the [Smith et al., 2012] IR satellite data based product.

<table>
<thead>
<tr>
<th></th>
<th>Time</th>
<th>Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPC</td>
<td>-0.42</td>
<td>_</td>
</tr>
<tr>
<td>GHNC</td>
<td>-0.4</td>
<td>-0.2</td>
</tr>
<tr>
<td>20CR</td>
<td>-0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>Smith</td>
<td>-0.4</td>
<td>0.25</td>
</tr>
</tbody>
</table>

TABLE II. A comparison of the spatial intermittency parameter $C_1$ and the effective outer scale ($L_{eff}$) for various data products at weather scales (ECMWF 3 hour resolution, 20CR, 6 hour resolution, CPC, one hour resolution, TRMM, 4 days resolution) and at macroweather scales (all at one month). ECMWF refers to the European Centre for Medium range Weather Forecasting interim reanalysis. The weather analyses and parameters were taken from [Lovejoy et al., 2012]. The “time” column refers to pixel scale spatial resolutions. For the macroweather regime, the $C_1$ of globally averaged values were slightly higher. Also, the CPC data in the macroweather regime gives $C_1 = 0.04$ (see Fig. 1b). The macroweather intermittencies ($C_1$) are significantly lower than the corresponding weather values, especially in time. The macroweather 20CR row reports two EW values from fig. 4c,d corresponding to 0-30° and 30-60°.
Since $C_1$ is very small, $H_t$ alone characterizes the long range statistical dependencies in the macroweather precipitation process and it turns out that the effective “memory” depends sensitively on $H$: the white noise value $-1/2$ corresponds to no memory (no predictability) while the value $H = 0$ to an infinite memory (perfect predictability), see [Lovejoy et al., 2015]. One way to quantify the memory is the fraction of the variance that can be explained by forecasting the future by one time step using the past data (i.e. a one month forecast with monthly data, a one year forecast with yearly data). For the global temperature $H \approx -0.2$ and about 35% of the variance can be explained, for ocean temperatures (SST’s) $H \approx -0.1$ and 65% can be explained. In contrast, for precipitation, with $H \approx -0.4$ only 4% can be explained (this might still be useful, especially if the skill could be increased with the help of co-correlates). For the temperature field, $H$ is found to vary relatively systematically, in particular, being strongly correlated with land or ocean location with the lower (more predictable) $H$’s generally over the ocean. It is therefore important to see if a similar systematic geographic distribution exists for precipitation. In dLL, averaging statistics over latitude bands failed to show significant variations.
Fig. 2b: The pixel scale world maps of the distribution of $H$ for monthly precipitation for the three data sets discussed in the paper (5°x5°, 3.75°x3.75° and 5°x5° resolutions for GHCN, 20CR, Smith data sets respectively, top to bottom; the 20CR resolution was degraded 2x2 pixels so as to be comparable to the resolutions of the other data sets). The exponents were estimated from the annual detrended data using the Haar analysis technique with exponents fit over the range 6 months to 12 years (top avoid possible biases at low frequencies due to anthropogenic effects or poor statistics). To avoid overcrowding, no geography was overlaid; it’s location can be judged from the top map. The red in the GHCN map corresponds to oceans, no data.

These temporal scaling statistics suggest that macroweather precipitation is reasonably well estimated, and this in spite of the quite different techniques used to estimate precipitation rates (instruments, reanalyses and satellite based). While the agreement of the statistics over wide ranges of time scales is a necessary condition for the
products to agree with each other, it is not sufficient: indeed, each product could be from a statistically independent realization of the same stochastic process. To gain further confidence in the quality, accuracy of the precipitation products, dLL therefore compared the product fields directly to each other by considering the difference of two products and studied the fluctuation statistics of the difference fields.

The dLL analysis showed that the agreement between the products was not so good. For example, for the GHCN and 20CR products, the agreement at scales below a year but also great than 10 years was low. Similarly, for the global scale 20CR and Smith products (the two that were not missing data), there was poor agreement until scales of 5 years or so and poor agreement at scales beyond about 30 years.

In space, the situation was somewhat different with poor agreement for scales below about 20°–30° latitude. It also seemed that the spatial scaling exponents were significantly different (see Table I), with the 20CR and Smith products being much smoother ($H_s$ about 0.4 to 0.5 larger than for the GHCN product). The smoothness may be an artifact of the limitations of the 20CR model and the smoothness of the satellite IR fields that were used to infer the Smith product.
The macroweather spatial and temporal scaling properties of precipitation that we have just described are qualitatively very similar to those of macroweather temperatures as analyzed, for example, in [Lovejoy and Schertzer, 2012b] and [Lovejoy and Schertzer, 2013]. The main differences between temperature and precipitation statistics is that the latter have generally lower temporal $H$ values, higher spatial $C_1$ values and somewhat longer transition scales $\tau_\circ$. Similar comments also pertain when comparing weather scale temperatures and precipitation (although in the weather regime, the intermittency parameter $C_1$ is much larger for precipitation than for temperature). This suggests that precipitation may be treated in the same theoretical framework as the temperature (and other atmospheric fields): that it can be modeled with the help of space-time cascade processes. In this picture, the scaling laws are emergent high level statistical (turbulent) laws expected to apply in the limit of high nonlinearity.

III. Space-time statistical factorization

A. Theoretical considerations

dLL and the other studies cited above indicate that in the macroweather regime there is good scaling in space and in time. This basically reflects the absence of strong scale breaking processes in either time or in space. Scaling symmetries also hold with high accuracy in the weather regime; [Lovejoy and Schertzer, 2013] is a review of 30 years of research supporting this conclusion, however the single Fig. 3 is enough to show – at least for thermal infra red satellite data (the type used by the Smith algorithm to infer precipitation) – that the symmetry is in fact quite exact. The figure uses nearly 1300 hours
of hourly geostationary MTSAT data (at 30 km resolution) from a 8000 × 13000 km² region centred on the equatorial Pacific (see [Pinel et al., 2014] for details). It can be seen that the zonal (i.e. EW), meridional (NS) and temporal spectra are nearly identical up to ≈ (5–10 days)⁻¹ and are nearly perfect power laws (most of the deviations from linearity in the figure can be accounted for by the finite resolution and finite data “window”, see the black line that theoretically takes these limits into account). The main exceptions are the two small spectral “bumps” at (1 day)⁻¹, (12 hours)⁻¹. These 1-D spectral densities were obtained by successively integrating out the complementary variables in the full (horizontal) space-time spectral density $P_{xyt}(k_x,k_y,\omega)$. This figure – and many others in [Pinel et al., 2014] – show that the spectrum satisfies the isotropic scaling symmetry:

$$P_{xyt}(\lambda k_x, \lambda k_y, \lambda \omega) = \lambda^{-s} P_{xyt}(k_x,k_y,\omega)$$

(7)

where, empirically, $s = \beta+2 \approx 3.4$ (note that at any given scale $P_{xyt}$ displays anisotropy; Eq. 7 simply implies that the anisotropy doesn't change with scale, see [Pinel et al., 2014] for the full analysis), and see [Pinel and Lovejoy, 2014] for their interpretation in terms of waves and turbulence.
FIG. 3. 1D spectra of MTSAT thermal IR radiances; the Smith product was developed with similar IR satellite radiance fields. In black: the theoretical spectrum using parameters estimated by regression from Eq. 7 and taking into account the finite space – time sampling volume. The spectra are $E_x(k_x) ≈ k_x^{-\beta_x}$, $E_y(k_y) ≈ k_y^{-\beta_y}$, $E_t(\omega) ≈ \omega^{-\beta_t}$ with $\beta_x ≈ \beta_y ≈ \beta_t ≈ 1.4 ± 0.1$; $s ≈ 3.4 ± 0.1$. The straight line is a reference line with slope $-1.5$ (blue). Pink, zonal spectrum; orange, the meridional spectrum; blue (with the diurnal spike and harmonic prominent) is the temporal spectrum. Reproduced from [Pinel et al., 2014].

By necessity, the scale symmetry (Eq. 7) can only hold up to planetary scales ($L_w$); this implies a breakdown in the time domain at scales $\tau_w$ which is interpreted as the lifetime of planetary scale structures. In Fig. 3 we find $\tau_w ≈ 5 – 10$ days; [Lovejoy and Schertzer, 2010] describes how this is determined by the turbulent energy flux $\varepsilon$ (power/mass): $\tau_w = \varepsilon^{-1/3} L_w^{2/3}$ where $\varepsilon$ itself is determined by the solar flux. This theory well describes the spectrum of the many atmospheric variables, ocean temperatures as well as the Martian weather and macroweather ([Lovejoy et al., 2014]).

B. Spatial intermittency, multifractality, cascades
We have mentioned that the spatial intermittency is much stronger than the temporal intermittency. One way to demonstrate this is to use the Haar fluctuations directly, estimating $\xi(q)$ from the exponents of the $q$th order moments and then estimate $K(q)$ as $-\xi(q) + q\xi(1)$ (see Eq. 6). Due to the missing data, this was the approach followed for the GHCN data (see Tables I and II). However, for the 20CR and Smith data, we can easily estimate $\varphi$ and hence $K(q)$ directly; for example, using Eq. 3 we can readily obtain:

$$\frac{\varphi}{\langle \varphi \rangle} = \frac{\Delta R}{\langle \Delta R \rangle}$$

(8)

If $\Delta R$ is estimated at the smallest available scale – for these complete, gridded data, we used the absolute second finite difference along the transect – then eq. 8 shows that the normalized high resolution flux is obtained by dividing by the mean fluctuation $<\Delta R>$. This high resolution flux can then be systematically degraded to lower resolution by averaging. The generic multifractal process is a multiplicative cascade; if such a process starts at outer scale $L_{eff}$, then the statistics follow:

$$\langle \Phi_{\lambda'}^q \rangle = \lambda'^{K(q)}; \quad \lambda' = L_{eff}/\Delta r$$

(9)

where $\lambda'$ is the scale ratio of the outer scale to the resolution scale of the degraded flux $\Delta r$ (see Eq. 5). In empirical analyses, $L_{eff}$ is not known a priori, it has to be estimated from the data; here we use a convenient scale ratio based on $L_w = 20000$ km, which is the largest distance on the earth (half a circumference), i.e. $\lambda = L_w/\Delta r$. If Eq. 9 holds, then, for all $q$ the lines of $\log \langle \Phi_{\lambda'}^q \rangle$ against $\log \lambda$ will cross at a scale corresponding to $\lambda_{eff} = L_w/L_{eff}$.

We mentioned in the introduction that there are two technical issues that have not always been carefully considered and which have led to incorrect macroweather exponent
estimates; they both involve incorrect flux estimates. For example, if daily precipitation is analyzed by estimating the flux as indicated (by absolute first or second differences), then one will obtain correct trace moment estimates of the weather regime exponents. However, the scaling (and exponents) of these daily resolution fluxes in the macroweather regime (e.g. when daily fluxes are averaged to monthly values), will be different from the scaling and exponents that would be obtained if we first took monthly averages and then estimated the flux at monthly resolutions before degrading the resolution further. It is the latter method that gives the correct macroweather exponent estimates. The reason for the difference is that taking absolute differences at the finest resolution (eq. 8) is a nonlinear transformation: weather and macroweather fluxes are physically different. Alternatively, if one has monthly data but one neglects to take the absolute differences (implicitly assuming that $H = 0$), then one has $\xi(q) = -K(q)$ (eq. 6 with $H = 0$ instead of $\approx -0.4$), so that one can easily find the values, $C_1 \approx 0.6$ (with $H$ implicitly = 0) instead of the correct values $H \approx -0.4$, $C_1 \approx 0$.

Figures 4a,b show the trace moment result when the various moments of order $2 \geq q \geq 0$ are estimated for the Smith anomaly data (i.e. the difference between the raw precipitation data and the long term averages, see section III.D) in the zonal and meridional directions respectively (even if the intermittency is low, higher order moments will generally diverge). It can be seen from the log-log linearity that the scaling is excellent up to about 10000 km (it is a little better in the zonal direction). In addition, the lines plausibly cross at a scale of the order of the size of the planet (see Table II); the fact that $L_{eff}$ can be a little larger than the size of the earth is because, even at planetary scales (20000 km), there is some residual variability due to the interaction of the precipitation field with other atmospheric fields –
$L_{\text{eff}}$ is simply the “effective” scale at which the cascade would have had to start in order to explain the statistics over the observed range. Fig. 4c, d show similar analyses for the 20CR anomaly data and 4e, f show the trace moments of the long term averages. This shows some differences between anomalies and long term averages with larger outer scales (meaning more variability at any given scale), but with no clear trends for $C_1$ (i.e. the rate at which the variability near the mean changes with scale).

The slopes in Figs. 4 determine $K(q)$ (Eq. 9); the latter is a convex function, equivalent to an infinite number of parameters. Fortunately, one can avail of a multiplicative analogue of the usual (additive) central limit theorem for random variables. This shows that under fairly general circumstances $K(q)$ is determined by only two parameters that define multifractal “universality classes”:

$$K(q) = \frac{C_1}{\alpha - 1} (q^\alpha - q); \quad 0 \leq \alpha \leq 2$$

where $\alpha$ is the Levy index and $C_1$ is the codimension of the mean [Schertzer and Lovejoy, 1987]. From Eq. 10, we see that $C_1 = K'(1)$; this provides a convenient way of estimating the parameters (one can also use $\alpha = K''(1)/K'(1)$). Table II (bottom rows) shows the resulting parameter estimates for the macroweather regime and compares these with those in the weather regime (top rows). The estimates of the parameter $\alpha$ were all in the range 1.7 to 1.85 (see also table III). Also, note that the difference between the exponents of the RMS and mean (important for interpreting the slopes in the RMS Haar graphs), $\xi(2)/2 - H = K(2)/2 = A_\alpha C_1$ where $A_\alpha = (2^{\alpha-1} - 1)/(\alpha - 1)$. Since empirically $1.85 > \alpha > 1.7$, we find $0.95 > A_\alpha > 0.9$ (near one) so that $C_1$ provides a good estimate of the error in using the RMS exponent $\xi(2)/2$ in place of $H$. From Table II we see that, in space, the difference $\xi(2)/2 - H$ can be
readily in the range 0.1 to 0.2: it is significant. Also notable in Table II is the comparison of weather and macroweather exponents and outer scales. We can see that while the outer scales are very similar, that weather scale precipitation is much more variable (the $C_1$ parameter is much higher) than macroweather precipitation.

Table III: This shows a comparison of various spatial exponent estimates using the 20CR data in latitudinal bands between 0-30°N and 30-60°N over the period 1871-2012.

<table>
<thead>
<tr>
<th></th>
<th>0-30°N</th>
<th></th>
<th>30-60°N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R(r,t)$ (anomaly)</td>
<td>$R_c(r)$ (mean)</td>
<td>$R_{raw}(r,t)$ (raw)</td>
</tr>
<tr>
<td>$C_1$</td>
<td>0.10</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.7</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>Outerscale (km)</td>
<td>90000</td>
<td>28000</td>
<td>45000</td>
</tr>
</tbody>
</table>

FIG. 4a. Zonal (EW) trace moments for the Smith anomaly data spanning the region from -45° to 45° latitude. Note the convergence close to 20000 km and the low $C_1$ value. $L = L_w/\Delta r$ with $L_w = 20000$ km. The top curve is for $q = 2$, with $q$ ranging down to 0.1.
FIG. 4b. Same as Fig. 4a but for the meridional (NS) direction.
Fig. 4c (top left), d (top right), e (bottom left), f (bottom right): The trace moment analyses of the 20CR anomalies in the east-west direction (at 1.875° resolution), averaging the data over the bands from the equator to 30°N (left) and from 30°-60°N (right). The outer reference scale is 180° which corresponds to nearly 20000km in the left, but only 14,000km in the right so that for the anomalies (c, d) \( L_w \approx 25,000 \text{km}, 14,000 \text{km} \) respectively. Also for the anomalies, the \( C_1 \) parameter is 0.12, 0.09 (0-30°N and 30°-60°N) respectively, i.e. slightly higher near the equator (see table 2). The bottom row (fig. 4e, f) show the corresponding spatial analyses of the long term average rates from 1871-2012. The parameters are \( C_1 \approx 0.17, 0.13 \) and outer scales 28000, 56000km respectively, see table III.

C. Space-time cascades, the EFIF model and macroweather factorization and the homogeneous spatial SLIM model

The statistical weather and macroweather behaviors (including the transition scale between them) can be modeled by assuming that a multiplicative cascade starts at a scale \( L_w \) in space, but at a scale \( \tau_c \) much larger than \( \tau_w \) in time, the Extended Fractionally Integrated Flux (EFIF) model. To see this, write the weather-macroweather flux \( \varphi_{w,mw} \) (subscript “\( w \), “\( mw \)”), here for precipitation, see Eqs. 3 and 9 in terms of its “generator” \( \Gamma \):

\[
\varphi_{w,mw}(r,t) = e^{\Gamma}
\]  

(11)

\( \Gamma \) is a scaling additive process and \( \varphi \) is multiplicative (see [Schertzer and Lovejoy, 1987]). If \( \Gamma \) has the appropriate scaling, then \( \varphi \) will satisfy Eq. 9 of a multiplicative cascade, it will follow the behavior in Figs. 4a,b. For this, we require:

\[
\Gamma = \gamma \star g; \quad g \propto \Theta(t)[r,t]^{-D(1-1/\alpha)}; \quad [r,t] = \left( |r|^2 + t^2 \right)^{1/2}
\]  

(12)

where “\( \star \)” indicates space-time convolution over the entire region for which the cascade occurs and \( [r,t] \) is a space-time scale function; we have chosen the simplest possible \( g \) consistent with the isotropic symmetry in Eq. 7. In Eq. 12, \( \gamma \) is a Levy noise subgenerator index \( \alpha \) made up of independent identically distributed random variables representing the
innovations, \( D = 3 \) is the dimension of (horizontal) space-time and \( \Theta(t) \) is the Heaviside function \((=0 \text{ for } t<0, =1 \text{ for } t\geq0)\) needed to take causality into account (see [Marsan et al., 1996] and ch. 9 in [Lovejoy and Schertzer, 2013]). We have nondimensionalized space and time using the planetary scale \( L_w \) distances and the lifetime of planetary structures \( \tau_w \).

From weather to macroweather: a dimensional transition

\[
\Gamma(r,t) = \int_{\Lambda_w}^{1} \gamma(r-r',t-t')g(r',t')dr'dt' + \int_{1}^{\Lambda_c} \gamma(r-r',t-t')g(r',t')dr'dt'
\]

\[
\Phi_\lambda = e^{i(\lambda\cdot \mathbf{r})} \quad \Gamma_{\text{w}} \quad \Gamma_{\text{mw}} \quad \Gamma = \gamma \ast g \quad g = (|\mathbf{r}|+\tau')^{1/D'}
\]

**FIG. 5.** A schematic diagram showing the weather and macroweather regions of integration for the cascade convolution (Eq. 12) in horizontal-time space \((x,y,t)\) nondimensionalized by using the size of the earth and the lifetime of planetary structures. \( B_1 \) is the space-time region corresponding to weather processes \((r<1, \ t<1; \ r=(x^2+y^2)^{1/2}) \) and \( B_w \) is the corresponding inner (dissipation) scale, size \( \Lambda_w^{-1} \) (a factor \( \Lambda_w \) times smaller). The cylindrical region between the red (dark, larger, filled) circles is macroweather region with \( r<1, \Lambda_c>>>1 \); in this region the temporal variability dominates the integral since the convolution kernel \( g \) is nearly independent of \( r \) (\( \gamma \) is an i.i.d. Levy noise subgenerator representing the innovations).
Figure 5 shows a schematic of the region of integration: it can be seen that for the convolution in Eq. 12, the domain of integration can be separated into two statistically independent terms as:

\[ \Gamma_{w,mw}(\mathcal{L}, t) = \Gamma_w(\mathcal{L}, t) + \Gamma_{mw}(\mathcal{L}, t) \quad (13) \]

where the weather and macroweather generators are:

\[ \Gamma_w(\mathcal{L}, t) = \int_{B_w} \int_{\Lambda_w} \gamma(r, r', t, t') g(r', t') dr' dt' \]
\[ \Gamma_{mw}(\mathcal{L}, t) = \int_{B_w} \int_{1/\Lambda_w} \gamma(r, r', t, t') g(r', t') dr' dt' \quad (14) \]

In the above, the weather generator \( \Gamma_w \) is obtained by integrating over the region between \( B_1 \) (the space-time region corresponding to weather processes \( r<1, t<1; r = (x^2 + y^2)^{1/2} \) and \( B_w \) which is the corresponding inner (dissipation) scale, size \( \Lambda_w^{-1} \) (a factor \( \Lambda_w \) times smaller; see Fig. 5). The macroweather generator \( \Gamma_{mw} \) is obtained from the cylindrical region between the red circles i.e. the region with \( r<1, \Lambda_c>t>1 \); in this region the temporal variability dominates the integral since the convolution kernel \( g \) is nearly independent of \( r \). \( \Lambda_w^{-1} \) is the inner scale of the weather processes and \( \Lambda_c = \tau_c/\tau_w \) is the outer scale of the temporal cascade. The first term on the right of Eq. 13 corresponds to a roughly isotropic space-time region so that all the degrees of freedom (the \( \gamma \)'s) are important in the left most integral and, due to the astute choice of exponent \( (D(1-1/\alpha)) \), the weather regime statistics follow the multiplicative cascade law eq. 9 with \( K(q) \) given by eq. 10 (the value of \( C_1 \) is determined by the proportionality constant for \( g \) in eq.12, for full details, see the review, ch. 5, Lovejoy and Schertzer 2013). However, the macroweather term (the second term on the right in Eq. 13) is pencil-like, and over the macroweather range of temporal
integration $t \to r$ so that $g = \Theta(t)r^{-D(1-\alpha)}$ and hence it is mostly the temporal degrees of freedom that are effective, the spatial degrees of freedom are essentially “quenched”; the system is essentially reduced in dimension from a 2+1 horizontal space, time to a 1 dimensional system (time), a “dimensional transition” [Lovejoy and Schertzer, 2010].

Exponentiating the generators, we obtain the following multiplicative relationship for the fluxes:

$$\varphi_{w,mw}(r,t) = \varphi_w(r,t)\varphi_{mw}(r,t); \quad \varphi_{w,mw} = e^{\Gamma_{w,mw}}; \quad \varphi_w = e^{\Gamma_w}; \quad \varphi_{mw} = e^{\Gamma_{mw}}$$

(15)

Due to the quenching, we see that the macroweather flux has practically no variability at nondimensional space or time scales $<1$ (i.e. smaller than $L_w$, shorter than $\tau_w$), thus:

$$\varphi_{w,mw}(r,t) = \varphi_w(r,t)\varphi_{mw}(r,t) = \varphi_w(r,t)\varphi_{mw}(t)$$

(16)

where the far right approximation is due to the quenching. If we now average this equation over nondimensional times $=1$ (i.e. over time scales up to $\tau_w$), due to the space-time coupling in the weather regime, this temporal averaging effectively averages out the spatial variability as well so that:

$$\varphi_{w,mw,\tau}(r,t) \approx \varphi_{mw}(t); \quad \tau_w < \tau < \tau_c$$

(17)

where $\varphi_{w,mw,\tau}$ is the temporally averaged flux over a scale $\tau \gg \tau_w$. Writing $\varphi_{mw}(r,t) \approx \varphi_{mw}(t)$ is an abuse of notation that will be clarified in the next subsection. It should be interpreted as indicating that the type of temporal statistical process is the same at all spatial locations (independent of $r$), not in the strict deterministic sense that the actual value of $\varphi_{mw}$ is identical at different locations, we discuss this further below.
From this analysis we see that a single weather/macroweather cascade model will yield completely spatially smooth macroweather fields. This led [Lovejoy and Schertzer, 2013] to propose that models of atmospheric variability valid in the macroweather regime must include an additional slowly varying space-time climate process \( \varphi_c(r,t) \) (that only varies significantly at long time scales \( \tau \gg \tau_c \)), the consequence of physically different climate processes. The full (weather-macroweather-climate) model is thus:

\[
\varphi_{w,c}(r,t) = \varphi_{w,mw}(r,t) \varphi_c(r,t)
\]

(18)

since by definition for time scales \( \tau < \tau_c \), the climate process has \( \varphi_c(r,t) = \varphi_c(r) \), over the macroweather regime, we find:

\[
\varphi_{w,c,t}(r,t) = \varphi_{m,w}(t) \varphi_c(r); \quad \tau_w < \tau < \tau_c
\]

(19)

where \( \varphi_{w,c,t}(r,t) \) is the full space-time weather - climate process averaged over time scale \( \tau \). This is the macroweather space-time factorization property of the fluxes, it is compatible with the scaling and it is predicted by this simple model. Since the observables (e.g. the precipitation rate) are obtained by fractional integration (power law filters, order \( H \)) of the fluxes, we obtain equivalent space-time factorizations of the observables. Physically, the interpretation is that the macroweather spatial variations are associated with different climate zones but that each zone has the same statistical type of temporal variability whose origins are in high frequency weather dynamics. In other words, the climate spatially modulates the atmosphere at macroweather scales even though the climate temporal variations are at longer scales.

**D. A simple scaling macroweather model: multifractal SLIM**
Motivated by the weather-climate EFIF based model described above and in order to allow for practical applications such as macroweather precipitation forecasting, let us make a pure macroweather model with the minimum ingredients that will at least approximately reproduce our empirical analyses. In the temporal domain, we take the simplest nonintermittent model - a Gaussian – so that we pay special attention to the second order statistics characterized by autocorrelation functions (“c”). With this Gaussian approximation, and since $-1 < H_t < 0$, we are led to the fractional Gaussian noise (fGn) process:

$$\phi_{mw}(r,t) = \frac{1}{\Gamma(1/2 + H_t)} \int_{-\infty}^{t} (t - t')^{-(1/2 - H_t)} \gamma(r,t') dt'$$  \hspace{1cm}(20)$$

where $\Gamma$ is the usual gamma function (not to be confused with the cascade generator). The innovations $\gamma$ are taken as zero mean Gaussian white noise processes ($\langle \gamma(r,t) \rangle = 0$) having an autocorrelations:

$$c_\gamma(\Delta r, \Delta t) = \langle \gamma(r - \Delta r, t - \Delta t) \gamma(r, t) \rangle = c_{slow}(\Delta r) \delta(\Delta t)$$  \hspace{1cm}(21)$$

so that for each position $r$, $\gamma$ is a temporally “delta correlated” Gaussian white noise process, the integration in eq. 20 is a temporal fractional integral of order $1/2 + H_t$ of $\gamma$. For each $r$, $\gamma$ is a different but correlated white noise with slowly varying spatial correlations $c_{slow}(\Delta r) \approx$ constant (since most of the spatial variability in the weather process is averaged out at macroweather scales, for emphasis this slow variability was suppressed in eq. 17 above). While at each point in space, $\phi_{mw}$ is an fGn process; it’s (order one) time integral is a (more familiar) fractional Brownian motion.

With this definition, the autocorrelation function of $\phi_{wm}$ is:
\[ c_{\phi_m} \left( \Delta r, \Delta t \right) = \langle \phi_m (r - \Delta r, t - \Delta t) \phi_m (r, t) \rangle = c_{\text{slow}} \left( \Delta r \right) \Delta t^{2H}, \quad (22) \]

(we have absorbed some numerical factors in \( c_{\text{slow}} \), see [Lovejoy et al., 2015] for details on this). From eq. 22 and since \( \langle \gamma (r, t) \rangle = 0 \) and \( H_t < 0 \), we see that long time averages (corresponding to fluctuations over time periods \( \Delta t \)- recall that since \( H_t < 0 \), that the averages are good definitions of fluctuations), of \( \phi_{wm} \) will converge to zero. Therefore, this model is for precipitation anomalies; we return to this below.

In the time domain, such a process has \( \langle \left| \Delta \phi_{m} (r, \Delta t) \right|^q \rangle = \Delta t^{q} \xi(q) \) with \( \xi(q) = H_t q \) so that \( K_t(q) = 0 \), the process is an not a cascade in time (corresponding to the observations that \( C_{1t} \approx 0 \)). However in space, we saw that it has reasonably intermittent multifractal behavior \( (C_{1x} \approx 0.15, \text{ see table II}) \), hence we assume that \( \varphi_c \) is the flux of a conservative “climate regime” cascade:

\[ \langle \phi_{c, \lambda} (r)^q \rangle = \lambda^{K_r(q)}; \quad \lambda = \frac{L_{\text{eff}}}{l} \tag{23} \]

where \( \lambda \) is the scale ratio of the outer cascade scale \( L_{\text{eff}} \) to the resolution \( l \). The corresponding correlation function is:

\[ c_{c, \lambda} (\Delta r) = \left( \frac{\left| \Delta r \right|}{L_{\text{eff}}} \right)^{-K_r(2)} = \lambda^{K_r(2)} \left( \frac{\left| \Delta r \right|}{l} \right)^{-K_r(2)}; \quad \left| \Delta r \right| > l; \quad l = L_{\text{eff}} \lambda^{-1} \tag{24} \]

(where for simplicity we have assumed that the process is statistically isotropic, the horizontal scale depends on the vector norm which is isotropic). The restriction \( \left| \Delta r \right| > l \) is because by definition the flux \( \varphi_{c, \lambda} \) at scale ratio \( \lambda \) is smooth at scales \( < l \). As above, the
overall model for the flux is thus \( \Phi_{w,c} = \Phi_{c,\lambda} \Phi_{\omega m} \) so that (inverting the order of spatial and temporal integration) for the rainrate we have:

\[
R(r,t) = \Phi_{w,c} * x \left| \mathcal{L}^{(D-H_s)} r \right| \Phi_{c,\lambda} \Phi_{\omega m} \left| \mathcal{L}^{(D-H_s)} r \right| = \frac{1}{\Gamma(1/2 + H_t)} \int_{-t}^{t} (t-t')^{-(1/2-H_t)} \gamma'(r,t) dt'
\]

\[
\gamma'(r,t) = \left[ \int_{\text{sphere}} \Phi_{c,\lambda} \left( r' \right) \left| \mathcal{L}^{(D-H_s)} r' \right| \gamma \left( r', t \right) dr' \right]
\]

where the asterisk indicates a convolution with respect to space only ("x") this is a fractional integration of order \( H_s \); \( D \) is the dimension of space (here the horizontal only is considered, \( D=2 \)). The overall precipitation rate time series is therefore also an fGn process with spatially varying amplitude. If we consider a single realization of the process, and consider a point \( r \) and average \( R \) over a resolution \( \tau < \tau_c \) ( \( R_t(r,t) = \frac{1}{\tau} \int_{t-\tau}^{t} R(r,t') dt' \) ) then at the point \( r \)

\[
\sigma_{R,t}(r) = \left( \overline{R_t(r,t)^2} \right)^{1/2} \approx \sigma_{\gamma}(r) \tau^H, \quad \sigma_{\gamma}(r) = \left( \overline{\gamma'(r)^2} \right)^{1/2}
\]

where \( \overline{\cdot} \) indicates a time average over a long time scale (ideally of order \( \tau_c \) so that the time average is a good approximation to the average over an ensemble of statistically identical time series, if we average for \( \tau > \tau_c \) then the approximation that \( \Phi_{c,\lambda} \) is constant in time is no longer valid). Notice that in the macroweather regime, the temporal resolution \( \tau \) of the data is fundamental; \( \sigma_{R,t} \) diverges as \( \tau \) approaches zero (recall \( H_t < 0 \), although the law itself breaks down at \( \tau = \tau_w \). The scaling, eq. 26 and many other properties of fGn processes are reviewed in [Lovejoy et al., 2015]. Using \( \sigma_{R,t}(r) \) we can obtain the "homogenized" process: \( R_\ast(r,t)/\sigma_{R,t}(r) \) whose statistics are the same everywhere. This justifies a common climatological homogenization practice used for example to produce
climate indices (such as drought indices). Note that this works as long as $H$ is independent of spatial location; while this is apparently reasonable for precipitation, it is a poor approximation for the temperature field which involves spatially varying $H$.

While each spatial location is an $f$Gn process, the processes at neighbouring locations are correlated. First consider the spatial autocorrelation of the temporal Gaussian white noise process $\gamma'$:

$$c_{\gamma'}(\Delta r, \Delta t) = |\Delta r|^{\xi_{\Delta r}} c_{\text{slow}}(\Delta r) \delta(\Delta t); \quad \xi_{\Delta r} = qH_x - K_x(q)$$

(27)

so that the overall macroweather rainrate process (i.e. valid for time scales $\tau_w < \Delta t < \tau_c$) has autocorrelation:

$$c_x(\Delta r, \Delta t) = c_x(\Delta r) c_t(\Delta t); \quad c_x(\Delta r) \propto |\Delta r|^{\xi_{\Delta r}} c_{\text{slow}}(\Delta r); \quad c_t(\Delta t) = \Delta t^{2H};$$

(28)

or more generally, for the space-time rainrate fluctuations, any moment $q$:

$$\left\langle \Delta R(\Delta r, \Delta t)^q \right\rangle \propto |\Delta r|^{\xi_{\Delta r}} \Delta t^{qH},$$

(29)

Note that in practice we have a single realization of $\varphi_c(r)$ so that we must approximate the ensemble averages by spatial averages.

The statistical factorization (eq. 29) implies that macroweather statistics are profoundly different from weather statistics. In the weather regime, we are used to reasonably well defined space-time equivalences, based physically on the fact that structures of a certain spatial extent have statistically well defined lifetimes that grow with increasing spatial scale (big structures live longer than small ones). Ignoring various spatial anisotropies and nondimensionalizing space and time with $L_w$, $\tau_w$, the weather structure function has:
\[ \langle \Delta R_w(\Delta r,\Delta t)^q \rangle \propto \| (\Delta r, \Delta t) \|^{\xi(q)} ; \quad \| (\Delta r, \Delta t) \| \approx \left( |\Delta r|^2 + |\Delta t|^2 \right)^{1/2} \]  

(30)

where \( \| \| \) is the scale function. The space-time equivalence is obvious since the fluctuations at pure space scale \( \| (\Delta r,0) \| \) and pure time scale \( \| (0,\Delta t) \| \) are the same whenever the nondimensional lags are equal: \( |\Delta r| = \Delta t \). Physically, \( \Delta t \) is the typical lifetime of structures of size \( |\Delta r| \). In comparison, in the macroweather regime (eq. 29), average fluctuations \( q = 1 \) will be constant whenever \( \langle \Delta R(\Delta r,\Delta t) \rangle = |\Delta r|^{H_{t}} \Delta t^{H_{x}} \). Recall that when \( H_{t} < 0 \), fluctuations are essentially averages of anomalies, so that \( \Delta t = \langle \Delta R(\Delta r,\Delta t) \rangle^{-1/H_{t}} |\Delta r|^{-H_{x}/H_{t}} \) represents the time scale over which anomalies of spatial extent \( |\Delta r| \) must be averaged in order yield an average fluctuation \( \langle \Delta R(\Delta r,\Delta t) \rangle \). If we consider the GHCN anomalies with \( H_{t} = -0.4, H_{x} = -0.2 \), we have \( -H_{t}/H_{x} \approx -0.5 \) so that \( \Delta t = \langle \Delta R(\Delta r,\Delta t) \rangle^{2.5} |\Delta r|^{-0.5} \). Due to the negative exponent, this is an inverse relation between temporal and spatial averaging for precipitation anomalies. For example, monthly data spatially averaged over 1000 km would have the same ensemble mean value as data temporally averaged over 16 months but spatially averaged over only 1000/16^{0.5} = 250 km. While this statistical relation may be valid, it doesn’t imply a physical link between space and time (we could add that due to the fairly strong spatial intermittency, the space-time relation will be different for different moments \( q \)).

The rainrate anomaly process \( R(\Delta t) \) is the homogeneous spatial extension (i.e. with constant \( H_{t} \)) of the basic Scaling Linear Macroweather (SLIM) model described in [Lovejoy et al., 2015]; it has been derived starting from the Extended Fractionally Integrated Flux
model. It is approximate since the EFIF is not exactly an fGn process (see appendix 10A of Lovejoy and Schertzer 2013), and because the assumption that the spatial variability mostly derives from a low frequency multiplicative climate regime process which is an admittedly ad hoc model choice. We could also note here that we have assumed that $H_t$ is constant in space; the relaxation of this assumption is important in application to macroweather forecasting and will be discussed elsewhere.

We saw above that temporally averaging the anomaly process $R(r,t)$ over duration $\tau$ diminishes its amplitude $\sigma_{R_t}$ by $\tau^{H_t}$ (eq. 26), therefore long term time averages $\overline{R_t(r,t)}$ tend to zero; the model is an anomaly model. To obtain the actual (raw, measured) rain rate $R_{raw}(r,t)$, the anomaly must be added to another low frequency climate process $R_c(r)$:

$$R_{raw}(r,t) = R(r,t) + R_c(r)$$

(31)

Temporally averaging the anomaly process $R(r,t)$ over duration $\tau$ diminishes its amplitude so that:

$$R_c(r) = \overline{R_t(r,t)}; \quad \tau = \tau_c$$

(32)

is $\tau$ is large enough (preferably close to $\tau_c$), then $R_c(r)$ can be estimated by long time averages. Like the anomaly amplitudes controlled by $\phi_c(r)$, at scales long than $\tau_0$, $R_c(r)$ starts to vary - it “wanders” since $H_{t,c}>0$. Further study of the link between the climate scale amplitude of the anomaly fluctuations ($\phi_c(r)$) and the long term mean values ($R_c(r)$) is outside our scope.

To summarize: the model for $R(r,t)$ describes macroweather precipitation variability as a fractional Gaussian noise process with spatially correlated amplitudes. The space-time statistics factor into spatial and temporal functions. Over periods over which $\phi_c$ can be
considered independent of time (i.e. $\Delta t < \tau_c$, less than decadal, centennial scales), the spatial correlations are determined by $\varphi_{c,\lambda}(r)$ which is a single realization of a multifractal climate process at resolution $l = L_{\text{eff}}^{-1}$ where $L_{\text{eff}}$ is the effective outer scale of the climate cascade process. At each spatial location, the time series is a fGn process parameter $H_t$ with standard deviation given by $\sigma_{R, t}(r)$ and depending on the temporal resolution via $\tau_{H_t}$. In this way, the local statistics of the precipitation anomaly are determined in a multiplicative way by the climate process $\varphi_{c,\lambda}(r)$. This justifies the simple expedient of normalizing (nondimensionalizing) the precipitation process at each spatial location by the standard deviation estimated from the time average at $r$: $(\sigma_{R, r}(t))$, see eq. 26.

E. Empirically testing the statistical space-time factorization property

The space-time spectral density (“$P_{xyt}$”) corresponding to the autocorrelation eq. 28 is:

$$P_{xy}(k, \omega) = F.T.\left(c_r(\Delta r, \Delta t)\right) \approx F.T.\left(c_x(\Delta r)\right) F.T.\left(c_t(\Delta t)\right) \approx P_x(k) P_t(\omega) = |k|^{-\beta_x} \omega^{-\beta_t}$$  \hspace{1cm} (33)

which is expected to be valid for $\tau_w^{-1} < \omega < \tau_c^{-1}$ and where:

$$\beta_x = D - 1 + 2 H_x - K \hspace{1cm} \beta_t = 2 H_t$$  \hspace{1cm} (34)

where “F.T.” means Fourier transform: this is the Wiener-Khintchin theorem relating the spectral density and the autocorrelation. In the above two equations, we have assumed that $c_{\text{slow}} \approx \text{constant}$, and quantities averaged over time scales $\tau_c > \tau > \tau_w$ i.e. in the macroweather regime.

In order to test this, we can simplify by considering zonal wavenumber frequency spectra. First, to simplify, consider only the zonal direction (wavenumber $k$).
If factorization holds, then: $P_{x,t}(k,\omega) = P_x(k)P_t(\omega)$ and the density $P_{x,t}(k,\omega)$ will have iso-density contours parallel to the $k, \omega$ axes. Fig. 6a shows the result for the 20CR reanalysis data, considering two latitudinal bands. Although the contours are noisy, they are fairly parallel to the axes hence they are compatible with statistical factorization. At the lowest 3 or so available frequencies (i.e. $\approx$ 30–40 year scales) factorization breaks down – the beginning of the climate regime – the vertically aligned contours near the centre of the plot. Elsewhere we show that this factorization behavior is reasonably well reproduced by GCM’s although with somewhat different amplitudes and spatial scaling laws. A figure very similar to this, showing the factorization of space-time macroweather temperature spectra, was given in [Lovejoy and Schertzer, 2013], in section 10.3.

In order to investigate this further, we can directly check the constancy of the ratio $P_{x,t}(k,\omega)/(P_x(k)P_t(\omega))$; the result is shown in fig. 6b. The various horizontal sections are for $k$ increasing by factors of two from bottom to top, we see that except for the lowest frequencies, the ratio is very constant, especially for the 30-60°N band (right), for example the variations about the bottom right line are ±20% while the overall variation of $P_{x,t}$ over the same range is $\approx 10^7$. 


FIG. 6a. The space-time spectral densities $P_x(k,\omega)$ for the 20CR monthly precipitation product over the period 1871-2012. The horizontal axis is for the frequencies (in units of cycles/141 years), the vertical axes are for the zonal wavenumbers (up to ±180° longitude). Contours of the spectral densities for latitudes 0° to 30°N, and 30° to 60°N are shown on bottom and top. Note the annual cycle and the harmonics (the regularly spaced spikes). Note that the contours lines are generally parallel to the axes as expected by the factorization property.
FIG. 6b. A direct test of the factorization hypothesis for the spectra analysed in fig. 6a. Each line shows the ratio of zonal wavenumber – frequency spectral density $P_x(k,\omega)$ to the product of the zonal only $P_x(k)$ and frequency only $P_t(\omega)$ densities for zonal wavenumbers increasing in octaves from bottom to top. Each section is displaced by a factor $e^{0.5}$ in the vertical and colors alternate for clarity. The frequency units are (months)$^{-1}$ and the wavenumber units are cycles per half-circumference (due to the latitudinal dependent map factor, about 20000 km on the left, about 14100 km on the right). The constancy of the ratio is remarkable when it is considered that the total variation (maximum/minimum) of $P_x(k,\omega)$ over the range of $(k,\omega)$ in fig. 6b is $\approx 10^7$.

We can also check the factorization directly in real space using joint space-time Haar fluctuations. In this case, for the RMS fluctuations we expect the joint $S(\Delta x, \Delta t)$ to be the product of a spatial function $S_x$ and temporal function $S_t$:

$$S(\Delta x, \Delta t) = S_x(\Delta x)S_t(\Delta t)$$

(34)
(in the analyses, we actually use the zonal angle so that the symbol $\Delta x$ is the longitude subtended by an interval).

In order to facilitate the interpretation of the results, we can use the fact that in the time domain the fluctuations decrease with scale, hence if we simply average the anomalies over longer and longer time scales, the resulting averages are virtually the same as the corresponding Haar fluctuation (the differencing in Eq. 1 has little effect when $H<0$). In space, we use the usual (spatial) Haar fluctuation of the temporally averaged data. For the GHCN data, this joint space-time analysis is shown in Fig. 7a. We see not only that the curves parallel for different amounts of averaging – the basic factorization prediction - but also that as we double the averaging time $\Delta t$ the curves are roughly equally spaced; this demonstrates the scaling in time. In addition, the structure functions are more or less linear on the log-log plot so that there is also scaling in space. Overall we have Eq. 21 with:

\[
S_x(\Delta x) \approx \Delta x^{\xi_x(2)/2}; \quad S_t(\Delta t) \approx \Delta t^H_t,
\]

(35)

with $\xi_x(2)/2 \approx -0.3$, $H_t \approx -0.4$. Taking $K_x(2)/2 \approx 0.1$ (Table II) we have $H_x \approx -0.2$. Notice that since the temporal fluctuations are estimated by simply averaging the anomalies ($q = 1$), the spacing in the vertical allows us to directly infer $H_t$. In contrast, the RMS spatial fluctuations have exponent $\xi_x(2)/2$ which is a poor approximation to $H_x$ due to the large spatial intermittency; it requires the correction $K_x(2)/2$ (see the estimates in Table II).

In Fig. 7b, we show the analogous plot for the Smith product. Once again, the curves for different averaging times (top to bottom) are roughly equally spaced as predicted by Eq. 21; and the reference lines are again spaced at factors $4^{-0.4}$ as in Fig. 7a (corresponding to $H_t = -0.4$, with a factor 4 of temporal averaging, the same reference lines as the GHCN,
Fig. 7a). However, the slopes are very different ($\xi_s(2)/2 \approx +0.4$, hence $H_s \approx 0.25$), rather than $(\xi_s(2)/2 \approx -0.3$ for the GHCN anomaly, Fig. 7a, so that although the Smith anomaly also respects the factorization property, it has very different spatial statistics than the GHCN data. Indeed, the Smith $H_s$ (Table I) is very close to those of the spatial satellite IR fields of the type that were used in its construction in the range 0.2 to 0.25; (see [Lovejoy et al., 2009]). Finally, fig. 7c shows the corresponding plot for the 20CR data. It is fairly close to the Smith data, although with slightly different parameters. In addition, the temporal scaling breaks down between about 10 and 20 years (128 and 256 months).

Also shown in Fig. 7b are the corresponding analyses for the raw rain rates averaged over 1, 2, 4, ... 512 months. We see that after about a year of temporal averaging the statistics have nearly converged to their long term values: the spatial variability of annual averaged precipitation is nearly the same as for centennial averages.

Note that, when performing averaging on anomalies (the bottom curves), as explained above, since $H_c<0$, the result is close to the corresponding Haar fluctuation. However this is not true when averaging the raw data so that the top curves are not joint space-time Haar fluctuations, but simply spatial Haar fluctuations at various temporal resolutions. In fact, there is a difference in the anomalies and raw statistics in space, but not in time (see dLL for details).
FIG. 7a. The log$_{10}$ of the RMS structure functions of GHCN (anomalies) in the zonal (EW) direction (units: mm/month) for averaging times increasing from top to bottom by factors of 2 (in months). The longitudinal angle subtended is indicated: $\Delta \theta$ in the figure (i.e. an angular “lag” rather than a distance, the data were from latitudes between $\pm 45^\circ$ so that the difference is not very important), $\Delta x$ in the text. One can see that up to the limit of the macroweather regime (in red – thick line –, about 256 months i.e. $\approx 20$ years), that the effect of averaging is essentially to systematically decrease the spatial structure function but without changing it’s shape (in this case, reasonably close to a power law fall-off). In addition, the dashed reference lines slopes $\xi_x(2)/2 \approx -0.3$, corresponding to $H_x \approx -0.2$ (since $C_1 \approx 0.1$, see Table II) are spaced so as to correspond to a factor $4^{-0.4}$ i.e. the theoretical spacing for two curves with a factor of 4 different in averaging time and with $H_t = -0.4$. This indicates that the joint structure function $S(\Delta x, \Delta t)$ satisfies Eq. 33, i.e. the predicted macroweather “factorization” property.
FIG. 7b. The RMS spatial structure function of the temporal analyses of the Smith product (units: mm/month). The bottom (black) is for the Smith anomalies, it is the analogue of the GHCN analysis (Fig. 7a), these approximate the joint space-time structure function $S(\Delta x, \Delta t)$. Once again, the curves for different averaging times (top to bottom) are roughly equally spaced as predicted by Eqs. 29, 33: the reference lines are spaced at factors $4^{-0.4}$ as in Fig. 7a (corresponding to $H_t = -0.4$, and a factor 4 of temporal averaging, the same temporal scaling as the GHCN). However, the slope is very different: $\xi_x(2)/2 \approx +0.4$ (hence $H_t \approx 0.25$) rather than $\xi_x(2)/2 \approx -0.3$ for the GHCN anomaly, see Fig. 7a, so that the Smith anomaly (corresponding to the dashed) has very different spatial statistics than the GHCN data. The top (pink) are the corresponding analyses for the raw rain rates averaged over 1, 2, 4, ... 512 months (top to bottom).
FIG. 7c. The same as fig. 7b but for the 20CR data (1871-2012), and averaging data over all latitudes between ±45°. The dashed lines have the indicated slopes corresponding to \( \xi_x(2)/2 = 0.35 \) and the vertical spacing corresponds to \( \xi_t(2)/2 = -0.45 \) (every factor of 4 in time scale). These are consistent with the estimates in table I, II of \( H_x = 0.2, C_{1x} = 0.15 \) and \( H_t = -0.42, C_{1t} = 0.03 \) (taking \( \alpha \approx 2 \) in both cases). For the anomalies (black, bottom set) the temporal scaling is reasonably well respected up to 128 months, but for averaging over 256, 512 months, the scaling is badly broken, these lines are indicate “256”, “512” are thin. The raw data is reasa

IV. Discussion and conclusions

Over the last decades, there have been numerous scaling analyses of precipitation and other atmospheric fields, so that several fundamental aspects of atmospheric dynamics have been greatly clarified. For example, over time scale ranges from weather scales up to \( \approx 100 \) kyrs (ice-age scales), in between the familiar weather and climate regimes, there is an intermediate macroweather regime, roughly spanning the range 10 days to 30 years
(industrial, 100 years or longer, preindustrial). The three regimes alternate in their basic characters. In the weather and climate regimes, average fluctuations tend to grow with scale \((H>0)\), they appear unstable. In contrast, in the macroweather regime they decrease with scale \((H<0)\), they appear stable. A recent analysis [Lovejoy, 2014] finds that this alternation continues through two larger scale (macroclimate, megaclimate) regimes out to time scales of over 500 Myrs.

Scaling analyses of macroweather precipitation have mostly used monthly station data and the analyses have often used difficult to interpret statistical methods (such as the Detrended Fluctuation Analysis technique), and this has hindered the emergence of a clear overall picture of annual, decadal and centennial scale precipitation variability. Surprisingly, with the exception of dLL, there have been no attempts to characterize the spatial macroweather variability, nor – the focus of this paper – the more fundamental joint space-time macroweather variability needed to construct macroweather models (in contrast, there have been many spatial scaling analyses of precipitation in the weather regime).

Due to the extreme precipitation variability (high multifractal intermittency), the statistics have nonclassical behaviors. An important (near) exception is the temporal macroweather variability that has a small \(C_1\) (this characterizes the intermittency near the mean) and is therefore not too far from being quasi-Gaussian, see Table II (although the extremes are apparently quite non-Gaussian: power laws). Indeed both at weather and macroweather scales of all the standard atmospheric fields, precipitation has the largest \(C_1\). This intermittency combined with the long range statistical dependencies implied by the scaling has seriously hampered attempts by conventional analysis methods (such as the
decadal trend estimates in the IPCC AR4) to demonstrate the existence of anthropogenic increases in precipitation, and this in spite of simple and direct physical, theoretical connections between increased temperatures and increased precipitation. Since the conventional station precipitation series such as the Global Historic Climate Network (GHCN) product are for land only, in addition to this, we also studied the globally complete Twentieth Century Reanalysis product (20CR, [Compo et al., 2011]) and the satellite based ([Smith et al., 2012]) global precipitation product.

The global space-time scaling up to planetary scales and lifetimes of planetary structures (fig. 3) allows the atmosphere to be modeled up to those scales using the stochastic Fractionally Integrated Flux (FIF) model. This success suggests its extension to much longer time scales, the Extended FIF (EFIF). The long time properties of EFIF ($\tau_\tau \tau_\tau_\tau$) (worked out in detail in appendix 10A of [Lovejoy and Schertzer, 2013]), are that one generically obtains (rough) temporal macroweather exponents $H$ in the range $-0.5<H<0$ (especially in the range $-0.4<H<-0.2$). The simplest model outlined in section III.C also predicts that a statistical space-time macroweather factorization property should generally hold, a property that was verified on temperature data in Lovejoy and Schertzer 2013. In a recent paper ([Lovejoy et al., 2015]) we showed how single macroweather time series (using the example of the mean global scale temperature) can be modeled using fractional Gaussian noise which is the simplest relevant scaling process with $H<0$ (its order one integral is the more familiar fractional Brownian motion); the resulting model was called the Scaling Linear Macroweather (SLIM) model. In section III.D as an approximation to EFIF (notably respecting space-time factorization), we therefore derived the spatial extension of SLIM to the case where $H_z$ is spatially uniform, this space-time SLIM model
involves temporal fGn processes at each spatial location with amplitudes spatially correlated according to a very low frequency spatial multifractal climate process.

In the spectral domain, the space-time factorization implies that the spectral density approximately verifies \( P_{xy}(k_x,k_y,\omega) = P_{xy}(k_x,k_y) P_t(\omega) \), a property that we demonstrated on the 20CR precipitation data. For simplicity considering only latitudinal variations (wavenumbers \( k_x \)) we found that over ranges of \( k_x, \omega \) where \( P_{xt}(k_x,\omega) \) varies by a factor \( 10^7 \), the ratio \( P_{xt}(k_x,\omega) / (P_x(k_x) P_t(\omega)) \) varies by typically \( \pm 20\% \). In real space, factorization predicts that joint structure functions should also decompose into separate spatial and temporal factors, a property that we confirmed on all three data sets. Physically, the interpretation is that spatial variations in macroweather statistics are controlled by different “climatic zones” which modulate the otherwise qualitatively similar (scaling) temporal variability. This property is already implicitly widely used for example when local climate states are “homogenized” by nondimensionalizing their statistics (including probabilities) by using standard deviations of local anomaly fluctuations.

The work described here clarifying and modeling space-time macroweather precipitation variability is a necessary step not only for understanding macroweather precipitation and the limitations of the corresponding precipitation products, it is also a necessary first step in the stochastic forecasting of macroweather fields. The potential of such forecasts is large since the temporal scaling implies that in macroweather there are strong long range memories that can potentially be exploited. This promises to overcome many of the limitations of conventional (deterministic) GCM climate forecasts.

ACKNOWLEDGEMENTS
We thank A. Bussy for help with the analysis of the Smith product. This work was unfunded, there are no conflicts of interest.

REFERENCES


Van der Hoven, I. (1957), Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour, *Journal of Meteorology, 14*, 160-164.
