Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Is the Weather Over Nigeria Chaotic? A Calibrated Surrogate- Data Audit of Daily Tropical Meteorological Series From Two Independent Data Products

Iyanuoluwa O. Ajayi

ajayi.iyanuoluwa@bouesti.edu.ng

Bamidele Olumilua University of Education, Science and Technology

Temitope Y. Ojebisi

Bamidele Olumilua University of Education, Science and Technology

Research Article

Keywords: deterministic chaos, surrogate data, Lyapunov exponent, nonlinear prediction, tropical meteorology, West Africa, reanalysis, hypothesis testing

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11669477/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

IS THE WEATHER OVER NIGERIA CHAOTIC? A CALIBRATED SURROGATE-DATA AUDIT OF DAILY TROPICAL METEOROLOGICAL SERIES FROM TWO INDEPENDENT DATA PRODUCTS

Iyanuoluwa O. Ajayi, Temitope Y. Ojebisi

Department of Physics, Bamidele Olumilua University of Education, Science and Technology, Ikere-Ekiti, Nigeria

Email: (*ajayi.iyanuoluwa@bouesti.edu.ng, ajayiiyanu16@gmail.com), ojebisi.temitope@bouesti.edu.ng*

ORCID NO.: 0009-0008-6399-0167(Ajayi)

Corresponding Author: Iyanuoluwa O. Ajayi (*ajayi.iyanuoluwa@bouesti.edu.ng, ajayiiyanu16@gmail.com)*

Abstract

Claims of low-dimensional deterministic chaos in daily weather series over Nigeria usually rest on a positive Lyapunov exponent or a finite correlation dimension, without a calibrated null distribution. We audit such claims with surrogate-data hypothesis testing on 42 years (1984–2025, N = 15 341 days) of daily temperature, humidity, precipitation, wind speed, radiation, dew point and pressure at ten sites from the Sahel to the Guinea coast, in two independent data products (NASA POWER/MERRA-2 and ECMWF ERA5), validated against NOAA station records. Three null models (including a periodically forced linear stochastic process), the Rosenstein largest Lyapunov exponent, a nonlinear prediction error, permutation entropy and statistical complexity, a multi-scale analysis (1–30 d) and known-answer controls were used; the empirical size and power of the tests were measured. The prediction-error test detected chaotic controls (reduction of 92–100%, 100% power) with a false-positive rate of 3.3%. No weather series (100 series, both products) was flagged as chaotic, and no series satisfied all four convergent criteria. Without surrogates, 83–89% of defensible analysis settings returned a positive exponent, equally for observed series and for non-chaotic surrogates; the north–south gradient of the exponent was reproduced by the surrogates. The weak prediction-error structure that remained (≤ 8%) disappeared after removing the seasonal modulation of variance. We conclude that daily weather series over Nigeria are, within the sensitivity of these tests, indistinguishable from forced linear stochastic processes, and we provide a calibrated protocol for future claims.

Keywords: deterministic chaos; surrogate data; Lyapunov exponent; nonlinear prediction; tropical meteorology; West Africa; reanalysis; hypothesis testing

1. Introduction

Whether the atmosphere is chaotic is not in doubt: Lorenz (1963) showed that a three-variable truncation of the convection equations already has sensitive dependence on initial conditions, and Lorenz (1969) argued that the multi-scale nature of the atmosphere limits deterministic predictability in practice. A different and much more fragile question is whether

the chaos of a *particular measured series*, for example the daily temperature at one station, can be identified from that series alone as a low-dimensional deterministic attractor. The reconstruction theorem of Takens (1981) and the algorithms built upon it (correlation dimension, Lyapunov exponents, nonlinear prediction) made it tempting, from the 1980s onward, to answer this question with a number: a finite correlation dimension or a positive Lyapunov exponent was read as the fingerprint of a 'climatic attractor' (Nicolis and Nicolis 1984; Grassberger 1986; Tsonis and Elsner 1988).

The critique of that practice is equally old. Osborne and Provenzale (1989) showed that a purely stochastic process with a power-law spectrum yields a finite, apparently low correlation dimension, and Provenzale et al. (1992) and Theiler et al. (1992) established the logical remedy: a diagnostic computed on data is only informative relative to its distribution under an explicit null hypothesis, which is built by Monte-Carlo generation of *surrogate* series that share the linear properties of the data but are otherwise random (Schreiber and Schmitz 1996, 2000; Lancaster et al. 2018). Reviews of chaos theory in the geosciences have repeatedly warned about short, noisy, quantised and non-stationary records and about the sensitivity of estimators to embedding and fitting choices (Sivakumar 2004). Despite this, diagnostics are still frequently reported without any null distribution, or with a very small number of surrogates and without a control showing that the test can tell a chaotic from a non-chaotic series.

The question has regional relevance. Daily meteorological series over Nigeria and the wider West African monsoon belt are used to argue for 'low-dimensional deterministic chaos' and for north–south gradients in the degree of chaos. The most recent and most extensive example is Echi et al. (2025), who analysed daily MERRA-2 precipitation, wind speed, air temperature and all-sky radiation (1982–2020) at stations in all geo-climatic zones of Nigeria, removed the seasonal cycle by 365-day differencing, and reported positive Rosenstein Lyapunov exponents (0.001–0.28 d^{-1} depending on variable and site), finite correlation dimensions (about 3.5–4.3 for temperature, wind and radiation), and spatial gradients of the exponent with latitude (higher in the north for precipitation and temperature, higher in the south for wind and radiation), concluding that the data show low-dimensional deterministic chaos across Nigeria. That study did include a surrogate test, but it was applied to the correlation dimension only, with five amplitude-adjusted Fourier-transform surrogates and a significance statistic for which no threshold or values are reported; the Lyapunov exponents were not compared with surrogate distributions, no known-answer control system was used to show what the pipeline returns for a non-chaotic series, and the fitting range of the exponent is documented only for one example. Related work on Nigerian climate has used entropy, Hurst exponents and recurrence quantification (Fuwape et al. 2018, for ERA-Interim extremes), again without a calibrated null.

Earlier work on nonlinear dynamics has shown how strongly nonlinear diagnostics depend on the scale and on the method of analysis, both in model systems where the chaotic regime can be established from the equations (Ajayi 2026, on a six-dimensional memcapacitor system) and in measured space-plasma series, where coupling between variables is scale-dependent and surrogate tests are required to separate nonlinear structure from linear correlation (Adelakun et al. 2026). The contrast is instructive. When the governing equations are known, a positive Lyapunov exponent can be computed from the model and the chaotic regime is established independently of any single time series. For a measured meteorological series the equations are unknown, the exponent must be *estimated* from one noisy realisation, and the claim therefore stands or falls with the statistical evidence that the estimate differs from what a non-chaotic process would give.

This paper provides that evidence for Nigerian daily weather. We do not propose a new chaos detector. The contribution is a systematic, pre-specified and calibrated application of established surrogate-data methodology to the type of series and diagnostics used in the literature we audit, with the following components: (i) ten sites spanning the Sahel to the Guinea coast and five core variables (temperature, relative humidity, precipitation, wind speed, solar radiation) over 42 years (N = 15 341 days), extended by four further variables (maximum and minimum temperature, dew point, surface pressure) at five sites; (ii) three null models, including a physically motivated one in which the annual cycle is treated as deterministic and only the anomaly is randomised; (iii) two diagnostics with complementary properties, the Rosenstein largest Lyapunov exponent and a zeroth-order nonlinear prediction error; (iv) known-answer controls, with an empirical study of the false-positive rate and the power of the tests; (v) a replication on an independent data product (ECMWF ERA5) and a validation of both products against station records; (vi) a 'multiverse' of defensible analysis choices that quantifies how often an uncalibrated analysis returns a positive exponent; (vii) a replication of the pre-processing used by the audited study (seasonal differencing); (viii) a multi-scale audit of 3-, 7- and 30-day means; and (ix) an independent ordinal-pattern diagnostic (permutation entropy and statistical complexity). All departures from the pre-specified protocol are disclosed.

The paper addresses four questions. **Q1:** Do the diagnostics, applied to real series, differ from those of non-chaotic surrogates? **Q2:** Is any such difference robust to the choice of null, the data product and the analysis settings? **Q3:** Does the calibration on controls show that the procedure would have detected chaos if it were present? **Q4:** Do the north–south gradients of the exponent reported in the literature survive when the exponent is compared with its null?

2. Data

2.1 Study sites and variables

Ten sites were selected to span the four principal climatic zones of Nigeria, from the Sahel (Sokoto, Maiduguri, Kano) through the Guinea savanna (Abuja, Ilorin) and the forest-savanna transition (Ibadan, Akure, Enugu) to the Atlantic coast (Lagos, Port Harcourt) (Table 1; Fig. 1). Five variables were analysed at each site: 2 m air temperature (T2M), 2 m relative humidity (RH2M), precipitation (PRECTOTCORR), 10 m wind speed (POWER WS2M, wind speed at 2 m; ERA5 wind_speed_10m, at 10 m, so absolute values differ between products and only dynamical properties are compared) and all-sky downward shortwave radiation (ALLSKY_SFC_SW_DWN; ERA5 shortwave_radiation_sum). These are the four variables of the audited study plus humidity, which is physically linked to temperature and precipitation. Maximum and minimum temperature, dew point and surface pressure were also downloaded, quality-checked and analysed in an extended audit.

Table 1. Study sites (coordinates as requested from the data providers) with climatological means of the NASA POWER series, 1984–2025.

Site	Lat. (°N)	Lon. (°E)	T (°C)	RH (%)	Precip. (mm/yr)	Wind (m/s)	SW (MJ/m²/d)	Days
Sokoto	13.06	5.24	27.7	44.1	703	2.27	21.7	15341
Kano	12.0	8.52	26.1	45.4	739	0.84	21.7	15341
Maiduguri	11.85	13.16	28.2	39.2	560	2.69	21.6	15341
Abuja	9.07	7.4	25.2	70.8	1381	1.54	19.3	15341
Ilorin	8.5	4.55	25.4	75.0	1280	1.99	18.3	15341
Ibadan	7.38	3.9	25.1	84.9	1706	1.56	16.9	15341
Akure	7.25	5.19	24.5	83.2	1649	1.33	17.2	15341
Lagos	6.52	3.38	26.3	84.7	1491	2.28	16.8	15341
Enugu	6.44	7.5	25.3	82.2	1560	1.77	17.2	15341
Port Harcourt	4.82	7.05	25.6	87.4	2477	0.92	15.1	15341

Precipitation: mean annual total; wind speed: POWER WS2M at 2 m.

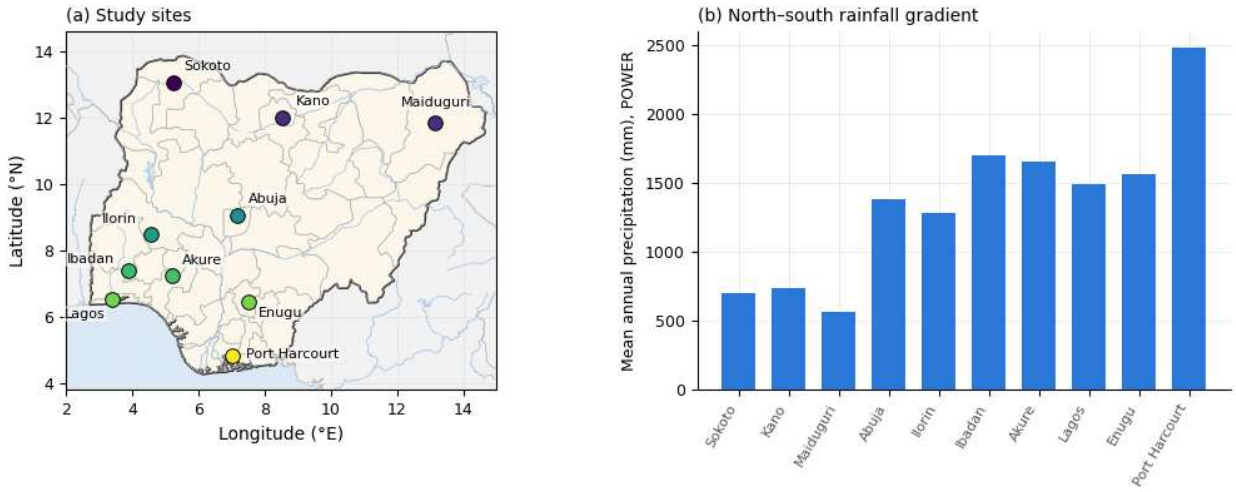


Fig. 1. (a) Locations of the ten sites (the coast of the Gulf of Guinea lies just south of Lagos and Port Harcourt). (b) Mean annual precipitation of the sites ordered by latitude (NASA POWER, 1984–2025).

2.2 Data

Primary Data: NASA POWER. Daily point series were retrieved from the NASA Prediction of Worldwide Energy Resources (POWER) daily API (community AG; https://power.larc.nasa.gov) for 1 January 1981 to 31 December 2025. POWER meteorological parameters are derived from the MERRA-2 reanalysis (Gelaro et al. 2017; native grid 0.5° × 0.625°) and surface solar radiation from CERES-based products (1° grid). This is the same product family as in the audited study, which makes the comparison direct.

Replication Data product: ECMWF ERA5. ERA5 (Hersbach et al. 2020; native 0.25° grid) daily aggregates were retrieved for the same coordinates through the Open-Meteo Historical Weather API (https://archive-api.open-meteo.com; model 'era5'; daily time zone Africa/Lagos) for 1984–2025. Open-Meteo returns the series of the nearest ERA5 grid cell (grid-cell coordinates are stored in the metadata files). ERA5 is produced by a different model and assimilation system from MERRA-2, so agreement between the two products is not guaranteed by common processing. The first download (2026-11-07, 21:05 UTC) was truncated for Ibadan and Akure by the API rate limit (HTTP 429); the two sites were re-downloaded with a rate-limit-aware script (2026-11-08, 06:50 UTC), and all ten sites are complete (no missing day, all nine variables).

Validation product: NOAA GHCN-Daily. Station records (Menne et al. 2012; https://www.ncei.noaa.gov/pub/data/ghcn/daily/) were retrieved for the Nigerian stations NIM00065046 (Kano), 065082 (Maiduguri), 065101 (Ilorin), 065201 (Lagos), 065250 (Port Harcourt) and five further stations (2026-11-07, 19:53 UTC). Values failing NOAA quality-control flags were removed, and TMAX/TMIN/PRCP were screened for physically impossible values. These records are far too sparse for phase-space analysis (the longest run

of consecutive observed days is 28 d), so they serve only to validate the reanalysis products and were not used in the audit.

Controls. Known-answer systems (logistic map, Lorenz-63, AR(1) noise, a sinusoid plus noise, exp(AR(1)), white noise) were generated numerically. They are used exclusively to calibrate the procedure and are always labelled as controls; no synthetic series enters any of the empirical results on weather.

Table 2. Data products and their role in the study

Product	Source / native grid	Variables	Period	Retrieved (UTC)	Role
NASA POWER (community AG)	MERRA-2 meteorology (0.5° × 0.625°); CERES-based surface solar (1°)	T2M, T2M_MAX, T2M_MIN, T2MDEW, RH2M, PRECTOTCORR, WS2M (2 m), PS, ALLSKY_SFC_SW_DWN	1981–2025 (analysed 1984–2025)	2026-11-07 19:31	Primary series
ECMWF ERA5 via Open-Meteo (model era5)	ERA5 reanalysis, 0.25°; daily aggregation, time zone Africa/Lagos	temperature (mean, max, min), dew point, relative humidity, precipitation, wind speed at 10 m, surface pressure, shortwave radiation sum	1984–2025	2026-11-07 21:05 (8 sites); 2026-11-08 06:50 (Ibadan, Akure)	Independent replication
NOAA GHCN-Daily (by-station files)	Station observations; NOAA quality flags removed	TMAX, TMIN, PRCP at 10 Nigerian stations	1984–2025 (sparse)	2026-11-07 19:53	Validation of the products
Known-answer controls	Logistic map, Lorenz-63, AR(1), annual sinusoid + noise, exp(AR(1)), white noise (numerical)	—	N ≈ 15 000 (6 000 in the size study)	—	Calibration only (never presented as weather results)

2.3 Quality control and pre-processing

The analysis window is 1 January 1984 – 31 December 2025 (N = 15 341 days) for all variables and products, because POWER solar radiation is structurally absent before 1984. In the window the POWER files contain one missing value in total (Abuja radiation, 15 March 1991), which was linearly interpolated and flagged; no other value was altered or removed. Series were checked for range (physical plausibility), constant runs, sign (precipitation ≥ 0), duplicated dates and continuity of the daily index (assertions in the cleaning code).

Descriptive statistics, the fraction of variance in the seasonal cycle and trend, the lag-1 autocorrelation, and the fraction of zero days are given for every series in the supplementary quality-control tables.

Two derived versions of every series are used. The **raw** series is the product as downloaded. The **anomaly** series is the residual after ordinary-least-squares removal of a linear trend and four annual harmonics (period 365.2425 d):

$$a(t) = x(t) - [\, \beta_0 + \beta_1 t + \Sigma_{k=1}^{4} (\gamma_k \cos(k\omega t) + \delta_k \sin(k\omega t)) \,], \quad \omega = 2\pi / 365.2425 \text{ d} \qquad (1)$$

The fitted component $f(t) = x(t) - a(t)$ is the 'seasonal cycle plus trend'. Both versions are analysed because the raw series is what is often analysed in practice, whereas the anomaly is the object of a clean dynamical question once the forced (astronomical) cycle is removed. The audited study used seasonal differencing, $x(t) - x(t - 365)$; this variant is analysed separately. Fig. 2 illustrates the decomposition for one series. Quantisation is a feature of the products (ERA5 temperature to 0.1 °C, humidity to 1 %, precipitation to 0.1 mm) and is kept; ties created by quantisation are handled.

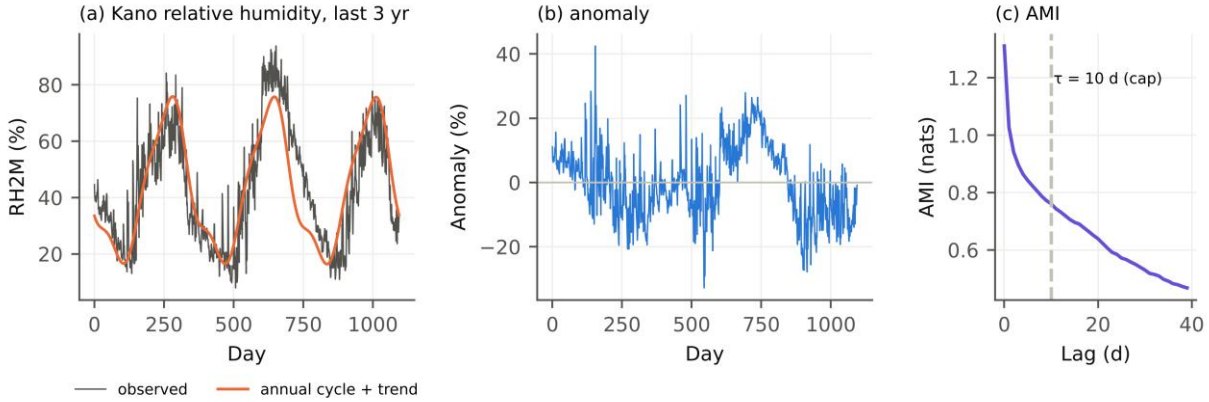


Fig. 2. Pre-processing of a daily series (NASA POWER relative humidity at Kano, last three years): (a) observed series and fitted annual cycle plus trend; (b) anomaly; (c) average mutual information with the selected delay (capped at 10 d).

3. Methods

3.1 Overview and protocol

The analysis follows a protocol that was fixed before the full audit was run (supplementary file PROTOCOL.md). In brief, for every series, diagnostics are computed on the data and on B = 99 surrogates generated under each null, with embedding parameters held identical, and the observed value is ranked against the surrogate distribution (Fig. 3).

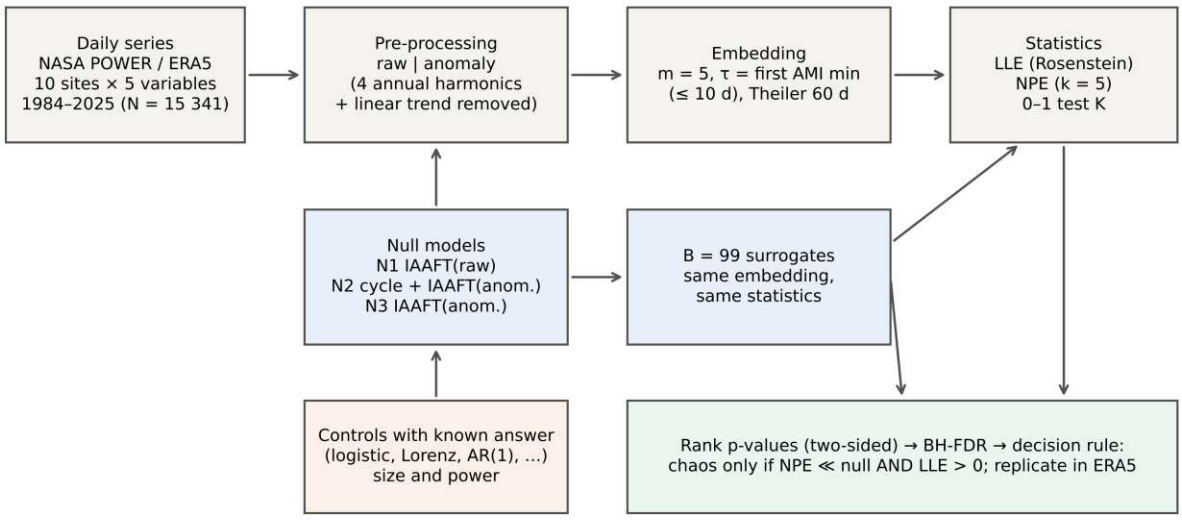


Fig. 3. Audit Workflow: pre-processing, delay embedding, diagnostics computed on the data and on surrogates generated under each null, calibration on known-answer controls, and the decision rule.

3.2 Delay embedding

A scalar series x(t) is mapped to delay vectors $\mathbf{x}(t) = [x(t), x(t+\tau), \ldots, x(t+(m-1)\tau)]$ (Takens 1981). The delay τ is the lag of the first local minimum of the average mutual information (Fraser and Swinney 1986; 32 equal-width bins, lags 1–40 d), capped at 10 d, which is already long relative to the synoptic correlation time. The embedding dimension was fixed at $m = 5$ for all series so that results are comparable across series and products; false-nearest-neighbour fractions (Kennel et al. 1992) were used as a diagnostic of this choice, and $m \in \{3, 5, 8\}$ is varied in the multiverse analysis. Pairs of delay vectors closer than a Theiler window of 60 d in time were excluded from all neighbour searches (Theiler 1986), which removes neighbours that are close only because of temporal autocorrelation.

3.3 Diagnostics

Largest Lyapunov exponent (LLE). The algorithm of Rosenstein et al. (1993) was used. For each of up to 2 500 reference points the nearest neighbour outside the Theiler window is found and the logarithm of the distance between the two trajectories, averaged over reference points, is followed over s = 0…20 d:

$$y(s) = \langle \ln \| \mathbf{x}(i+s) - \mathbf{x}(j(i)+s) \| \rangle_i \qquad (2)$$

The LLE is the slope of the least-squares line through y(s) over a fit window. The primary window is s = 1–10 d; 1–5 d and 5–20 d are sensitivity windows. Units are d^{-1} for the weather series. For a chaotic system y(s) has a linear section whose slope is the exponent; for a noisy or stochastic series y(s) typically rises quickly and then saturates, so the slope depends strongly on the window, which is precisely why a null distribution is needed.

Table 3. Protocol amendments made after the protocol was fixed

ID	Change	Reason	Effect
A1	Decision rule: NPE lower than null (FDR) and LLE > 0, instead of 'LLE higher than surrogates'	Chaotic controls have a lower Rosenstein exponent than their surrogates	None: zero flags under both rules
A2	A post-hoc null with seasonally modulated noise was discarded	Distorted the autocorrelation of the anomaly	None; results not used
A3	Seven series with nominal $p \leq 0.05$ re-run with B = 199	Check of borderline series	p-values are optimistic (selected on the outcome)
A4	Rank-based FDR has no power at B = 99; empirical size, direction and z-score FDR added	Resolution limit 2/(B+1)	Interpretation relies on size, direction and convergence
A5	ERA5 Ibadan and Akure downloaded later (rate limit)	HTTP 429 truncation	Audit run later with identical code
A6	Extended analyses (extra variables, multi-scale, seasonal-variance diagnostic, larger size study, station validation) added	Broader coverage	Settings fixed beforehand; all results reported
A7	Permutation entropy: deterministic tie-break replaced by random tie-break	Bias for series with runs of tied values	Biased output discarded and archived

Nonlinear prediction error (NPE). One-step-ahead prediction by the zeroth-order local method (Farmer and Sidorowich 1987; Casdagli 1989; Sugihara and May 1990): the next value of the series is predicted as the mean of the successors of the k = 5 nearest neighbours (outside the Theiler window) of the current delay vector. The series is standardised, so the error is expressed as RMSE/SD (1 = no better than predicting the mean). A deterministic low-dimensional system is more predictable from its own recent state than a linear stochastic process with the same spectrum, so a significantly *smaller* NPE than the surrogates is evidence for nonlinear determinism. 2 500 reference points were used for the weather series.

0–1 test for chaos (K). The Gottwald–Melbourne test (Gottwald and Melbourne 2004), regression/correlation implementation, median over 100 random frequencies $c \in (\pi/5, 4\pi/5)$

for observed series and 30 for the first ten surrogates of each run, is reported only as a diagnostic of its own discriminating power. As anticipated by Gottwald and Melbourne (2005) for noisy data, $K \approx 1$ for all series, including white noise, and it plays no role in the inference.

3.4 Surrogate (null) models

Surrogates are generated with the iterative amplitude-adjusted Fourier transform (IAAFT; Schreiber and Schmitz 1996), which preserves the amplitude distribution exactly and the power spectrum approximately (up to 300 iterations, stopped when the rank order is stable). Before Fourier-based surrogates are generated, the series is trimmed by up to 365 d at each end so that the end points match in value and in slope, which limits the spurious low-frequency power created by the implicit periodic extension (Theiler et al. 1992; Schreiber and Schmitz 2000). Three nulls are used:

a) **N1 (raw:IAAFT).** The raw series is replaced by an IAAFT surrogate. The null hypothesis is a static monotonic transformation of a linear Gaussian process, including a Gaussianised seasonal peak in the spectrum. This is the textbook null.
b) **N2 (raw:seasonal), the main null.** surrogate = f(t) + IAAFT(a(t)), i.e. the fitted annual cycle and trend (kept fixed) plus a linear-stochastic surrogate of the anomaly. The null hypothesis is a *periodically forced linear stochastic process*, which is the natural 'no chaos' description of the weather series once the forced annual cycle is treated as deterministic. N2 retains the seasonal waveform exactly, which N1 does not.
c) **N3 (anom:IAAFT).** The deseasonalised anomaly is replaced by an IAAFT surrogate. The null is a linear stochastic process for the anomaly.

For each series and null, B = 99 surrogates were generated, each with the same embedding parameters as the data. The diagnostics of the data are ranked among those of the surrogates. The two-sided rank p-value is

$$p = \min\{1, 2 \cdot \min[(1 + \#\{s \geq q\})/(B + 1), (1 + \#\{s \leq q\})/(B + 1)]\} \quad (3)$$

(Theiler et al. 1992; Davison and Hinkley 1997; Phipson and Smyth 2010), whose smallest attainable value is $2/(B + 1) = 0.02$. We also report the direction of the departure (observed value above or below the surrogate mean) and the surrogate z-score, $z = (q - \text{mean}(s)) / \text{SD}(s)$. Because the direction matters scientifically (a smaller NPE means more determinism, a smaller LLE does not), two-sided tests are used throughout.

3.5 Decision rule and multiplicity

A series is flagged as showing evidence of low-dimensional chaos only if (a) its NPE is significantly *lower* than the N2 surrogates after Benjamini–Hochberg false-discovery-rate control at $q = 0.05$ (Benjamini and Hochberg 1995), applied over all series of a data product, (b) the observed LLE is positive. The pre-specified protocol also required the LLE to exceed

the surrogate distribution; this was amended after the controls showed that it fails for true chaos. Because a single B = 99 run cannot produce $p < 0.02$, and BH needs $p \leq 0.05 \cdot k/n$ for the k-th smallest of n tests, the rank-based FDR procedure has essentially no power at this B (amendment A4). We therefore additionally report (i) the number of nominal rejections compared with the number expected by chance and with the empirical false-positive rate measured on controls, and (ii) an FDR procedure based on the normal approximation to the surrogate distribution (z-scores), clearly labelled as approximate.

3.6 'Multiverse' of analysis choices

To quantify how often an uncalibrated analysis returns a positive exponent, the LLE was computed without surrogates for every combination of $m \in \{3, 5, 8\}$, $\tau \in \{1, 3, \tau_{AMI}\}$ d and fit window $\in$ {1–5, 1–10, 5–20} d on raw and anomaly series of all POWER series and, for comparison, on one N2 surrogate (a non-chaotic series by construction) of every raw series. The share of settings with LLE > 0 is reported for the observed raw series, the observed anomaly series and the surrogates.

3.7 Controls, size and power

The same code was applied to systems with known behaviour (all with $N \approx 15\,000$ and B = 99): the logistic map at r = 4 (chaotic, LLE = ln 2 per iteration), the x-component of Lorenz-63 ($\sigma = 10$, $\rho = 28$, $\beta = 8/3$; RK4, dt = 0.01, sampled every 0.05 time units), the same Lorenz series with additive Gaussian noise of 50% of its standard deviation, AR(1) noise ($\varphi = 0.9$), a sinusoid with annual period plus white noise, exp(AR(1)) (a static nonlinear transformation of linear noise, non-chaotic) and white noise. Embedding parameters for the controls are given in the supplement. The empirical false-positive rate (size) of each test at $\alpha = 0.05$ was estimated on repeated realisations of three non-chaotic processes (AR(1) with $\varphi = 0.9$, sinusoid + AR(1), exp(AR(1))) and its power on the logistic map (N = 6 000, B = 39 surrogates per realisation; 50 realisations per process; Clopper–Pearson 95% intervals).

3.8 Extended variables

Four further variables (maximum 2 m temperature T2M_MAX, minimum 2 m temperature T2M_MIN, 2 m dew-point temperature T2MDEW and surface pressure PS) were processed identically (same window, same decomposition, same null N2, B = 99) at five sites chosen to span the latitudinal range (Sokoto, Kano, Abuja, Lagos, Port Harcourt), in both products. Only the main null N2 was used for these series.

3.9 Multi-scale audit

Diagnostics of a nonlinear system depend on the time scale at which it is observed, and dependence on scale is a theme of our earlier work (Adelakun et al. 2026). Non-overlapping block means over s = 3, 7 and 30 d were formed from the raw series and from the fitted

cycle-plus-trend, the anomaly of the aggregated series being their difference (N = 5 113, 2 190 and 511 samples). The N2 test was repeated on each aggregated series (B = 99), with m = 5, τ the first minimum of the average mutual information capped at ⌈10/s⌉ samples, a Theiler window of ⌈60/s⌉ samples (minimum 2), an LLE fit window of 1–5 samples and 2 500 reference points (or half the series length). Aggregation by averaging acts as a low-pass filter that removes much of the daily weather noise, so it increases the signal-to-noise ratio of any slow deterministic component at the cost of fewer samples.

3.10 Permutation entropy and statistical complexity

As an independent diagnostic that does not rely on neighbour searches, the ordinal-pattern distribution of the raw series was analysed (Bandt and Pompe 2002): for pattern order d = 5 and delays of 1 and 3 d, the normalised Shannon entropy H of the 5! = 120 ordinal patterns and the Jensen–Shannon statistical complexity C (Rosso et al. 2007) were computed. White noise has H = 1 and C = 0; low-dimensional chaotic maps combine intermediate H with high C; stochastic processes with long memory occupy the high-H, low-C region. H and C were compared with N2 surrogates (B = 99; two-sided rank p-value). Ties (identical values, frequent in quantised and zero-inflated series) were broken at random by adding a jitter of 10^{-9} of the range to the series; a first implementation with a deterministic tie-break was found to impose a fixed ordinal pattern on every run of tied values and was discarded (amendment A7).

The anomaly was also divided by a seasonal scale (two annual harmonics fitted to the absolute anomaly) and the standardised anomaly was tested against IAAFT surrogates of itself, to check whether weak prediction-error reductions under N2 stem from the annual modulation of the variance, which N2 does not reproduce. **Seasonal differencing.** The pre-processing of the audited study, $x(t) - x(t - 365\ \mathrm{d})$, was applied to the POWER series, which were then tested against IAAFT surrogates of the differenced series. **Latitudinal gradients.** Spearman rank correlations with latitude were computed for the observed exponent, the mean surrogate exponent, the exponent z-score and the prediction-error reduction. **Station validation.** POWER and ERA5 were compared with GHCN-Daily on paired days. **Monte-Carlo uncertainty.** The variability of the observed NPE and LLE with respect to the random selection of reference points was estimated from 30 independent draws. All analyses were implemented in Python.

4. Results

Before any weather series was analysed, the estimators were checked against known results (supplementary file validation_B99.log). For the logistic map at r = 4 the Rosenstein estimator returned 0.688 per iteration against the exact value ln 2 = 0.693. For the Lorenz-63 x-series the estimate depends strongly on the fit window: 1.65, 1.51, 0.99 and 0.88 per time

unit for windows of 1–8, 5–15, 10–25 and 15–35 samples, against the true 0.906. Only windows that avoid the short-time transient and the saturation reproduce the true exponent. This window dependence is the first indication of why a slope without a null distribution is hard to interpret. A periodic orbit (logistic r = 3.5) contaminated by a 10^{-8} dither, whose true exponent is non-positive, gave 1.25: once distances are dominated by noise, the log-divergence curve rises and a positive slope is obtained without any sensitive dependence. The 0–1 test returned K = 0.998 for the logistic map at r = 4 and 0.988 for Lorenz-63 (correct) and -0.001 for the periodic orbit (correct), but K ≈ 1 for every noisy control, including white noise (Table 4, column K), so the test is uninformative for noisy sampled data and is not used for inference.

Table 4 and Fig. 4 show the behaviour of the two diagnostics under their nulls (IAAFT, B = 99). For the chaotic controls the prediction error is far below that of the surrogates: 99.9% lower for the logistic map and 92.1% for Lorenz-63, both at the smallest attainable p = 0.02. With 50% additive noise the reduction falls to 6.1%, but is still detected (p = 0.02). In the non-chaotic controls the reduction ranges from -1.8% to 5.5%, with p between 0.20 and 0.56. The NPE test therefore separates chaos from non-chaos in the controls, and its sensitivity limit for a strongly sampled chaotic signal in heavy noise is of order 5% (Fig. 5). The Lyapunov-exponent test behaves differently. For the chaotic controls the observed exponent is *lower* than the surrogate exponent (logistic: 0.69 versus 1.65), because the IAAFT surrogate of a deterministic signal is broadband noise whose nearest-neighbour distances diverge faster. In addition, for exp(AR(1)), which is not chaotic, the LLE test returns p = 0.02. This is not a fluke: in the size study (Table 5), the LLE test rejects the true null in a large fraction of realisations of this process, whereas the NPE test keeps close to the nominal rate.

Table 4. Known-answer controls (N ≈ 15 000; B = 99 IAAFT surrogates; two-sided rank p-values, minimum 0.02). NPE = nonlinear prediction error (RMSE/SD); LLE = Rosenstein largest Lyapunov exponent over the first fit window of each system (per sample); K = 0–1 test statistic.

System	NPE obs.	NPE surr.	NPE reduction (%)	p (NPE)	LLE obs.	LLE surr.	p (LLE)	K
Logistic map, r = 4 (chaotic)	0.001	1.098	99.9	0.02	0.687	1.650	0.02	0.999
Lorenz-63, x (chaotic)	0.020	0.248	92.1	0.02	0.084	0.139	0.02	0.993
Lorenz-63 + 50% noise (chaotic)	0.653	0.695	6.1	0.02	0.045	0.050	0.02	0.998
AR(1), φ = 0.9	0.486	0.477	-1.8	0.20	0.069	0.070	0.44	0.998

Annual sinusoid + noise	0.480	0.487	1.4	0.30	-0.002	0.000	0.02	0.999
exp(AR(1))	0.715	0.756	5.5	0.36	0.055	0.045	0.02	0.998
White noise	1.105	1.095	-0.9	0.56	0.090	0.094	0.14	0.998

NPE reduction = 100 × (mean surrogate NPE − observed NPE) / mean surrogate NPE. Fit windows and embedding parameters for each control are listed in the supplementary file Table_S_control_settings.

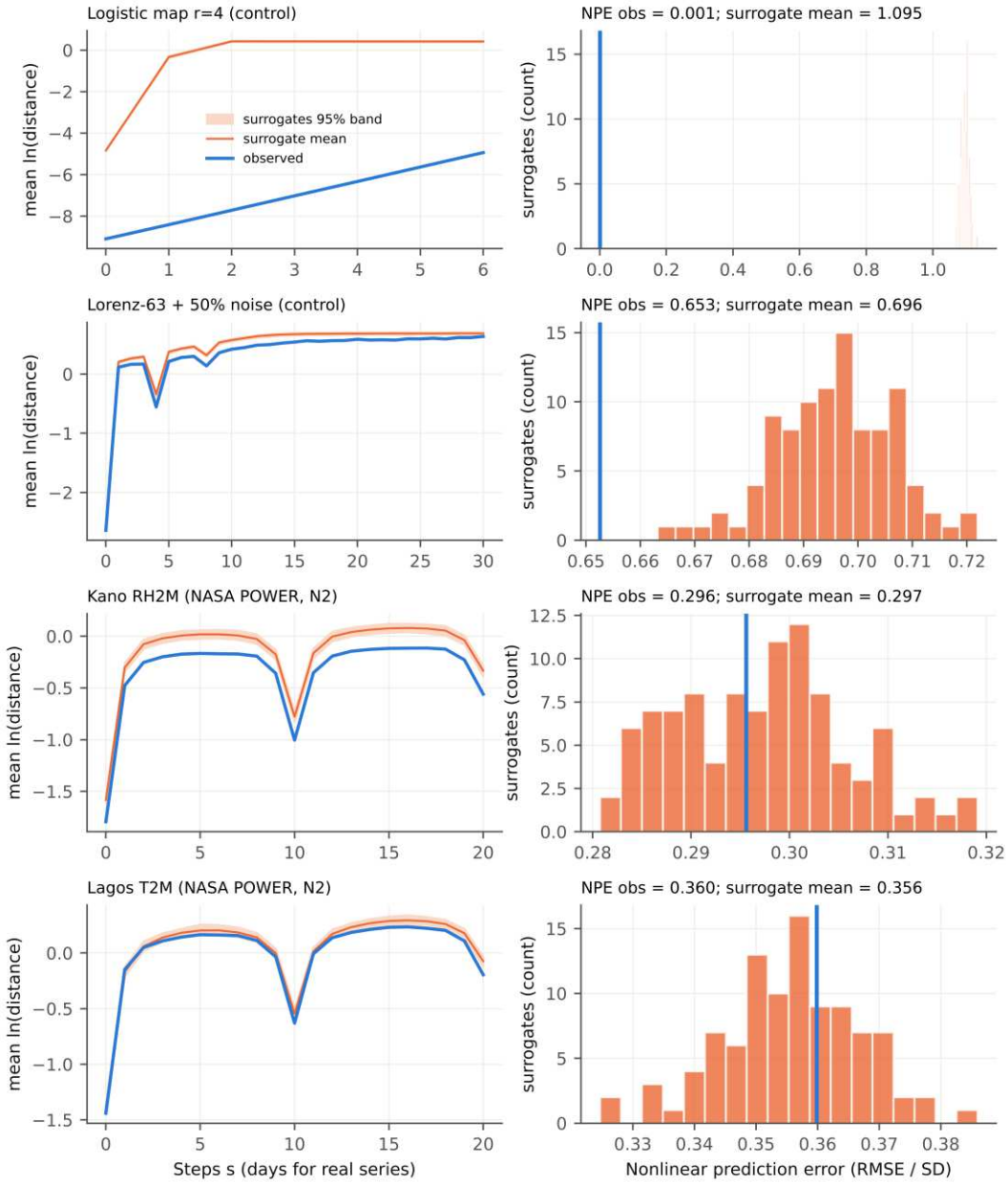


Fig. 4. Worked examples of the two diagnostics. Left: mean log-divergence curves of nearest-neighbour pairs for the observed series (blue) and the 99 surrogates (orange: mean; band: 2.5–97.5 percentile). Right: histogram of the surrogates' nonlinear prediction error with the observed value (blue line). Rows: two chaotic controls, and two real series (NASA POWER Kano relative humidity; Lagos temperature) tested against null N2. For the controls,

the observed series is far more predictable than its surrogates; for the real series the observed value lies within the surrogate distribution.

The empirical size study (Table 5, Fig. 6; 50 realisations per process) quantifies this. For the three non-chaotic processes the NPE test rejects the true null at rates of 0.0–6.0% at $\alpha = 0.05$ (pooled 3.3%; two-sided test with B = 39 surrogates, whose coarse resolution makes it conservative), and its power on the logistic map is 100%. The LLE test rejects the true null for exp(AR(1)) in 74% of realisations (95% CI 59.7–85.4%). The IAAFT null preserves the amplitude distribution but not the dependence structure of the transformed process, and the exponent estimate responds to the marginal skewness. The LLE test is therefore not a valid test of chaos for skewed series, which matters for precipitation.

Table 5. Empirical size and power of the surrogate tests (rejection rate at $\alpha = 0.05$ in repeated realisations; N = 6 000; B = 39 IAAFT surrogates per realisation).

Process	Test	Rejections	Realisations	Rate (%)	95% CI (%)
AR1_phi0.9 (null true)	LLE (Rosenstein)	0	50	0.0	0.0–7.1
AR1_phi0.9 (null true)	NPE (prediction error)	3	50	6.0	1.3–16.5
exp(AR1) (null true)	LLE (Rosenstein)	37	50	74.0	59.7–85.4
exp(AR1) (null true)	NPE (prediction error)	2	50	4.0	0.5–13.7
logistic r=4 (chaos)	LLE (Rosenstein)	50	50	100.0	92.9–100.0
logistic r=4 (chaos)	NPE (prediction error)	50	50	100.0	92.9–100.0
sine+AR1 (null true)	LLE (Rosenstein)	4	50	8.0	2.2–19.2
sine+AR1 (null true)	NPE (prediction error)	0	50	0.0	0.0–7.1

Null true for all processes except the logistic map (chaotic; power). 95% Clopper–Pearson confidence intervals. NPE test: lower or higher than surrogates (two-sided).

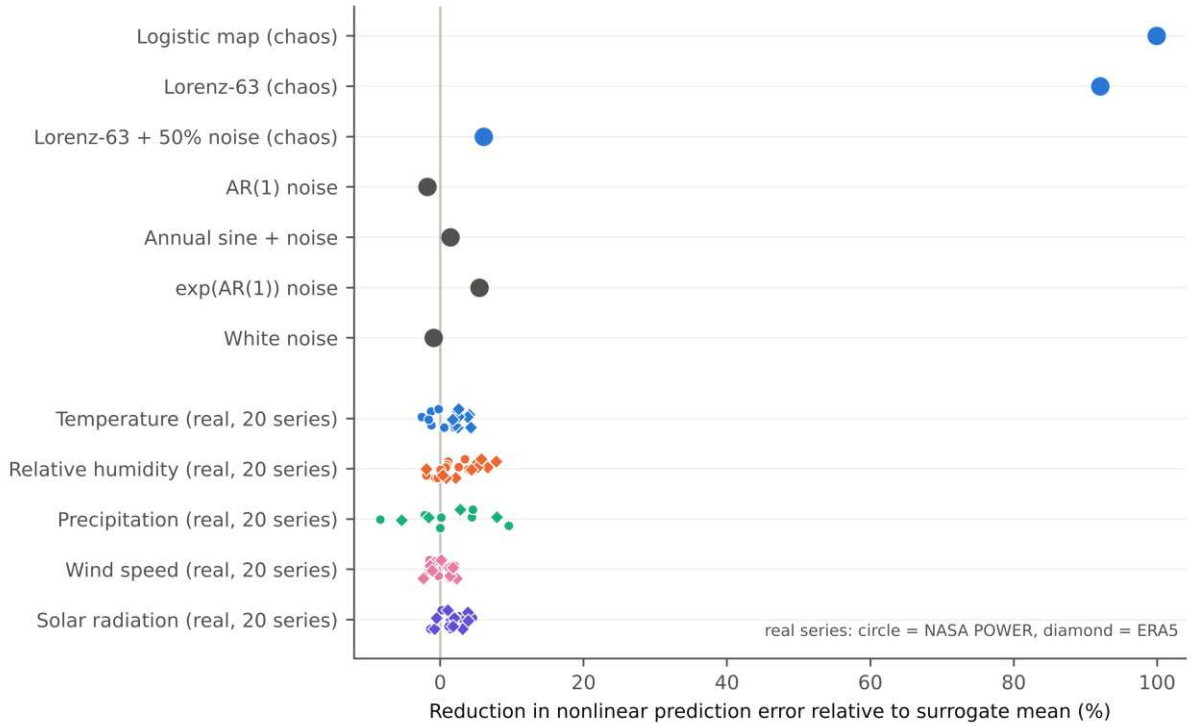


Fig. 5. Reduction of the nonlinear prediction error relative to the surrogate mean, for the known-answer controls (top, blue = chaotic, grey = non-chaotic) and for the 100 real series (NASA POWER circles, ERA5 diamonds; null N2). Chaotic controls lie between 6% and 100%; real series lie between -50% and +10% (negative = less predictable than the null), and all except precipitation are within ±8% of zero.

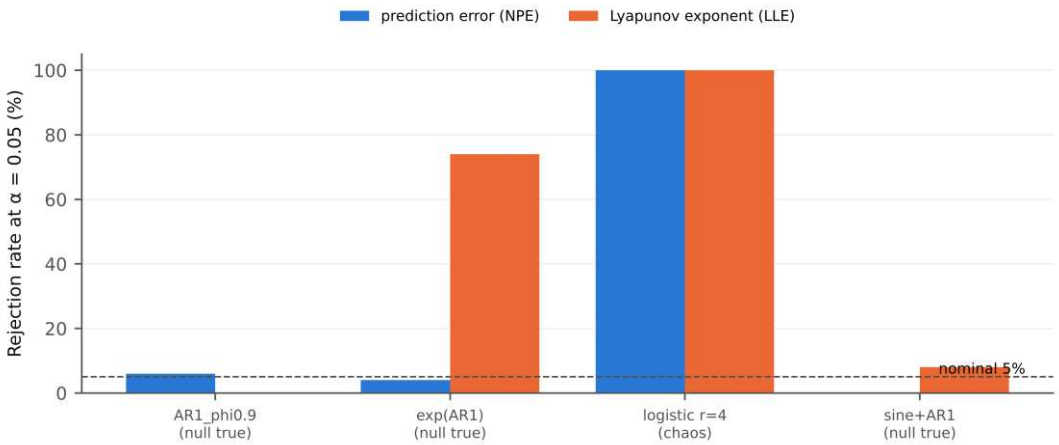


Fig. 6. Empirical rejection rates of the two tests at α = 0.05 for three non-chaotic processes (size; the dashed line is the nominal 5%) and the chaotic logistic map (power).

The embedding delays selected by the average mutual information were 10 d (the cap) for temperature, relative humidity and solar radiation in most series, and 2–10 d for precipitation and wind (Table 6), reflecting the strong persistence of the former and the intermittent character of the latter. Under the main null N2, and in the primary fit window (1–10 d), the observed LLE was positive in 6 of 50 POWER series (solar radiation 2, precipitation 1, wind speed 3) and 13 of 50 ERA5 series (precipitation 10, wind speed 3) (Table 7); temperature

and humidity exponents are negative in every series of both products. Nominal (unadjusted) rejections of N2 by the prediction-error test in the direction of *more* determinism were 2 of 50 (POWER) and 5 of 50 (ERA5), against 2 and 5 in the direction of *less* determinism. The empirical size of the two-sided NPE test at $\alpha = 0.05$ was 3.3% (Table 5), so about 1.7 rejections out of 50 are expected by chance, 2.5 at the nominal level. The POWER total (4) is consistent with that expectation (binomial $P(X \geq 4) = 0.24$ at $p = 0.05$), whereas the ERA5 total (10) exceeds it ($P = 0.0002$); The ERA5 excess in temperature and humidity vanishes when the seasonal modulation of variance is removed. No series survived Benjamini–Hochberg control with rank p-values (a result limited by resolution, amendment A4); with the z-score approximation 0 POWER and 2 ERA5 series were NPE-lower at $q = 0.05$. The number of series flagged as chaotic under the amended rule, which requires a significantly lower NPE and a positive exponent, is 0 in POWER and 0 in ERA5 (nominal, unadjusted $p \leq 0.05$: 0 and 0); under the original protocol rule it is 0 and 0.

Table 6. Embedding delay τ selected by the first minimum of the average mutual information (capped at 10 d), by variable and product (10 sites; m = 5 for all series).

Product	Variable	τ min (d)	τ median (d)	τ max (d)
ERA5	Solar radiation	10	10.0	10
ERA5	Precipitation	2	2.0	3
ERA5	Relative humidity	10	10.0	10
ERA5	Temperature	10	10.0	10
ERA5	Wind speed	2	8.0	10
NASA POWER	Solar radiation	3	6.5	10
NASA POWER	Precipitation	3	9.0	10
NASA POWER	Relative humidity	9	10.0	10
NASA POWER	Temperature	10	10.0	10
NASA POWER	Wind speed	3	9.5	10

Table 7. Main audit against null N2 (annual cycle and trend + IAAFT anomaly; B = 99).

Product	Variable	LLE > 0	Median NPE reduction (%)	Range (%)	NPE lower p ≤ 0.05	NPE higher p ≤ 0.05	NPE lower, FDR (rank)	NPE lower, FDR (z)	Flagged as chaos
ERA5	Solar radiation	0	1.9	-0.8 to 3.9	0	0	0	0	0
ERA5	Precipitation	10	-21.3	-50.2 to 7.9	0	5	0	0	0
ERA5	Relative humidity	0	4.6	-1.9 to 7.8	2	0	0	2	0
ERA5	Temperature	0	2.5	1.7 to 4.3	3	0	0	0	0
ERA5	Wind speed	3	0.7	-2.4 to 2.3	0	0	0	0	0

NASA POWER	Solar radiation	2	1.3	-1.4 to 4.6	1	0	0	0	0
NASA POWER	Precipitation	1	-1.1	-36.2 to 9.5	0	2	0	0	0
NASA POWER	Relative humidity	0	0.9	-1.9 to 3.9	1	0	0	0	0
NASA POWER	Temperature	0	0.2	-2.6 to 2.7	0	0	0	0	0
NASA POWER	Wind speed	3	-0.4	-2.5 to 2.0	0	0	0	0	0

LLE > 0: number of series with positive Lyapunov exponent in the 1–10 d window. NPE reduction = 100 × (surrogate mean − observed)/surrogate mean (negative = observed series less predictable than surrogates). Rank FDR: Benjamini–Hochberg on rank p-values (resolution-limited); z FDR: Benjamini–Hochberg on normal-approximation p-values from the surrogate z-scores. Flagged = NPE lower (rank FDR) and LLE > 0.

Two features stand out. First, for precipitation the observed series is *less* predictable than its surrogates in most series (ERA5 median reduction -21%, range -50% to 8%; POWER median -1.1%), and the ERA5 exponents are positive at all ten sites against 1 of ten in POWER. Precipitation is zero-inflated and bursty, so its marginal distribution and its temporal clustering of wet days carry structure that a linear surrogate does not reproduce, and that structure shows up as *more* unpredictability, not less. Positive exponents for precipitation therefore reflect intermittency, not determinism. Second, in ERA5, temperature and relative humidity have a small but systematic reduction of the prediction error, between 1.7 to 4.3 % for temperature and -1.9 to 7.8 % for humidity, which is an order of magnitude smaller than for the chaotic controls (92–100%) and of the same size as the weakest detectable control (Lorenz-63 with 50% noise, about 6%). The exponents of these series are negative, and the structure disappears once the seasonal variance is removed. The z-scores (Figs. 7 and 8) and the site-by-variable map of the reduction (Fig. 9) show that these are small, spatially diffuse departures from the null rather than a pocket of strongly deterministic sites.

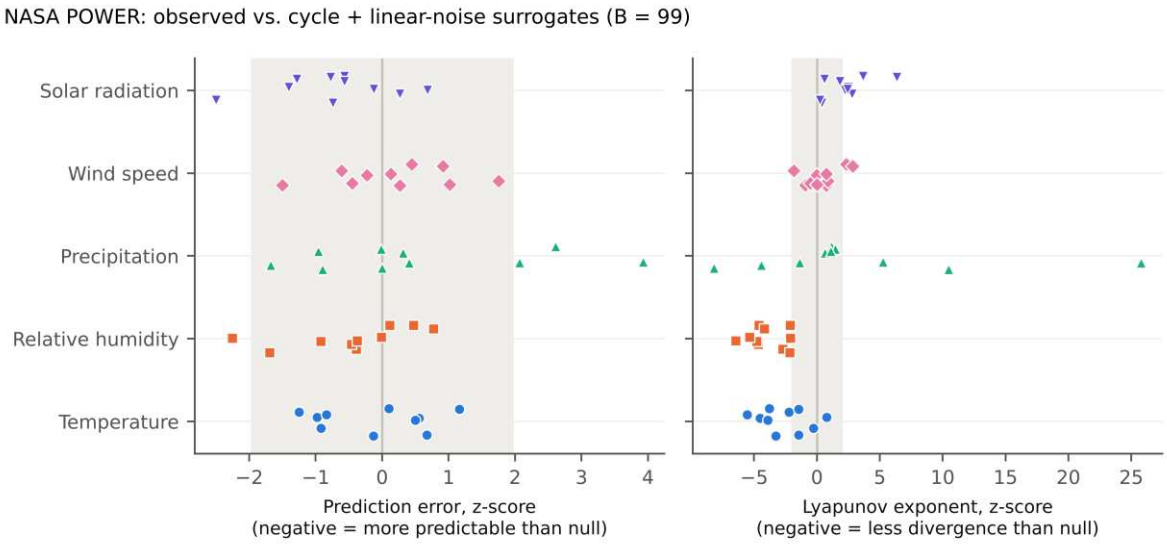


Fig. 7. NASA POWER: z-scores of the observed prediction error (left) and Lyapunov exponent (right) relative to the N2 surrogate distribution (B = 99), 10 sites per variable. Shaded band: ±1.96 SD. Chaotic systems would fall far to the left in the left panel (Fig. 5 shows the scale).

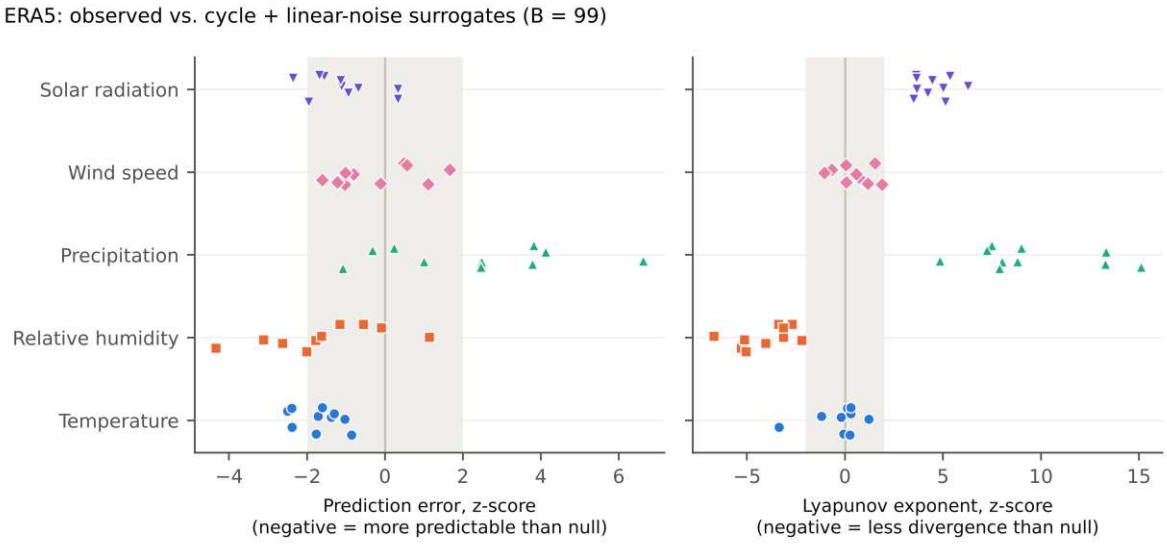


Fig. 8. As Fig. 7, for ERA5. The left-shift of temperature and humidity (small, systematic) and the right-shift of precipitation (less predictable than the null) are the only departures.

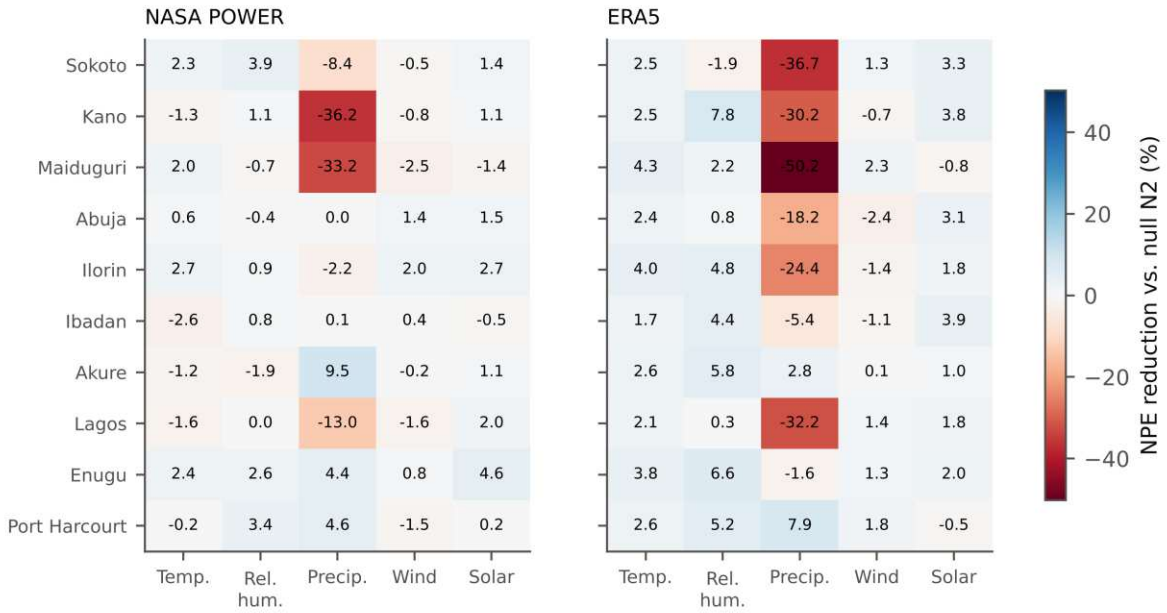


Fig. 9. Reduction of the prediction error (%) relative to N2 surrogates, by site (ordered by latitude) and variable, for the two products. Red: more predictable than the null; blue: less predictable.

The textbook null N1 (IAAFT surrogate of the raw series) rejects far more often in the direction of determinism than the seasonal null N2 (Table 8): NPE-lower nominal rejections 12 and 14 of 50 (POWER, ERA5) under N1, against 2 and 5 under N2 and 2 and 4 under N3. The difference is entirely attributable to the annual cycle: the raw series has a strongly non-Gaussian, periodic waveform that the IAAFT surrogate reproduces only through its amplitude distribution and a smeared spectral peak, so a more 'predictable' observed series is guaranteed. Rejection of a linear Gaussian null is thus the expected outcome for any seasonal series and carries no information on chaos (Fig. 10). Once the seasonal waveform is held fixed (N2) the excess disappears. The Lyapunov-exponent test rejects in 25 (POWER) and 30 (ERA5) of 50 series under N2, but a rejection of this test cannot be read as determinism given its inflated false-positive rate for skewed series

Table 8. Effect of the null model on the number of rejections (50 series per product and null).

Product	**Null**	**n**	**NPE lower $p \leq 0.05$**	**NPE higher $p \leq 0.05$**	**LLE $p \leq 0.05$**	**LLE > 0**	**NPE lower FDR (rank)**	**NPE lower FDR (z)**	**Flag (nominal)**	**Flag (FDR)**
ERA5	anom:iaaft	50	4	1	22	11	0	1	1	0
ERA5	raw:iaaft	50	14	3	23	13	0	10	2	0
ERA5	raw:seasonal	50	5	5	30	13	0	2	0	0
NASA POWER	anom:iaaft	50	2	1	24	12	0	0	0	0
NASA POWER	raw:iaaft	50	12	0	32	6	0	1	1	0

NASA POWER	raw:seasonal	50	2	2	25	6	0	0	0	0

N1: IAAFT(raw); N2: annual cycle + trend + IAAFT(anomaly); N3: IAAFT(anomaly). Flag = NPE lower than null and LLE > 0.

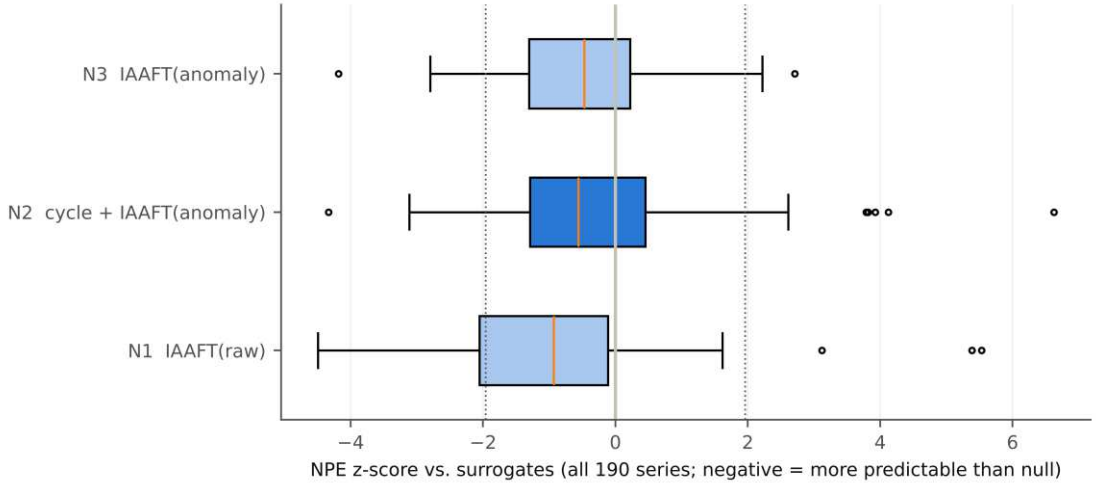


Fig. 10. Distribution of the prediction-error z-scores of all 100 series (both products) under the three nulls. The textbook null N1 yields strongly negative z-scores (an artefact of the annual cycle), the seasonal null N2 and the anomaly null N3 centre on zero with a narrow spread (outliers are precipitation).

POWER (MERRA-2) and ERA5 agree about the *absence* of chaos but not about every detail of the weak structure (Table 9, Fig. 11). The prediction-error reductions of temperature are strongly correlated across products (r = 0.80), as are those of precipitation (r = 0.81, driven by site contrast in intermittency), whereas humidity (-0.08) and solar radiation (0.26) are uncorrelated, and wind speed is anti-correlated (-0.77; note that POWER wind is at 2 m and ERA5 at 10 m). The exponents agree in sign in 100% (temperature) and 100% (humidity) of the sites but only in 10% for precipitation (Table 9). The conclusion that no series passes the chaos criterion in either product is robust; the *magnitude* of weak structure is not, which is itself evidence that it is an attribute of the product (model resolution, data assimilation, daily aggregation) rather than a property of the atmosphere alone.

Table 9. Agreement between NASA POWER and ERA5 across the ten sites (null N2).

Variable	**Sites**	**r (NPE red.)**	**ρ (NPE red.)**	**r (LLE)**	**Same sign of LLE (%)**
Solar radiation	10	0.26	0.12	-0.03	80
Precipitation	10	0.81	0.9	0.13	10
Relative humidity	10	-0.08	0.08	0.84	100
Temperature	10	0.8	0.76	0.91	100
Wind speed	10	-0.77	-0.85	0.11	60
All	50	0.79	0.41	0.04	70

r: Pearson; ρ: Spearman; n = 10 pairs per variable, 50 for 'All'.

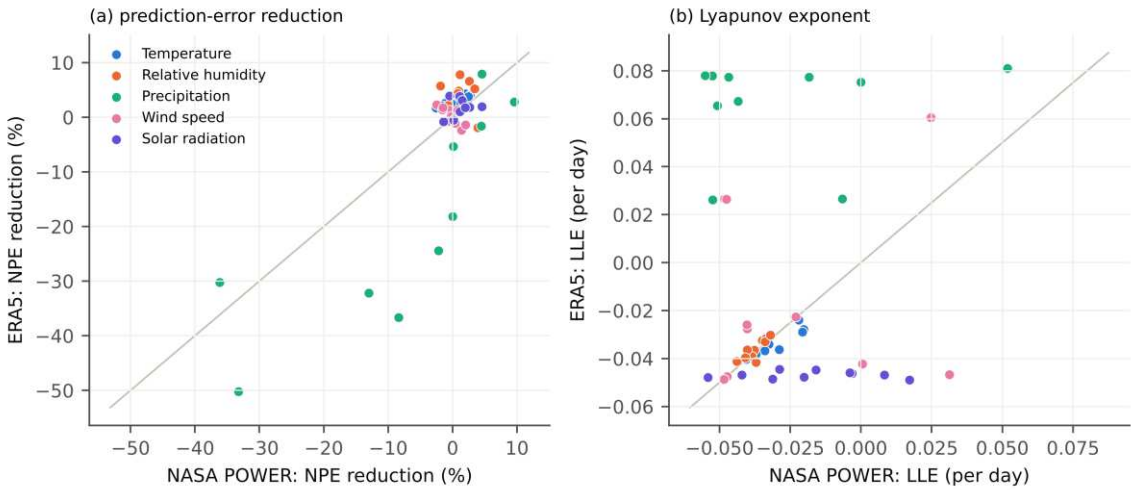


Fig. 11. Replication across data products: (a) prediction-error reduction and (b) Lyapunov exponent of the same site and variable in NASA POWER (x) and ERA5 (y). The 1:1 line is grey.

How often does an uncalibrated analysis return a positive exponent?

Across the 2610 defensible combinations of embedding dimension, delay and fit window (raw and anomaly series, ten sites, five variables), the Rosenstein exponent is positive in 83–89% of settings, depending on variable and series type (Table 10, Fig. 12). Applied to non-chaotic N2 surrogates the share is practically identical (85–89%). By window, the share of positive exponents for the observed raw series is 98% for the 1–5 d window, 90% for 5–20 d and 71% for the 1–10 d window used as primary window in the audit (surrogates: 95%, 93% and 74%), i.e. the sign of the exponent is decided by the window: our primary window happens to be the least favourable to a positive value, which is why the temperature and humidity exponents are negative there. A positive exponent is therefore the generic outcome for every series, chaotic or not, and carries no evidential weight on its own. Exponents of the observed series reached 0.40 d^{-1}, above the largest value reported as chaotic in the literature audited here (0.28 d^{-1}), and the per-variable medians (0.013–0.040 d^{-1}) lie within the reported range (0.001–0.28 d^{-1}).

Table 10. 'Multiverse' analysis: share of analysis settings ($m \in \{3,5,8\}$, $\tau \in \{1,3,\tau_{AMI}\}$, fit window $\in$ {1–5, 1–10, 5–20} d) giving a positive exponent, for observed series (POWER) and non-chaotic surrogates.

Variable	Observed raw, LLE>0 (%)	Observed anomaly, LLE>0 (%)	N2 surrogate, LLE>0 (%)	Median LLE (d^{-1})	Max LLE (d^{-1})	Settings
Solar radiation	87.407	87.243	88.123	0.013	0.282	513
Precipitation	87.356	89.259	89.286	0.04	0.399	531
Relative humidity	84.444	82.593	85.185	0.029	0.297	540
Temperature	85.556	82.593	86.667	0.033	0.319	540

Wind speed	87.302	88.889	88.066	0.02	0.285	486

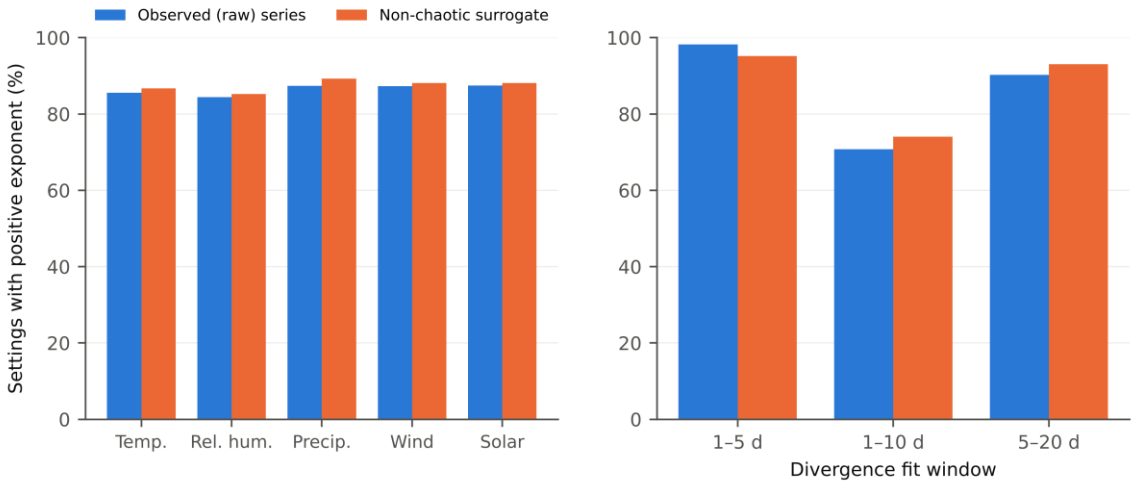


Fig. 12. (a) Share of settings with a positive exponent for observed raw series (blue) and non-chaotic N2 surrogates (orange), by variable. (b) The same by fit window. The two are indistinguishable.

The audited study reported higher exponents in the north for precipitation and temperature and in the south for wind speed and radiation. In our data the *observed* exponent does increase with latitude for temperature (Spearman ρ = 0.83 in POWER and 0.93 in ERA5), relative humidity (0.61; 0.73) and, in ERA5, precipitation (0.79) (Table 11, Fig. 13), i.e. northern series have less negative (or more positive) exponents. But the exponent of the *non-chaotic N2 surrogates* has the same gradient (ρ = 0.67 and 0.87 for temperature; 0.88 for ERA5 precipitation). The gradient is therefore produced by linear-stochastic structure (the spectrum and the amplitude distribution of the series, which change from the humid coast to the dry Sahel) and is reproduced by processes that are not chaotic. When the exponent is expressed relative to its own surrogate distribution (z-score), the gradient disappears in ERA5 ($|\rho| \leq 0.56$, all $p > 0.05$) and is reduced, but still present for temperature (ρ = 0.81, p = 0.005), in POWER. Wind speed and solar radiation show no gradient of the exponent at all ($|\rho| \leq 0.21$ in both products), unlike the south-ward increase reported for them. The prediction-error reduction, which is the more reliable diagnostic, has no latitudinal gradient except for precipitation, where it reflects intermittency (ρ = -0.68 and -0.73).

Table 11. Spearman correlation with latitude across the ten sites (null N2).

Product	**Variable**	**Sites**	**ρ LLE obs.**	**p**	**ρ LLE surr. mean**	**p**	**ρ LLE z**	**p**	**ρ NPE red.**	**p**
NASA POWER	Temperature	10	0.83	0.003	0.67	0.033	0.81	0.005	0.15	0.676
NASA POWER	Relative humidity	10	0.61	0.06	0.76	0.011	0.49	0.15	-0.01	0.987

NASA POWER	Precipitation	10	0.39	0.26	0.59	0.074	0.26	0.467	-0.68	0.029
NASA POWER	Wind speed	10	0.01	0.987	0.03	0.934	0.2	0.58	-0.02	0.96
NASA POWER	Solar radiation	10	-0.05	0.881	-0.05	0.881	0.58	0.082	-0.22	0.533
ERA5	Temperature	10	0.93	0.0	0.87	0.001	-0.54	0.108	-0.01	0.987
ERA5	Relative humidity	10	0.73	0.016	0.78	0.008	-0.15	0.676	-0.31	0.385
ERA5	Precipitation	10	0.79	0.006	0.88	0.001	-0.16	0.651	-0.73	0.016
ERA5	Wind speed	10	0.15	0.676	0.14	0.701	0.56	0.09	-0.21	0.556
ERA5	Solar radiation	10	0.21	0.556	0.15	0.676	0.31	0.385	0.41	0.244

LLE: 1–10 d window. 'LLE surr. mean': mean exponent of the 99 N2 surrogates. 'LLE z': observed exponent in units of surrogate SD. Ten sites per product, spatially dependent; the p-values are indicative.

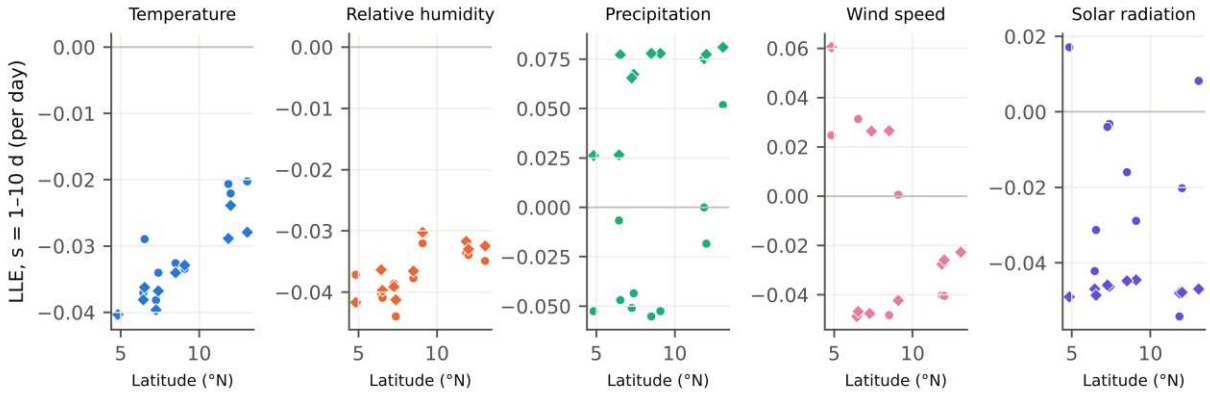


Fig. 13. Observed Lyapunov exponent (1–10 d window) versus latitude, by variable (circles: NASA POWER; diamonds: ERA5).

4.8 Replicating the pre-processing of the audited study (seasonal differencing)

In order to test whether the conclusions depend on the way the seasonal cycle is removed, the POWER series were seasonally differenced, $x(t) - x(t - 365)$, as in Echi et al. (2025), and analysed with the same diagnostics against IAAFT surrogates of the differenced series (B = 99). The result confirms the main audit (Table 12, Fig. 14). Temperature and humidity have negative exponents at all ten sites (median -0.039 and -0.042 d^{-1}), and the prediction error is not systematically lower than that of the surrogates (nominal NPE-lower rejections: 3 of 50). The exponent of precipitation is *higher* than that of the surrogates at 10 of ten sites, yet its prediction error is significantly lower than that of the surrogates at only 1 of ten sites (median reduction 3.9%): the signature of the intermittent, zero-inflated distribution again, and not of determinism. Differencing a series at lag 365 d is a linear filter that doubles the variance of independent noise and leaves the linear structure of the anomalies; it does not make the series 'stationary' in the dynamical sense and does not remove the seasonal modulation of variance.

The exponent of the differenced series has a latitudinal gradient for temperature (ρ = 0.77) and humidity (0.72) that is again of the type produced by linear processes.

Table 12. Seasonally differenced NASA POWER series (x(t) − x(t − 365 d)) against IAAFT surrogates (B = 99), 10 sites per variable.

Variable	Series	LLE > 0	Median LLE (d^{-1})	LLE higher, $p \leq 0.05$	LLE lower, $p \leq 0.05$	NPE lower, $p \leq 0.05$	Median NPE red. (%)	ρ LLE–latitude
Solar radiation	10	5	-0.008	5	0	1	0.2	-0.18
Precipitation	10	5	-0.002	10	0	1	3.9	0.21
Relative humidity	10	0	-0.042	3	1	1	-0.7	0.72
Temperature	10	0	-0.039	0	2	0	-1.0	0.77
Wind speed	10	4	-0.008	2	0	0	0.0	-0.21

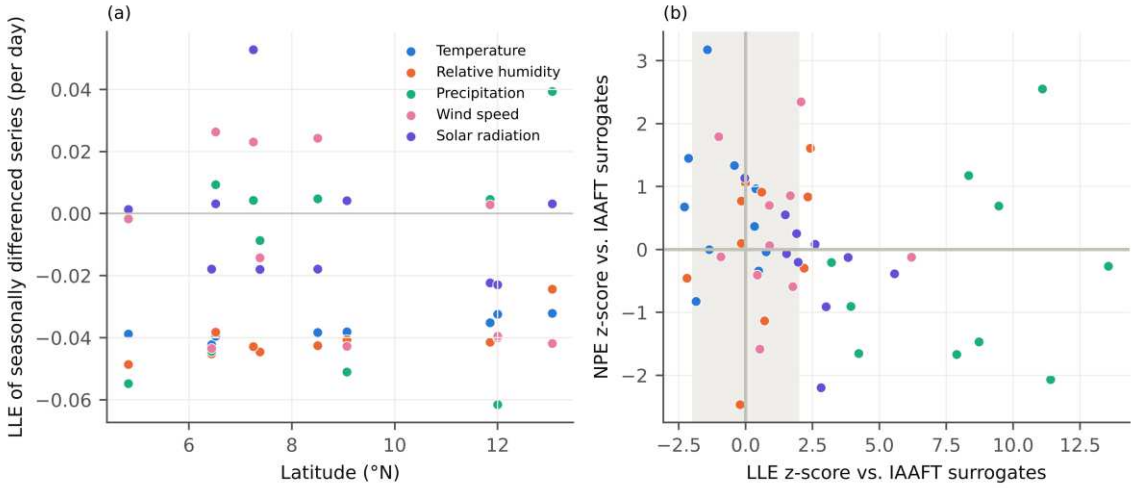


Fig. 14. Seasonally differenced POWER series. (a) Exponent versus latitude. (b) z-scores of the exponent and of the prediction error relative to IAAFT surrogates of the differenced series (shaded: ±1.96).

Validation of the products against station records

Station records are too sparse for dynamical analysis but sufficient to validate the products (Table 13, Fig. 14). Daily maximum temperature in POWER is biased by -3.3 to 0.7 °C relative to the stations (largest at Lagos, -3.3 °C); ERA5 biases are -3.0 to 0.3 °C. The correlation of the *anomalies* (seasonal cycle removed) with the stations is only 0.39–0.59 for POWER and 0.70–0.88 for ERA5: the reanalysis products reproduce the day-to-day variability of a point station only partly, and ERA5 does so better. Daily precipitation at Lagos and Port Harcourt (the only stations with at least 200 paired days, 997 and 338 days) is only weakly correlated with the gauge (r = 0.07–0.31). Both products show a jump in the annual-mean temperature between 2022 and 2023, much larger in POWER for the northern sites (Sokoto, Maiduguri, Kano: +1.8 to +2.8 °C) than in ERA5 (+0.8 to +1.0 °C). This is a

probable artefact in the later part of the POWER series; it is a source of non-stationarity that the linear trend in the anomaly definition absorbs only in part, and it does not create determinism, since all the tests are run against surrogates that contain the same step. Maiduguri, Kano and Ilorin have too few paired precipitation days (< 200) to be validated for that variable, and Maiduguri for temperature as well.

Table 13. Validation of NASA POWER and ERA5 against NOAA GHCN-Daily (paired days only; bias and RMSE in °C or mm d^{-1}; r: Pearson correlation).

Station	Variable	Product	Paired days	Bias (model − obs.)	RMSE	r (raw)	r (anomaly)
Kano	Tmax	POWER	1149	0.72	2.51	0.828	0.594
Kano	Tmax	ERA5	1149	-0.78	1.36	0.962	0.884
Kano	Tmin	POWER	1335	0.21	2.26	0.904	0.528
Kano	Tmin	ERA5	1335	0.7	1.91	0.942	0.709
Ilorin	Tmax	POWER	1893	-2.29	2.91	0.805	0.458
Ilorin	Tmax	ERA5	1893	0.27	1.28	0.921	0.704
Ilorin	Tmin	POWER	1675	-0.46	1.64	0.756	0.542
Ilorin	Tmin	ERA5	1675	1.07	1.56	0.862	0.702
Lagos	Tmax	POWER	4926	-3.29	3.67	0.777	0.394
Lagos	Tmax	ERA5	4926	-2.97	3.21	0.899	0.705
Lagos	Tmin	POWER	4044	0.75	1.44	0.633	0.544
Lagos	Tmin	ERA5	4044	1.5	1.83	0.716	0.656
Lagos	Precipitation	POWER	997	-4.81	31.65	0.157	0.157
Lagos	Precipitation	ERA5	997	-4.37	32.73	0.073	0.073
Port Harcourt	Tmax	POWER	1362	-2.4	3.04	0.657	0.407
Port Harcourt	Tmax	ERA5	1362	-1.59	1.98	0.883	0.705
Port Harcourt	Tmin	POWER	1209	0.26	1.19	0.67	0.582
Port Harcourt	Tmin	ERA5	1209	1.25	1.63	0.717	0.706
Port Harcourt	Precipitation	POWER	338	-0.63	24.39	0.312	0.312
Port Harcourt	Precipitation	ERA5	338	-1.42	28.31	0.069	0.069

Anomaly correlation: after removing a linear trend and four annual harmonics from both series (temperatures); for precipitation the raw correlation is repeated. Stations with fewer than 200 paired days are omitted.

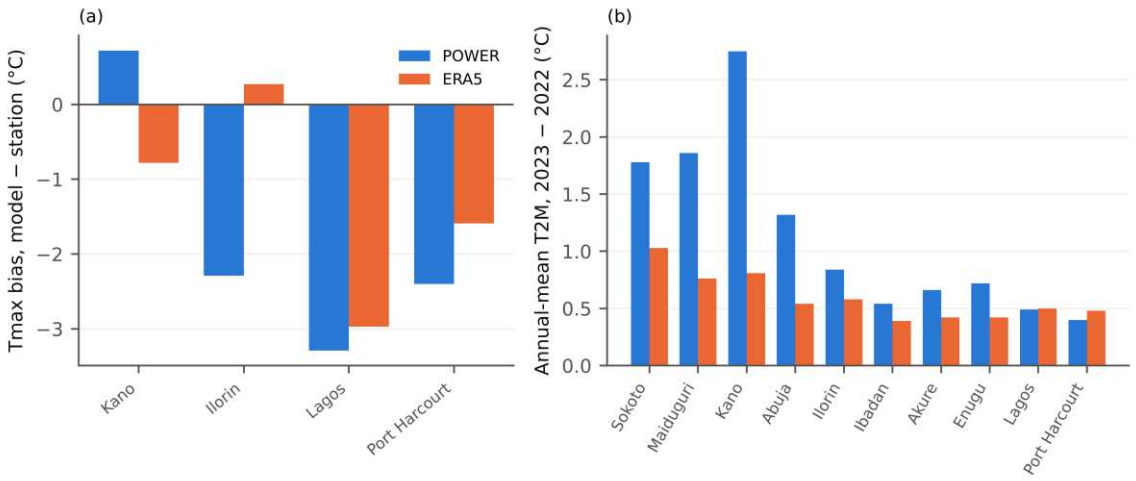


Fig. 15. (a) Bias of daily maximum temperature in POWER and ERA5 relative to the stations. (b) Change of the annual-mean 2 m temperature from 2022 to 2023 at the ten sites in the two products.

The 7 series with nominal NPE-lower $p \leq 0.05$ under N2 were re-run with B = 199 (Table 14). The reductions of 3.7–8.1% confirm the small effect sizes, while the p-values (0.01–0.04) are optimistic because these series were selected for being the most extreme among 100; with 100 series tested, one or two p-values below 0.02 are expected under the null. All exponents are negative (-0.042 to -0.029 d^{-1}), so none of these series would be flagged under the amended rule.

Table 14. Series with nominal NPE-lower $p \leq 0.05$ under N2, re-run with B = 199 (post-hoc, selected on the outcome).

product	site	var	null	B	NPE obs	NPE surr	NPE reduction %	P NPE	LLE obs
ERA5	Enugu	Relative humidity	seasonal	199	0.361	0.387	6.7	0.01	-0.036
ERA5	Enugu	Temperature	seasonal	199	0.454	0.471	3.7	0.03	-0.038
ERA5	Ilorin	Temperature	seasonal	199	0.43	0.449	4.1	0.02	-0.034
ERA5	Kano	Relative humidity	seasonal	199	0.211	0.229	8.1	0.01	-0.033
ERA5	Maiduguri	Temperature	seasonal	199	0.314	0.328	4.4	0.01	-0.029
NASA POWER	Enugu	Solar radiation	seasonal	199	0.933	0.977	4.5	0.04	-0.042
NASA POWER	Sokoto	Relative humidity	seasonal	199	0.256	0.267	4.0	0.01	-0.035

A seasonal-variance diagnostic (post hoc, not part of the protocol) tests whether the weak NPE reduction of ERA5 temperature and humidity (20 series) under N2 reflects the annual modulation of the variance of the anomaly (larger in the dry harmattan season than in the monsoon), which N2 does not reproduce. The anomaly was divided by a seasonally varying

scale and tested against IAAFT surrogates of itself (B = 99; Table 15). The median reduction is -0.3% (range -2.6 to 3.3%), compared with a median of 2.6% under N2 for the same series; 0 of 20 series are rejected at $p \leq 0.05$ after standardisation, against 5 of 20 nominally rejected under N2 (the reduction under N2 is positive in 19 of the 20 series). This indicates that most of the apparent structure under N2 is attributable to seasonal heteroscedasticity, a linear-stochastic but non-stationary feature.

Table 15. Seasonal-variance diagnostic: NPE reduction of the seasonally standardised anomaly relative to IAAFT surrogates of itself, ERA5 temperature and relative humidity (B = 99; post hoc).

Site	Variable	NPE reduction, standardised anomaly (%)	p	NPE reduction under N2 (%)
Abuja	Relative humidity	0.4	0.72	0.8
Akure	Relative humidity	3.3	0.06	5.8
Enugu	Relative humidity	0.6	0.72	6.6
Ibadan	Relative humidity	-2.3	0.32	4.4
Ilorin	Relative humidity	-1.3	0.3	4.8
Kano	Relative humidity	1.6	0.28	7.8
Lagos	Relative humidity	2.5	0.18	0.3
Maiduguri	Relative humidity	-1.7	0.22	2.2
Port Harcourt	Relative humidity	-0.3	0.84	5.2
Sokoto	Relative humidity	1.7	0.6	-1.9
Abuja	Temperature	0.1	0.92	2.4
Akure	Temperature	-1.2	0.34	2.6
Enugu	Temperature	-2.4	0.14	3.8
Ibadan	Temperature	-2.6	0.08	1.7
Ilorin	Temperature	1.0	0.44	4.0
Kano	Temperature	1.1	0.38	2.5
Lagos	Temperature	-2.0	0.26	2.1
Maiduguri	Temperature	-1.1	0.48	4.3
Port Harcourt	Temperature	-0.4	0.78	2.6
Sokoto	Temperature	-0.7	0.7	2.5

The NPE and the exponent are computed from a random sample of 2 500 reference points, and therefore carry Monte-Carlo noise (Table 16). The standard deviation over 30 independent draws is 0.0042–0.0064 for temperature and humidity (1–2% of the NPE),

against 0.076 (8%) for precipitation, and 0.0009–0.0024 d^{-1} for the exponent. The test is not affected, because the observed series is subsampled exactly like each surrogate (exchangeability), but the effect sizes in Table 7 have an uncertainty of about ±2 percentage points for temperature and humidity and much more for precipitation.

Table 16. Monte-Carlo variability of the observed diagnostics from the random selection of 2 500 reference points (30 draws), compared with using all reference points.

Product	Site	Variable	NPE mean	NPE SD	NPE (all refs)	LLE mean	LLE SD
ERA5	Kano	Rel. humidity	0.2222	0.0042	0.2211	-0.032	0.0013
ERA5	Lagos	Temperature	0.3998	0.0064	0.3997	-0.0349	0.0009
ERA5	Maiduguri	Temperature	0.3236	0.0042	0.3246	-0.0266	0.0011
ERA5	Port_Harcourt	Precipitation	1.0068	0.0757	1.0033	0.0268	0.0014
NASA POWER	Kano	Rel. humidity	0.2913	0.0053	0.29	-0.0337	0.0012
NASA POWER	Lagos	Temperature	0.3611	0.0063	0.3606	-0.0293	0.0011
NASA POWER	Maiduguri	Temperature	0.2872	0.0045	0.2877	-0.0204	0.0014
NASA POWER	Port_Harcourt	Precipitation	0.8805	0.0758	0.8593	-0.0509	0.0024

Extended variables: maximum and minimum temperature, dew point and pressure

The four additional variables (five sites each, two products; 40 series) give the same picture (Table 17, Fig. 16). Nominal NPE-lower rejections of N2 were 3 and NPE-higher 0 of 40 (expected by chance about 2), the median reduction of the prediction error lay between -1.0% and 1.8%, and no series satisfied the flagging rule. The exponents of maximum and minimum temperature and of dew point are negative at all but 1 of 30 series, whereas surface pressure, a smooth, highly persistent variable, has positive exponents in 5 of 10 series without any gain in predictability. Minimum temperature has the same latitude-dependent exponent as the mean temperature ($\rho = 0.9$ in both products, five sites), a linear-process gradient.

Table 17. Extended audit of four additional variables against null N2 (B = 99; five sites per product).

Product	Variable	Series	LLE> 0	Median NPE red. (%)	Range (%)	NPE lower $p \leq 0.05$	NPE higher $p \leq 0.05$	Flagged (nominal)	ρ LLE–lat.
ERA5	Surface pressure	5	1	-1.0	-2.9 to 0.9	0	0	0	-0.6
ERA5	Dew point	5	0	-0.7	-2.8 to 4.7	1	0	0	0.7
ERA5	Max.	5	1	0.7	-1.2	0	0	0	-0.1

	temperature				to 2.6				
ERA5	Min. temperature	5	0	1.1	-1.6 to 2.2	0	0	0	0.9
NASA POWER	Surface pressure	5	4	0.4	-2.5 to 1.1	0	0	0	-0.6
NASA POWER	Dew point	5	0	-0.4	-4.2 to 6.1	1	0	0	0.0
NASA POWER	Max. temperature	5	0	1.8	0.0 to 3.2	0	0	0	-0.2
NASA POWER	Min. temperature	5	0	0.9	-0.2 to 7.9	1	0	0	0.9

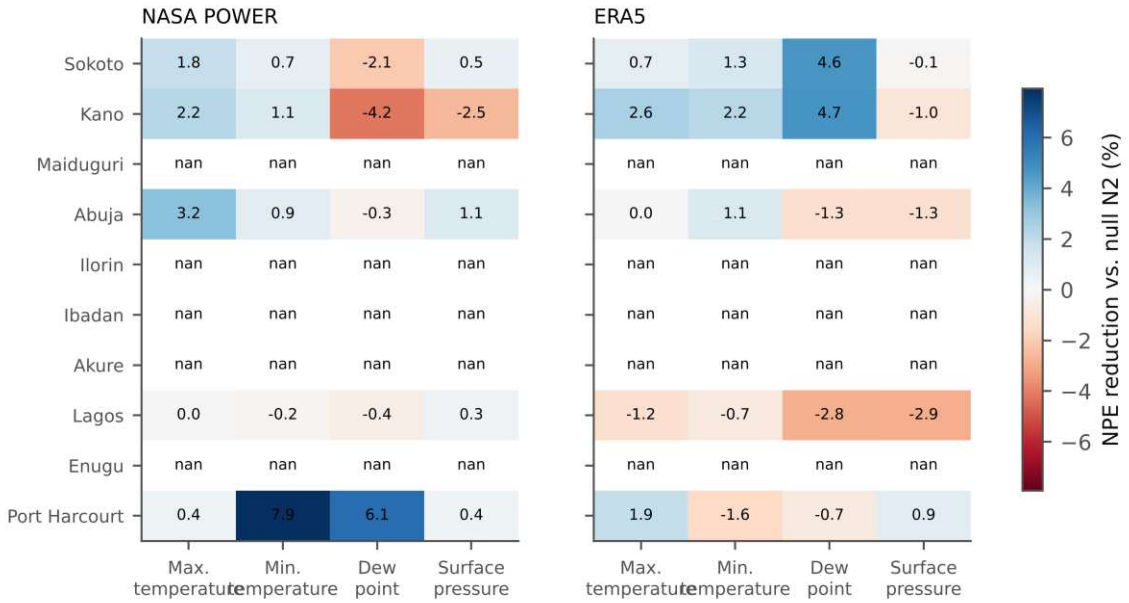


Fig. 16. Prediction-error reduction (%) relative to N2 for the four additional variables at five sites (ordered by latitude), in the two products.

Scale dependence: daily to monthly resolution

Averaging the series over 3, 7 and 30 days removes much of the day-to-day noise and could expose a slow deterministic component, if there were one. It does not reveal chaos, although a weak structure appears at the monthly scale (Table 18, Fig. 17). The number of nominal NPE-lower rejections of N2 over the 100 series is 6, 6 and 14 at 3, 7 and 30 d, with 8, 4 and 1 in the opposite direction. The only variable whose median reduction increases with the scale is relative humidity (POWER 0.5% at 3 d and 7.4% at 30 d; ERA5 1.3% and 12.4%), followed by POWER temperature (5.5% at 30 d). With N ≈ 500 monthly values this is also the scale at which the sampling variability of the prediction error is largest (range -9.1 to 20.8 for ERA5 humidity). The monthly series change their variance strongly across the year (dry

harmattan versus wet monsoon), so we repeated the 30-day test on the seasonally standardised anomaly: the median reduction falls to 1.1% (POWER humidity), 1.0% (ERA5 humidity), 1.5% (POWER temperature) and 4.1% (ERA5 temperature), with 1 of 40 series rejected. The 'exponent' in the short fit window of 1–5 samples is positive in all 100 series at 7 d and at 30 d, but in only 6 at 3 d (Table 18): the log-divergence of the aggregated series is still rising at those lags, and the positive slope is another instance of the window dependence not an indication of chaos.

Table 18. Multi-scale audit: N2 test on 3-, 7- and 30-day means (B = 99; 10 sites per cell).

Product	Variable	Scale (d)	N	Series	LLE > 0	Median NPE red. (%)	Range (%)	NPE lower p ≤ 0.05	NPE higher p ≤ 0.05
ERA5	Precipitation	3	4911	10	2	-5.9	-44.2 to 5.5	0	5
ERA5	Precipitation	7	2105	10	10	-1.7	-31.7 to 6.9	0	3
ERA5	Precipitation	30	491	10	10	0.4	-8.8 to 15.1	0	0
ERA5	Relative humidity	3	4993	10	0	1.3	-2.9 to 4.2	2	0
ERA5	Relative humidity	7	2139	10	10	3.0	-7.1 to 9.0	3	0
ERA5	Relative humidity	30	499	10	10	12.4	-9.1 to 20.8	6	0
ERA5	Solar radiation	3	5095	10	0	2.4	-0.3 to 4.9	3	0
ERA5	Solar radiation	7	2183	10	10	3.0	0.5 to 5.7	1	0
ERA5	Solar radiation	30	509	10	10	0.8	-6.7 to 5.3	0	0
ERA5	Temperature	3	4988	10	0	0.6	-1.3 to 1.5	0	0
ERA5	Temperature	7	2137	10	10	0.1	-3.1 to 4.7	1	0
ERA5	Temperature	30	498	10	10	-2.5	-12.9 to 11.0	1	0
ERA5	Wind speed	3	5084	10	1	0.6	-1.5 to 1.7	0	0
ERA5	Wind speed	7	2179	10	10	1.3	-1.6 to 3.8	0	0
ERA5	Wind speed	30	508	10	10	1.0	-5.5 to 8.8	0	0
NASA POWER	Precipitation	3	4998	10	1	1.3	-32.3 to 4.9	0	2
NASA POWER	Precipitation	7	2142	10	10	-5.3	-26.7 to 3.6	0	1

NASA POWER	Precipitation	30	499	10	10	-1.6	-27.3 to 18.0	0	1
NASA POWER	Relative humidity	3	5002	10	0	0.5	-3.7 to 5.1	1	1
NASA POWER	Relative humidity	7	2143	10	10	2.9	-1.8 to 6.3	1	0
NASA POWER	Relative humidity	30	500	10	10	7.4	1.0 to 17.6	3	0
NASA POWER	Solar radiation	3	5057	10	2	-0.6	-2.0 to 2.0	0	0
NASA POWER	Solar radiation	7	2167	10	10	2.6	-1.9 to 4.8	0	0
NASA POWER	Solar radiation	30	505	10	10	-0.4	-8.3 to 11.3	1	0
NASA POWER	Temperature	3	5023	10	0	-0.1	-2.5 to 2.0	0	0
NASA POWER	Temperature	7	2152	10	10	1.7	-2.5 to 3.5	0	0
NASA POWER	Temperature	30	502	10	10	5.5	-5.0 to 12.5	2	0
NASA POWER	Wind speed	3	4979	10	0	0.9	-2.5 to 2.3	0	0
NASA POWER	Wind speed	7	2134	10	10	-0.1	-3.1 to 5.5	0	0
NASA POWER	Wind speed	30	497	10	10	2.0	-8.2 to 13.1	1	0

LLE: slope over 1–5 samples of the aggregated series.

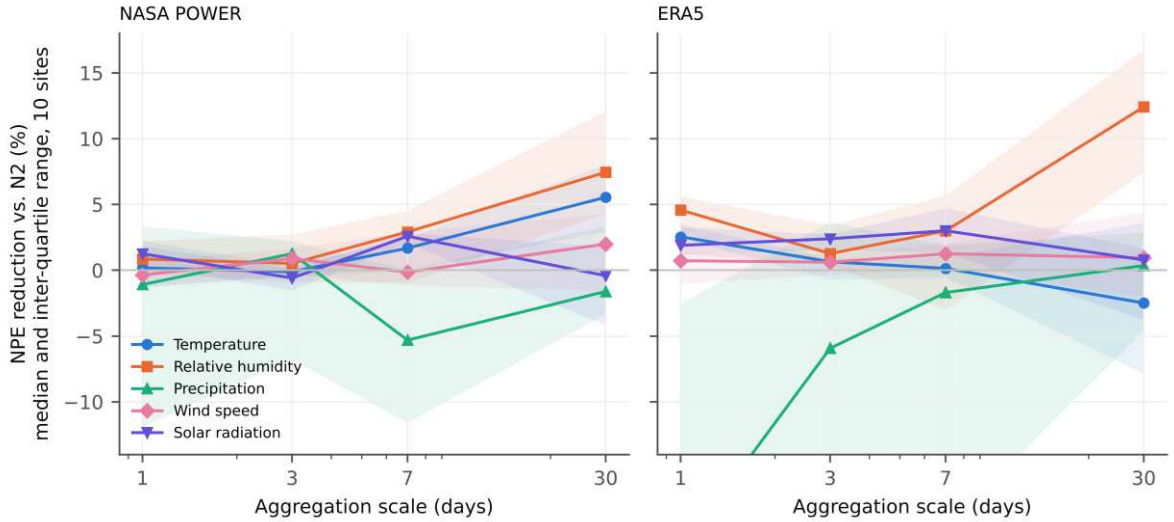


Fig. 17. Prediction-error reduction relative to N2 as a function of the aggregation scale (1 d = main audit). Lines: median over the ten sites; bands: inter-quartile range. The ERA5 precipitation median at 1 d (−21%) lies below the plotted range.

Ordinal-pattern diagnostics: permutation entropy and statistical complexity

Permutation entropy places every real series in the high-entropy, low-complexity corner of the complexity–entropy plane (Fig. 18a), with H between 0.90 and 0.998 and C below 0.14, in the region of linear-correlated noise (AR(1) with $\varphi = 0.9$ has H = 0.93, C = 0.10) and far from the two chaotic controls (logistic map H = 0.68, C = 0.40; Lorenz-63 H = 0.41, C = 0.31). No weather series approaches the region occupied by low-dimensional chaos. Against the seasonal null, however, H is not exchangeable with the surrogate values (Table 19, Fig. 18b): for solar radiation H is lower than N2 at all 20 series (both products; delay 1 d), for precipitation at 11 of 20 series, whereas relative humidity (17 of 20) and temperature (6 of 20) are *more* disordered than their surrogates. The direction of the departure thus differs between variables, and for temperature and humidity it is opposite to what determinism would imply. Solar radiation is bounded from above by the clear-sky envelope and falls abruptly under convective cloud, an asymmetric (time-irreversible) behaviour that a surrogate with the same spectrum and amplitude distribution does not reproduce, and that makes consecutive values more ordinally structured without being low-dimensional chaos. Ordinal patterns are therefore a sensitive detector of departures from the linear null, which is useful, but the direction and the position in the complexity–entropy plane do not support chaos.

Table 19. Permutation entropy H (pattern order 5) of the raw series against N2 (B = 99): number of series (of 10) with $p \leq 0.05$ in each direction, delays 1 and 3 d.

Product	Variable	Series	Median H obs. (d=1)	Median H N2 (d=1)	H lower (d=1)	H higher (d=1)	H lower (d=3)	H higher (d=3)
ERA5	Solar radiation	10	0.9842	0.9924	10	0	10	0
ERA5	Precipitation	10	0.9955	0.9959	3	1	2	4
ERA5	Relative humidity	10	0.9694	0.9361	0	10	0	10
ERA5	Temperature	10	0.9668	0.9739	10	0	1	4
ERA5	Wind speed	10	0.9731	0.9718	4	3	3	1
NASA POWER	Solar radiation	10	0.9933	0.9989	10	0	6	0
NASA POWER	Precipitation	10	0.979	0.9883	8	1	7	2
NASA POWER	Relative humidity	10	0.9687	0.9435	2	7	0	8
NASA POWER	Temperature	10	0.9618	0.9506	2	6	0	10
NASA POWER	Wind speed	10	0.9702	0.968	2	6	1	4

Controls on the same scale: white noise H = 0.999, C = 0.002; AR(1) $\varphi = 0.9$, H = 0.931, C = 0.103; logistic map r = 4, H = 0.679, C = 0.399; Lorenz-63, H = 0.409, C = 0.314.

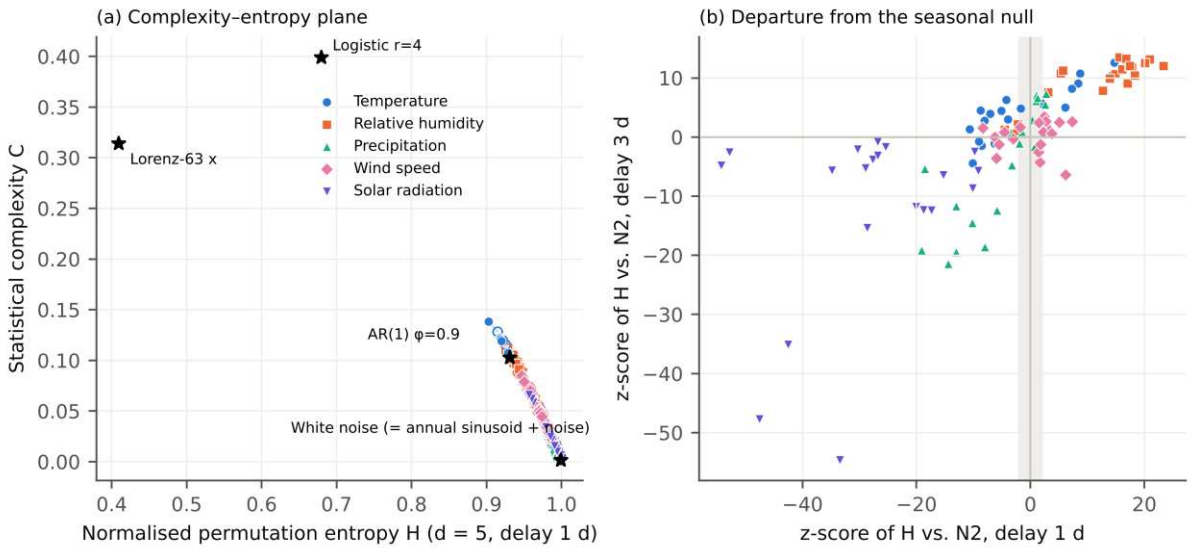


Fig. 18. (a) Complexity–entropy plane for all 100 real series (filled symbols: observed; open symbols: N2 surrogate mean; lines connect them) with known-answer controls (stars). (b) z-scores of the permutation entropy with respect to N2 at delays of 1 and 3 d (shaded: ±1.96).

Table 20 and Fig. 19 summarise the evidence across all diagnostics for the 100 core series. Positive exponents are common in precipitation (11 of 20) and wind (6), but never together with a lower prediction error. A lower prediction error occurs in 7 of 100 series at the daily scale and in 14 at 30 d; lower permutation entropy occurs in 51 (dominated by solar radiation, 20 of 20). The criterion that all four independent indicators agree (NPE lower at the daily and at the 30-day scale, PE lower, exponent positive) is met by 0 of 100 series, and the formal rule (FDR-controlled lower NPE with positive exponent) by 0. The diagnostics disagree with each other far more than they would if a common deterministic cause were responsible, which is itself informative: each captures a different linear-process feature (seasonal variance, bounded amplitude, intermittency).

Table 20. Evidence matrix: number of series (both products, 10 sites each) in which each diagnostic points in the indicated direction ($p \leq 0.05$, nominal) for the five core variables.

Variable	Series	LLE> 0	NPE ↓ (d)	NPE ↑ (d)	PE ↓	PE ↑	NPE ↓ (30 d)	NPE ↑ (30 d)	All four	Flag (FDR)
Temperature	20	0	3	0	12	6	3	0	0	0
Relative humidity	20	0	3	0	2	17	9	0	0	0
Precipitation	20	11	0	7	11	2	0	1	0	0
Wind speed	20	6	0	0	6	9	1	0	0	0
Solar radiation	20	2	1	0	20	0	1	0	0	0
All variables	100	19	7	7	51	34	14	1	0	0

↓: observed prediction error or permutation entropy lower than the null (more ordered); ↑: higher. 'All four': NPE lower (daily), NPE lower (30 d), PE lower and LLE > 0. 'Flag (FDR)': formal decision rule.

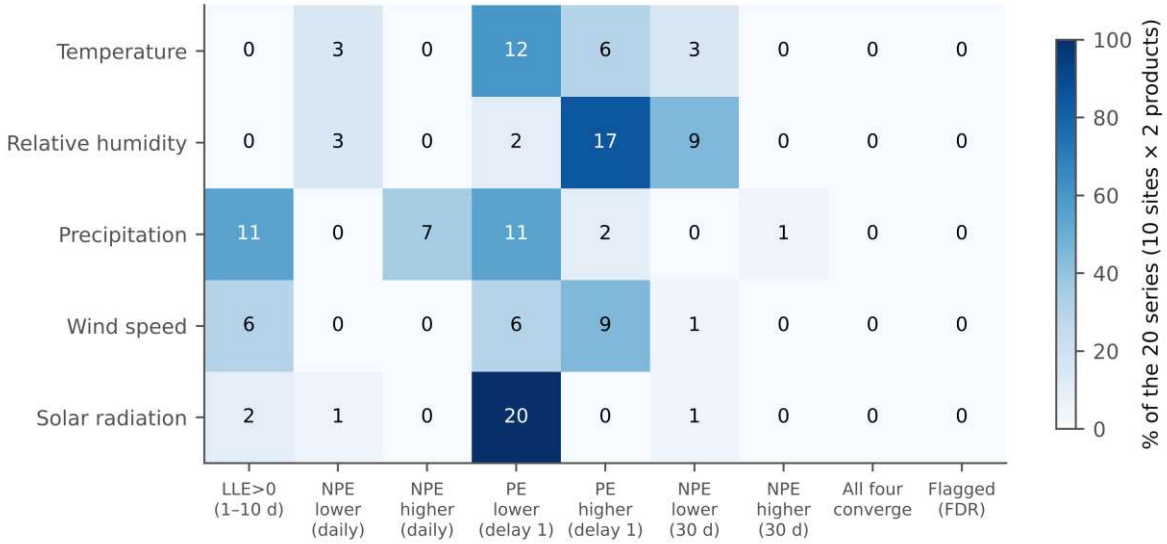


Fig. 19. Evidence matrix (Table 20). Cell labels: number of series; colour: share of the 20 series of that variable.

5. Discussion

5.1 Answers to the Research questions

(Do the diagnostics differ from those of non-chaotic surrogates?). Not in the way chaos requires. Against the seasonal null, 4 of 50 POWER series (consistent with chance) and 10 of 50 ERA5 series differ in the prediction error, half of them in the direction of *less* determinism, and no series passes the joint criterion. **Q2 (Is this robust?)** Yes with respect to the choice of null (the textbook null rejects often, but only because of the annual cycle), the data product, the removal of the seasonal cycle (anomaly, differencing), the time scale and the diagnostic. The magnitude of weak departures is product-dependent. **Q3 (Would the procedure have detected chaos?)** Yes for the logistic map and for Lorenz-63 (100% power, reductions of 92–100%) and still for Lorenz-63 hidden in 50% noise (6%, $p = 0.02$); the empirical false-positive rate was 3.3% (Table 5). **Q4 (Do the latitudinal gradients survive?)** No: the gradient of the observed exponent is reproduced by non-chaotic surrogates.

5.2 Why a positive exponent is not evidence of chaos

Three results explain the ubiquity of positive exponents in the literature. First, the Rosenstein estimator measures the rate at which nearest neighbours diverge on average over a chosen range of lags, and for a persistent stochastic process neighbours also diverge: the log-distance rises because the stochastic increments accumulate, not because of sensitive dependence. The multiverse shows that 83–89% of defensible settings produce a positive slope for observed series and surrogates alike. Second, the known-answer controls show that the estimate

depends on the window (Lorenz-63: 1.65 to 0.88 per time unit, true value 0.91) and that a noisy periodic orbit gives a positive exponent. Third, the surrogate test with the exponent as the statistic is invalid for skewed series (74% false rejections for exp(AR(1)), Table 5) and unreliable in direction even for true chaos (the chaotic controls have a *lower* exponent than their surrogates). The prediction error has none of these problems: it is bounded, it responds to determinism in the right direction and it is calibrated by the size study. The practical message is that an exponent can be reported, but only together with its null distribution, a control with known dynamics, and the window sensitivity.

5.3 The nature of the weak structure that remains

Not all departures from the null are noise. The prediction error of ERA5 temperature and humidity is up to about 8% lower than N2, permutation entropy departs from N2 in both directions, and the monthly means of humidity do the same at the longest scale. Four observations show that these are departures from the *specific* null, not signs of chaos. (i) The prediction-error reduction disappears when the seasonal modulation of the variance is removed (Table 15): the weather in the dry harmattan season is less variable than in the wet season, so a stationary linear surrogate is the wrong description of the data even though the process is stochastic. (ii) The exponents of these series are negative and the series lie in the high-entropy, low-complexity corner, nowhere near the chaotic controls (Fig. 18a). (iii) The structure depends on the data product (POWER temperature has no systematic reduction, ERA5 has) and is therefore partly a property of the reanalysis (resolution, assimilation, treatment of the diurnal cycle), as the weak correlation of the reanalysis anomalies with station data confirms. (iv) For solar radiation the entropy is *lower* than that of the surrogates at all sites in both products, a robust, physically interpretable non-Gaussian, time-irreversible feature (clear-sky ceiling, abrupt cloud events) that is again a property of the amplitude dynamics and not low-dimensional chaos. That the weak structure ($\leq 8\%$) is of the same size as the weakest chaotic control (6%) means that these tests cannot exclude a deterministic component that is buried in noise at the 5–10% level, but there is also no positive evidence for one, and the structure that is present has a simpler explanation.

5.4 The latitudinal gradient is a property of linear structure

The exponent of temperature increases with latitude ($\rho = 0.83$ in POWER, 0.93 in ERA5), as does that of humidity. A reader of the exponent alone might conclude that 'chaos increases towards the Sahel'. The non-chaotic surrogates have almost the same gradient ($\rho = 0.67$ and 0.87). The reason is the transition from the coast, where a humid, cloudy and weakly forced climate gives strongly persistent anomalies with small amplitude, to the Sahel, where the seasonal cycle is large and the anomalies are dominated by the harmattan–monsoon alternation. The spectrum and the amplitude distribution, which the surrogate preserves, carry the gradient. After the exponent is standardised by its own surrogate distribution the gradient

vanishes in ERA5 and is reduced in POWER. Any claim of a spatial pattern of chaos must therefore be made on the standardised statistic, not on the exponent.

6. Conclusions

A calibrated, pre-specified surrogate-data audit of daily temperature, humidity, precipitation, wind speed, radiation, dew point and pressure at ten Nigerian sites over 42 years, in two independent data products, finds no series that behaves as a low-dimensional chaotic system. The key results are: (1) the known-answer controls show that the prediction-error test detects chaos with 100% power and about 3% false positives, whereas the Lyapunov-exponent test is invalid for skewed series; (2) a positive exponent is the generic outcome of uncalibrated analysis (83–89% of settings) for observed and non-chaotic surrogate series; (3) the north–south gradient of the exponent is produced by linear stochastic structure; (4) the weak structure that remains is attributable to seasonal variance modulation, bounded amplitude and intermittency and depends on the data product; (5) no scale between 1 and 30 d reveals determinism; and (6) permutation entropy and complexity place all series in the stochastic region. The atmosphere is chaotic in principle, but daily weather series over Nigeria do not contain a detectable low-dimensional chaotic signature beyond that of forced linear noise, and claims to the contrary need a calibrated null distribution.

Data availability

All data used are public: NASA POWER (https://power.larc.nasa.gov), ERA5 through the Open-Meteo Historical Weather API (https://open-meteo.com; contains modified Copernicus Climate Change Service information) and NOAA GHCN-Daily (https://www.ncei.noaa.gov/pub/data/ghcn/daily/). The downloaded files exactly as retrieved, the cleaned series, the quality-control tables, every per-series result and all figures and tables are provided in the Supplementary Material, together with the retrieval scripts (Google Colab notebooks).

Code availability

All analysis code (Python), the protocol, the data-source log and the fixed random seeds are available upon request from the corresponding Author.

Author contributions (CRediT)

Iyanuoluwa Olaniyi Ajayi: Conceptualisation; Methodology; Software; Formal analysis; Investigation; Data curation; Writing – original draft; Writing – review and editing:

Temitope Y. Ojebisi: Investigation; Data curation; Writing – review and editing; Supervision.

Funding

The research do not receive any funding.

Competing interests

The authors declare no competing interests

Acknowledgements

We acknowledge the NASA POWER project, the Copernicus Climate Change Service and ECMWF (ERA5), Open-Meteo, and NOAA NCEI (GHCN-Daily) for making the data openly available.

References

Adelakun AO, Ajayi IO, Adebisi AC (2026) Scale-dependent coupling and directional causality in solar–wind dynamics. Eur Phys J Plus 141:741. https://doi.org/10.1140/epjp/s13360-026-07955-4

Ajayi IO (2026) Chaotic dynamics, basin structure and global dynamical analysis of a six-dimensional series memcapacitor system. Discover Electronics 3(1):147. https://doi.org/10.1007/s44291-026-00297-6

Bandt C, Pompe B (2002) Permutation entropy: a natural complexity measure for time series. Phys Rev Lett 88:174102

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Casdagli M (1989) Nonlinear prediction of chaotic time series. Physica D 35:335–356

Davison AC, Hinkley DV (1997) Bootstrap Methods and their Application. Cambridge University Press, Cambridge

Echi IM, Tikyaa EV, Eweh EJ, Tyovenda AA, Igbawua T (2025) Investigation of the spatial distribution of deterministic chaos in some meteorological variables across Nigeria. African Scientific Reports 4(2):280. https://doi.org/10.46481/asr.2025.4.2.280

Farmer JD, Sidorowich JJ (1987) Predicting chaotic time series. Phys Rev Lett 59:845–848

Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from mutual information. Phys Rev A 33:1134–1140

Fuwape IA, Oluyamo SS, Rabiu AB, Ogunjo ST (2018) Chaotic signature of climate extremes. arXiv:1901.02065

Gelaro R, McCarty W, Suárez MJ, et al. (2017) The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J Clim 30:5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gottwald GA, Melbourne I (2004) A new test for chaos in deterministic systems. Proc R Soc Lond A 460:603–611. https://doi.org/10.1098/rspa.2003.1183

Gottwald GA, Melbourne I (2005) Testing for chaos in deterministic systems with noise. Physica D 212:100–110

Grassberger P (1986) Do climatic attractors exist? Nature 323:609–612

Hersbach H, Bell B, Berrisford P, et al. (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146:1999–2049. https://doi.org/10.1002/qj.3803

Kantz H, Schreiber T (2004) Nonlinear Time Series Analysis, 2nd edn. Cambridge University Press, Cambridge

Kennel MB, Brown R, Abarbanel HDI (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 45:3403–3411

Lancaster G, Iatsenko D, Pidde A, Ticcinelli V, Stefanovska A (2018) Surrogate data for hypothesis testing of physical systems. Phys Rep 748:1–60

Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130–141

Lorenz EN (1969) The predictability of a flow which possesses many scales of motion. Tellus 21:289–307

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG (2012) An overview of the Global Historical Climatology Network-Daily database. J Atmos Ocean Technol 29:897–910

NASA Langley Research Center (LaRC) POWER Project. Prediction Of Worldwide Energy Resources (POWER), daily point data, community AG. https://power.larc.nasa.gov (accessed 7 November 2026)

Nicolis C, Nicolis G (1984) Is there a climatic attractor? Nature 311:529–532

Open-Meteo.com. Historical Weather API (ERA5). https://open-meteo.com (accessed 7–8 November 2026)

Osborne AR, Provenzale A (1989) Finite correlation dimension for stochastic systems with power-law spectra. Physica D 35:357–381

Phipson B, Smyth GK (2010) Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. Stat Appl Genet Mol Biol 9:Article 39

Provenzale A, Smith LA, Vio R, Murante G (1992) Distinguishing between low-dimensional dynamics and randomness in measured time series. Physica D 58:31–49

Rosenstein MT, Collins JJ, De Luca CJ (1993) A practical method for calculating largest Lyapunov exponents from small data sets. Physica D 65:117–134

Rosso OA, Larrondo HA, Martín MT, Plastino A, Fuentes MA (2007) Distinguishing noise from chaos. Phys Rev Lett 99:154102

Schreiber T, Schmitz A (1996) Improved surrogate data for nonlinearity tests. Phys Rev Lett 77:635–638

Schreiber T, Schmitz A (2000) Surrogate time series. Physica D 142:346–382

Sivakumar B (2004) Chaos theory in geophysics: past, present and future. Chaos Solitons Fractals 19:441–462

Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344:734–741

Takens F (1981) Detecting strange attractors in turbulence. In: Rand DA, Young L-S (eds) Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics 898. Springer, Berlin, pp 366–381

Theiler J (1986) Spurious dimension from correlation algorithms applied to limited time-series data. Phys Rev A 34:2427–2432

Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for nonlinearity in time series: the method of surrogate data. Physica D 58:77–94

Tsonis AA, Elsner JB (1988) The weather attractor over very short timescales. Nature 333:545–547