Editors' Suggestion

Detecting and Attributing Change in Climate and Complex Systems: Foundations, Green's Functions, and Nonlinear Fingerprints

Valerio Lucarini[1,*] and Mickaël D. Chekroun[2,3]
[1]School of Computing and Mathematical Sciences, University of Leicester, Leicester LE17RH, United Kingdom
[2]Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot 76100, Israel
[3]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, California, USA

(Received 5 December 2022; revised 27 February 2024; accepted 18 October 2024; published 9 December 2024)

Detection and attribution (DA) studies are cornerstones of climate science, providing crucial evidence for policy decisions. Their goal is to link observed climate change patterns to anthropogenic and natural drivers via the optimal fingerprinting method (OFM). We show that response theory for nonequilibrium systems offers the physical and dynamical basis for OFM, including the concept of causality used for attribution. Our framework clarifies the method's assumptions, advantages, and potential weaknesses. We use our theory to perform DA for prototypical climate change experiments performed on an energy balance model and on a low-resolution coupled climate model. We also explain the underpinnings of degenerate fingerprinting, which offers early warning indicators for tipping points. Finally, we extend the OFM to the nonlinear response regime. Our analysis shows that OFM has broad applicability across diverse stochastic systems influenced by time-dependent forcings, with potential relevance to ecosystems, quantitative social sciences, and finance, among others.

DOI: 10.1103/PhysRevLett.133.244201

Detection and attribution of climate change—The climate is a complex system comprising five subsystems—the atmosphere, hydrosphere, cryosphere, biosphere, and land surface—with a myriad of physical, biological, and chemical processes interacting over a wide range of spatiotemporal scales [1,2]. The resulting dynamics is multiscale, with different subsystems having a dominant role depending on the scales the climate is looked at [3–6], with subtle interactions across scales that are yet to be understood [7]. The role of tipping points has emerged as a key climate change feature [8–12], in particular, due to the interplay of forcings, instabilities, and feedbacks [7,13,14].

The complexity of climate makes it difficult to disentangle climate variability and the forced climate change signal. Nonetheless, the successive reports of the Intergovernmental Panel on Climate Change (IPCC) have confirmed two key scientific advances obtained via detection and attribution (DA) studies: (1) statistical evidence of the change of the current state of the climate system with respect to the conditions prevailing in the 19th and early 20th century and (2) statistical evidence that such a change can be attributed (for the most part) to anthropogenic causes [15,16]. More recently, DA studies have also investigated climate change at a local level [17] including the relationship between climate change and extreme events [18]. The current focus on extremes in a changing climate [19,20] and on tipping points has recently led the scientific community to use expressions like "climate crisis" or "climate emergency" instead of climate change [21,22].

DA studies wish to link in a causal sense the observed climate change with various acting forcings [23–25]. The optimal fingerprinting method (OFM) [15,26–28] aims at expressing a (vector-valued) observed climate change signal Y_k, $k = 1, \ldots, S$ as a linear combination of response patterns,

$$Y_k = \sum_{p=1}^{M} \tilde{X}_k^p \beta_p + \mathcal{R}_k, \quad \tilde{X}_k^p = X_k^p + \mathcal{Q}_k^p, \quad k = 1, \ldots, S, \quad (1)$$

where $\tilde{X}_k^p$'s are the M—typically $\leq \mathcal{O}(10)$—fingerprints, each associated with one forcing. One seeks the optimal solution in the β_p's to the (strongly underdetermined, $S \gg M$ in all cases of interest) multivariate regression problem above, where the $\mathcal{Q}_k^p$'s account for sampling or model uncertainties, and the residual $\mathcal{R}_k$ is the natural climate variability [23–28]. Usually $\mathcal{R}_k$ and $\mathcal{Q}_k^p$ are modeled as independent, normally distributed stochastic vectors with zero mean and given covariance matrices. The estimate X_k^p for the pth fingerprint is obtained by averaging across an ensemble of forced model runs, each driven by

*Contact author: v.lucarini@leicester.ac.uk

the same protocol for the pth forcing, while the other forcings are switched off [29,30].

An example follows. Let Y_k be a gridded surface temperature (T_S) anomaly field measured at S locations. Assuming that the pth forcing is the increase in the CO_2 concentration ($[CO_2]$), $\tilde{X}_k^p$ features larger anomalies at high latitudes as a result of polar amplification. If the qth forcing is a localized increase in the (reflecting) aerosol concentration, then $\tilde{X}_k^q$ is strongly spatially heterogeneous with negative values where such concentration is higher.

DA relies on Pearl causality [31]. This angle allows for estimating causal effects through an interventionist protocol [32,33]. Each fingerprint is computed using climate models—which provide the counterfactual reality—where the intervention is the selective application of one forcing. Observations-only-based DA requires strong assumptions on the link between climate change and climate variability [30].

The attribution of the signal to the pth forcing depends on the β_p confidence interval. If this includes 1 and excludes 0, we can reject the null hypothesis of no influence. The narrower this interval, the more robust the DA outcome. One has a favorable signal-to-noise ratio if either the signal is highly sensitive to the forcing or if the natural variability is weak. Attribution is also impacted by model uncertainties in estimating the fingerprints. Strong interactions between forcings can significantly skew the β's estimate. Finally, overlooking key forcings can jeopardize DA, leading to erroneous attribution.

This Letter—Traditionally, OFM relied on an empirical justification based on Eq. (1). We demonstrate that Eq. (1) can be rigorously derived from linear response theory (LRT) applied to nonequilibrium systems, a well-established framework in statistical mechanics for both deterministic [34,35] and stochastic systems [36]. The Green's function formalism plays a key role here in linking cause (forcing) and effect (observed signal).

This Letter offers a threefold contribution to climate physics. First, we establish a robust physical and dynamical foundation for OFM. This foundation allows for more advanced fingerprinting analysis, facilitating a deeper understanding of cause-and-effect relationships within the climate system. Second, our derivation provides a theoretical framework for the degenerate fingerprinting method used for tipping point detection, a crucial component of early warning systems [37–39]. Third, by leveraging nonlinear response theory [40,41], we propose an extension of OFM accounting for the nonlinear effects tied to the interplay between multiple climate forcings.

To demonstrate the real-world applicability of our theoretical advancements, we conduct detailed analyses on two climate models with varying complexity. This includes the one-dimensional Ghil-Sellers (G-S) energy balance model [42] and the open-source, low-resolution coupled climate planet simulator (PLASIM) model [43–46].

Stochastic climate dynamics and response theory—Following Hasselmann's program [47,48]—see also the recent review [49]—we consider climate models formulated as stochastic differential equations (SDEs), where the impact of unresolved scales on the scales of interest has been stochastically parametrized [50–54]. We study climate change by looking at how perturbations acting on the dynamics affect the ensemble statistics. Hence, the unperturbed dynamics is described by the following d-dimensional Itô SDE [55,56]:

$$\mathrm{d}\boldsymbol{x} = \mathbf{F}(\boldsymbol{x})\mathrm{d}t + \Sigma(\boldsymbol{x})\mathrm{d}\boldsymbol{W}_t. \tag{2}$$

Here, $\mathbf{F}$ (drift term) is a smooth vector field on $\mathbb{R}^d$, $\boldsymbol{W}_t$ denotes a p-dimensional Wiener process ($p \geq 1$), Σ is the diffusion coefficient matrix (of size $d \times p$), and $\boldsymbol{x}$ describes the climate state. The choice of the climate subcomponents included in our model and of the explicitly resolved spatial scales impacts the drift and noise laws [47–50,57,58]. Correspondingly, d can range from $\mathcal{O}(10^1)$ to $\mathcal{O}(10^8)$ [13]; yet our framework can be seamlessly applied.

We consider here time-dependent forcings impacting the resolved scales; see Ref. [59] for a more general case. We study

$$\mathrm{d}\boldsymbol{x} = \left(\mathbf{F}(\boldsymbol{x}) + \sum_{p=1}^{M} \epsilon_p g_p(t)\mathbf{G}_p(\boldsymbol{x})\right)\mathrm{d}t + \Sigma(\boldsymbol{x})\mathrm{d}\boldsymbol{W}_t, \tag{3}$$

where the "large-scale" perturbations are embodied by one or more "pattern forcing" $\mathbf{G}_p$ of $\mathbf{R}^d$, each modulated by a scalar-valued bounded function $g_p(t)$ and by a parameter $\epsilon_p \ll 1$.

The climatology of any smooth observable Ψ is the expected value under the reference climate: $\langle\Psi\rangle_0 = \int \mathrm{d}\boldsymbol{x}\rho_0(\boldsymbol{x})\Psi(\boldsymbol{x})$. Here, $\rho_0(\boldsymbol{x})$ represents the stationary probability density of the unperturbed system [Eq. (2)], solving its associated Fokker-Planck equation (FPE) [56,60]. Good choices for Ψ include essential climate variables [61], which are key physicochemical quantities observed at different spatial scales, as well as metrics used to evaluate Earth system models (ESMs) [62].

Response theory allows one to predict the system's response to perturbations from properties of the unperturbed system, as exemplified by the celebrated fluctuation-dissipation theorem (FDT) [63,64]. In a seminal article, Leith [65] suggested that if FDT can be applied in climate science, then climate change projections can be performed using the statistical properties of the natural climate variability. LRT for nonequilibrium systems [34–36] indicates that the first-order correction $\delta^{(1)}[\Psi](t)$ to the statistics $\langle\Psi\rangle_0$ is given by a suitable lagged correlation between functions of the unperturbed system [66]. The Green's function formalism enables useful insights as discussed below. It states that

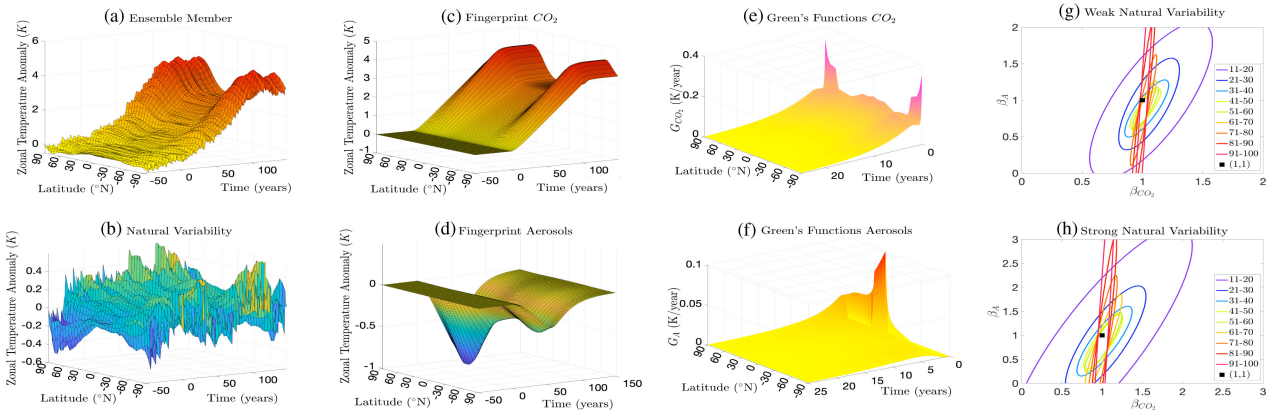

FIG. 1. OFM and link with response theory for the G-S model. The climate change signal [from one ensemble member, in (a)] is decomposed according to Eq. (1): fingerprints associated with the CO_2 and aerosol forcings [shown in (c) and (d)] plus natural variability shown in (b). The fingerprints shown in (c) and (d) are computed via response theory using Green's functions (e) and (f) in Eq. (6), respectively. The corresponding β coefficients in Eq. (1) and 95% confidence regions are computed per decade, for each ensemble member, in the cases of a weaker (g) and stronger (h) imposed natural variability. See text and Appendix B for details.

$$\delta^{(1)}[\Psi](t) = \sum_{p=1}^{M} \epsilon_p \int_{-\infty}^{t} \mathrm{d}s\, \mathcal{G}^p_\Psi(t-s) g_p(s), \qquad (4)$$

where the $\mathcal{G}^p_\Psi$ are the (causal) Green's functions given in Appendix A. The quality of approximation of the actual change in statistics by $\delta^{(1)}[\Psi](t)$ depends on many factors. It deteriorates with strong perturbations and/or if the unperturbed system has slow decay of correlations [40,41,59,67,68], such as encountered when approaching tipping points [59].

Connecting OFM to LRT via Green's functions—We now show how the structural equation of OFM—Eq. (1)—can be derived from LRT. Let Ψ_k be a climatic observable with reference climatology $\langle \Psi_k \rangle_0$. We want to assess the contribution of the large-scale forcing, $g_p(t)\mathbf{G}_p(\boldsymbol{x})$, to the anomaly signal $Y_k(t) = \Psi_k(t) - \langle \Psi_k \rangle_0$. Trivially,

$$Y_k(t) = \Psi_k(t) - \langle \Psi_k \rangle_{\rho^t_\epsilon} + \langle \Psi_k \rangle_{\rho^t_\epsilon} - \langle \Psi_k \rangle_0, \qquad (5)$$

where $\langle \Psi_k \rangle_{\rho^t_\epsilon}$ is the ensemble average with respect to ρ^t_ϵ, which evolves according to the FPE associated with Eq. (3). The difference term between the ensemble averages in Eq. (5) is the focus of response theory. By approximating it by Eq. (4), we obtain

$$Y_k(t) = \sum_{p=1}^{M} \tilde{X}^p_k(t) + \mathcal{R}_k(t), \qquad k = 1, \ldots, N,$$

$$\text{with} \quad \tilde{X}^p_k(t) = \epsilon_p \int_{-\infty}^{t} \mathrm{d}s \mathcal{G}^p_{\Psi_k}(t-s) g_p(s). \qquad (6)$$

The terms $\tilde{X}^p_k(t)$ account for the forced variability, while $\mathcal{R}_k(t) = \Psi_k(t) - \langle \Psi \rangle_{\rho^t_\epsilon}$ is a stochastic vector whose correlations are described by the statistical state ρ^t_ϵ. When $\Sigma = 0$, this statistical state is carried by the system's pullback attractor at time t [69–71], also known as the snapshot attractor [72].

Let us compare Eqs. (6) and (1). The fingerprints $\tilde{X}^p_k(t)$ are constructed as convolution products of Green's function relative to the considered forcing with the corresponding time modulation. Hence, it makes perfect sense to approximate them—as customarily done in most DA analyses—as ensemble average of the response signals obtained by applying selectively the corresponding forcing. This mirrors an efficient method used for estimating Green's functions [73–75].

In Eq. (6) there is a conspicuous absence of the β_p's. This is actually a key element of our derivation: LRT predicts that all the β_p's should be unitary, apart from (marginal) uncertainty, when DA is performed with a single "perfect" model.

Detection and attribution of climate change: Examples—Energy balance model The G-S model [42] can be seen as the "hydrogen atom model" of the climate system and is a one-dimensional space-time reaction-diffusion model. It describes the evolution of surface temperature T_S across different latitudes. To incorporate natural variability, we force the model with white noise. Details on the model and the DA experiments are provided in Appendix B.

The model is discretized in space, evaluating T_S at $d = 37$ latitudes. It is also subjected to $M = 2$ forcings. The first forcing mimics a $[CO_2]$ increase, ramping from a reference value to a final value over 100 years. The second forcing represents a localized increase in atmospheric aerosols, peaking around 50 years in the low-to-mid latitudes of the Northern Hemisphere and inducing a net cooling effect.

The climate change signal is the decadal $S = d = 37$-dimensional vector of T_S anomalies (across latitudes) for 10 decades following the start of the forcing. Along the lines discussed in Appendix D, we compute Green's

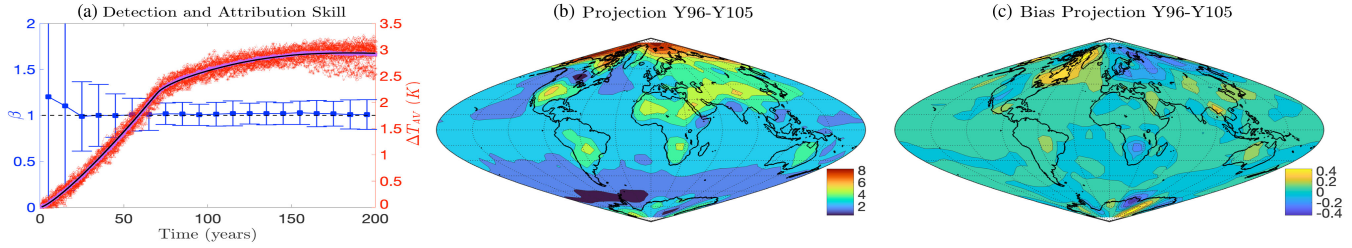

FIG. 2. DA of climate change for PLASIM forced experiments. (a) 95% confidence interval for the β factor (blue) of the CO_2 forcing fingerprint computed using Green's functions. The uncertainty decreases as the signal-to-noise ratio increases, as shown by the globally averaged T_S change ΔT_{AV} (red), ensemble average (thick magenta line; the black line shows the prediction), and ensemble members (thin red lines). (b) Decadal average of the $\delta^{(1)}[T_S]$ (in kelvin) field projection [Eq. (4)] during year 96 to 105 following the start of the forcing. (c) Difference (in kelvin) between the $\langle T_S \rangle$ anomaly of the forced runs and (b) for the year 96–105 average. The coastline is only indicative as it is at higher resolution than the land-sea mask used by the model. See text and Appendix C for details.

function for T_S at each latitude for both forcings; see Figs. 1(e) and 1(f). We use Green's functions to compute the fingerprints via Eq. (6); see Figs. 1(c) and 1(d). Following Eq. (1) we then apply the OFM for each individual forced run; see Fig. 1(a). We derive for each decade estimates of the coefficients β_{CO_2} and β_A (for aerosols). We repeat the protocol for $N = 100$ forced runs, thus obtaining confidence intervals for the β's, see Figs. 1(g) and 1(h).

As the CO_2 and aerosol forcings strengthen over the first half of the century, the β's 95% confidence region shrinks. We obtain attribution to both forcings between years 30 and 50. However, as the aerosol signal weakens in the latter half of the century, attribution for the CO_2 forcing becomes clearer, while the aerosol signal becomes indistinguishable from natural variability. This distinction arises because, despite their partially offsetting global effects, the CO_2 and aerosol forcings have different spatial patterns. Additionally, weaker natural variability simplifies the attribution process by reducing the noise level, as shown in Fig. 1(g). It is noteworthy that the confidence intervals for β consistently center around 1.

PLASIM We perform DA using the climate model PLASIM, which features $d = \mathcal{O}(10^5)$ degrees of freedom [43,46]. We study the effect of $M = 1$ forcing, namely, the increase in $[CO_2]$. Along the lines of [73,74], our protocol entails an annual 1% increase of $[CO_2]$ from 360 ppm up to doubling, while afterward $[CO_2]$ is kept constant. We run an ensemble of $N = 40$ forced simulations.

The climate change signals of interest are the decadal averages of the T_S evaluated at the $S = 64 \times 32 = 2048$ grid points. We use LRT to compute the fingerprints. We then derive an estimate of β_{CO_2} for each of the forced runs, thus deriving the β_{CO_2} 95% confidence interval; see Appendixes C and D. Figure 2(a) shows that, as the global warming progresses, the β confidence interval shrinks, while being always centered around 1. The attribution of the anomaly signal to the CO_2 forcing is statistically robust already after the second decade. LRT-based DA is effective because LRT-based climate change projections are accurate: Fig. 2(b) shows the decadal T_S anomaly one century after the start of the forcing computed via LRT, while Fig. 2(c) shows the (modest) bias with respect to the corresponding ensemble average of the forced runs.

Nonlinear fingerprints—The standard OFM relies on the hypothesis of linearity of the response. Response theory can be extended to higher-order terms [40,41]. The second-order correction to the expected value of a general observable Ψ can be written as [76,77]

$$\delta^{(2)}[\Psi](t) = \sum_{p,q=1}^{M} \epsilon_p \epsilon_q \int_{\infty}^{t} \int_{\infty}^{t} d\tau_1 \, d\tau_2 \mathcal{G}_{\Psi}^{p,q}(t-\tau_1, t-\tau_2) \times g_p(\tau_1) g_q(\tau_2), \tag{7}$$

where $\mathcal{G}_{\Psi}^{p,q}(t_1, t_2)$ is the second-order Green's function (causal in both time arguments) describing the joint effect of the pth and the qth forcing. One then has $\delta[\Psi](t) = \delta^{(1)}[\Psi](t) + \delta^{(2)}[\Psi](t) + h.o.t.$, where $h.o.t.$ accounts for $o(\epsilon^2)$ terms. By accounting for $\delta^{(2)}[\Psi](t)$ in the ensemble fluctuations $\langle \Psi \rangle_{\rho_\epsilon^t} - \langle \Psi_k \rangle_0$ in Eq. (5), we generalize Eq. (6) as follows:

$$Y_k(t) = \sum_{p=1}^{M} \tilde{X}_k^p(t) + \sum_{\ell=1}^{M^2} \tilde{Z}_k^\ell(t) + \mathcal{R}_k(t), \quad k = 1, \ldots, N, \tag{8}$$

where $\tilde{Z}_k^\ell(t) = \epsilon_p \epsilon_q \int d\tau_1 d\tau_2 \mathcal{G}_{\Psi}^{p,q}(t-\tau_1, t-\tau_2) g_p(\tau_1) g_q(\tau_2)$ with $\ell = p + M(q-1)$. Higher-order contributions can be constructed in a similar fashion. Indeed, Eq. (7) and its generalizations can be viewed as Volterra integrals, where Green's functions are interpreted as Volterra kernels [78,79].

We then propose a generalization of the OFM able to deal with interacting responses by seeking a *linear* regression for

$$Y_k = \sum_{p=1}^{M} \tilde{X}_k^p \beta_p + \sum_{\ell=1}^{M^2} \tilde{Z}_k^\ell \gamma_\ell + \mathcal{R}_k, \qquad k = 1, \ldots, N,$$
$$\tilde{X}_k^p = X_k^p + \mathcal{Q}_k^p, \qquad \tilde{Z}_k^\ell = Z_k^p + \mathcal{P}_k^\ell. \tag{9}$$

The factors β's and γ's are unitary in the perfect model scenario, see Eq. (7). The extra M^2 fingerprints compared to Eq. (1) can be constructed from selected climate model runs by suitable variations of the ϵ_p's [41,80]. A data-driven route leads to constructing the nonlinear response operator as feed forward neural networks [81]. After training, the kernels can be computed from the weights and biases of the network [82].

Fingerprinting near tipping points—Using the formalism of the Kolmogorov modes [83], which are the stochastic counterparts of the Koopman modes [84–86], we write $\mathcal{G}^p_\Psi$ as a sum of terms, each linked with a specific mode of natural climate variability. This expansion clarifies the link between free and forced climate variability [65,87],

$$\mathcal{G}^p_\Psi(t) \approx \sum_{j=1}^{V} \sum_{\ell=0}^{m_j-1} \alpha_j^{\ell,p}(\Psi) \frac{1}{\ell!} e^{\lambda_j t} t^\ell, \qquad t \geq 0, \quad (10)$$

with $\mathcal{G}^p_\Psi(t) = 0$ when $t \leq 0$. This formula follows from the expansion of (temporal) correlations in terms of Kolmogorov modes and relies on neglecting the continuous spectrum [49,59,83]. Here V is, in general, infinite and m_j is the multiplicity of the jth eigenmode φ_j. These are the eigenfunctions of the Kolmogorov operator $\mathcal{K}_\Sigma$ associated with Eq. (2) [see Eq. (A4) in Appendix A]. The λ_j are the corresponding eigenvalues, also known as Ruelle-Pollicott resonances [88]. Often a finite number of these modes are approximated from time series in a reduced state space using Markov approximations [89,90] or delayed-embedding techniques [86].

They encapsulate the decay rate and frequency of oscillation of the natural modes of variability (Sec. 2.3 in [83]), while the factors $\alpha_j^{\ell,p}$ depend on the applied forcing and choice of observable [49]. The spectral gap $\gamma = \mathrm{Re}(\lambda_1)$ associated with the slowest decaying mode φ_1 indicates the proximity to tipping. Rough dependence of statistics on parameters is found as $\gamma \to 0$ [89]. At a tipping point, $\gamma \to 0$, and Eq. (10) indicates that any Green's function decays subexponentially, irrespective of the observable and forcing, unless the corresponding α coefficient vanishes. At criticality, the response may diverge [59,89,91,92]. Hence, if either the climate system or the model used for fingerprinting is close to a tipping point, DA via OFM might experience major uncertainties and biases.

If each isolated eigenvalue has unitary multiplicity, the time-lagged correlation for Ψ is $\mathrm{Corr}_\Psi(t) = \sum_{j=1}^{N} a_j(\Psi) e^{\lambda_j |t|}$ with $a_j(\Psi)$ given in Corollary 1 of [83]. If $\gamma \to 0$, $\mathrm{Corr}_\Psi(t)$ decays subexponentially when Ψ has a nonvanishing projection onto φ_1. This is the so-called critical slowing down associated with tipping behavior [10,93], which was originally discovered for continuous phase transitions [94]. If $\Psi \propto \varphi_1$, $\mathrm{Corr}_\Psi(t) \propto e^{\lambda_1 |t|}$ and φ_1 is the degenerate fingerprinting's critical mode [37], encoding the "natural" tipping observable.

Discussion and perspectives—OFM has been instrumental in shaping modern climate change science and has had significant societal impact [15,16]. However, recent criticisms highlight potential underestimates of uncertainties and even question the statistical basis of OFM [95–97]. This Letter addresses this issue by deriving OFM equations for DA from LRT for nonequilibrium systems, thus providing a solid physical and dynamical foundation for OFM. In particular, the causality principle embodied in Green's functions aligns closely with Pearl's interventionist angle [31], which is key for DA studies. Our derivation extends OFM applicability to virtually any complex system—ecosystems, quantitative social sciences, finance—where attributing observed changes to multiple forcings is desired.

This Letter offers several key insights for fingerprint analysis in climate science as listed below.

Systematic attribution In the perfect model scenario where all forcings are included, this Letter explains why LRT provides accurate attribution in linear regimes. It also clarifies why the weighting factors (β's) are equal to 1 in such a scenario. However, a crucial limitation is identified: mismatches between the model's and the climate system's natural variability patterns (Kolmogorov modes) can distort Green's functions, affecting fingerprint accuracy across different timescales. This is especially concerning near tipping points. Future research will focus on linking natural and forced variability through Kolmogorov mode analysis.

Combining models Combining fingerprint estimates from different models can be problematic due to differences in their Kolmogorov modes, potentially leading to errors, especially near tipping points. Instead, using an ensemble approach with a single model aligns with the mathematical framework.

Critical modes and early warning signals Since Kolmogorov modes reflect proximity to critical transitions, they offer a reliable basis for degenerate fingerprinting, which is key for defining early warning indicators for tipping points [37] and for defining the critical mode of variability.

Nonlinear extension to the OFM Nonlinear response theory allows for a more powerful OFM framework. Even with significant nonlinearities due to strong forcing interactions, the method remarkably leads to a linear regression problem. These nonlinearities manifest as additional fingerprints, providing practical benefits for climate research and other complex systems. Future work will explore these applications in detail.

To demonstrate the theory's potential, we applied its key findings to perform DA of simulated climate change on two models: the simple yet historically relevant and physically informative one-dimensional G-S energy balance model and the more realistic three-dimensional coupled climate model PLASIM. The results showcase how LRT and OFM work hand in hand. Green's functions enable computation

of accurate fingerprint across timescales, while the signal-to-noise ratio heavily influences attribution robustness. Our approach paves the way for applying LRT to compute fingerprints in state-of-the-art ESMs [62]. The effectiveness of LRT for climate change projections in ESMs has already been established [75]. Successfully implementing LRT in this way could significantly improve our understanding of the current climate crisis.

Acknowledgments—The authors wish to thank Reyk Börner for support on the numerical simulations performed with PLASIM and two anonymous reviewers for providing very useful constructive criticism and insightful comments. V. L. acknowledges the support by the EU Horizon 2020 project TiPES (Grant No. 820970), the EU Horizon Europe project ClimTIP (Grant No. 101137601), the MSCA ITN CriticalEarth (Grant No. 956170), and by EPSRC project LINK (Grant No. EP/Y026675/1). M. D. C. acknowledges the European Research Council under the EU Horizon 2020 research and innovation program (Grant No. 810370). This work has been also supported by the Office of Naval Research (ONR) Multidisciplinary University Research Initiative (MURI) Grant No. N00014-20-1-2023 and by the National Science Foundation Grant No. DMS-2407484.

Data availability—All data needed to evaluate the conclusions of this work can be found at [124]. The PLASIM-LSG code is available at [98].

This work initiated from discussions between V. L. and M. D. C. Both authors conceptualized the approach and the mathematical details. V. L. performed the numerical calculations behind Fig. 1 and performed the data analysis leading to Figs. 2 and 3. V. L. and M. D. C. interpreted the main results and wrote the Letter.

The authors declare that they have no competing interests.

[1] J. P. Peixoto and A. H. Oort, *Physics of Climate* (AIP Press, New York, 1992).
[2] V. Lucarini, R. Blender, C. Herbert, F. Ragone, S. Pascale, and J. Wouters, Mathematical and physical ideas for climate science, Rev. Geophys. **52**, 809 (2014).
[3] M. Ghil, M. Chekroun, and E. Simonnet, Climate dynamics and fluid mechanics: Natural variability and related uncertainties, Physica D (Amsterdam) **237**, 2111 (2008).
[4] J. Neelin, *Climate Change and Climate Modeling* (Cambridge University Press, Cambridge, England, 2010).
[5] D. Stammer, A. Bracco, P. Braconnot, G. Brasseur, S. Griffies, and E. Hawkins, Science directions in a post COP21 world of transient climate change: Enabling regional to local predictions in support of reliable climate information, Earth's Future **6**, 1498 (2018).
[6] A. S. von der Heydt, P. Ashwin, C. D. Camp, M. Crucifix, H. A. Dijkstra, P. Ditlevsen, and T. M. Lenton, Quantification and interpretation of the climate variability record, Global Planet. Change **197**, 103399 (2021).
[7] H. Liu, I. Koren, O. Altaratz, and M. D. Chekroun, Opposing trends of cloud coverage over land and ocean under global warming, Atmos. Chem. Phys. **23**, 6559 (2023).
[8] T. Lenton, H. Held, E. Kriegler, J. Hall, W. Lucht, S. Rahmstorf, and H. Schellnhuber, Tipping elements in the Earth's climate system, Proc. Natl. Acad. Sci. U.S.A. **105**, 1786 (2008).
[9] N. Boers, N. Marwan, H. M. J. Barbosa, and J. Kurths, A deforestation-induced tipping point for the South American monsoon system, Sci. Rep. **7**, 41489 (2017).
[10] N. Boers and M. Rypdal, Critical slowing down suggests that the western greenland ice sheet is close to a tipping point, Proc. Natl. Acad. Sci. U.S.A. **118**, e2024192118 (2021).
[11] J. Lohmann and P. D. Ditlevsen, Risk of tipping the overturning circulation due to increasing rates of ice melt, Proc. Natl. Acad. Sci. U.S.A. **118**, e2017989118 (2021).
[12] P. Ditlevsen and S. Ditlevsen, Warning of a forthcoming collapse of the atlantic meridional overturning circulation, Nat. Commun. **14**, 4254 (2023).
[13] M. Ghil and V. Lucarini, The physics of climate variability and climate change, Rev. Mod. Phys. **92**, 035002 (2020).
[14] D. H. Rothman, Thresholds of catastrophe in the Earth system, Sci. Adv. **3**, e1700906 (2017).
[15] IPCC, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, England, 2014).
[16] IPCC, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, England, 2021).
[17] P. A. Stott, N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang, and F. Zwiers, Detection and attribution of climate change: A regional perspective, WIREs Clim. Change **1**, 192 (2010).
[18] D. R. Easterling, K. E. Kunkel, M. F. Wehner, and L. Sun, Detection and attribution of climate extremes in the observed record, Weather Clim. Extremes **11**, 17 (2016), Observed and projected (longer-term) changes in weather and climate extremes.
[19] M. Ghil, P. Yiou, S. Hallegatte, B. D. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov, O. Mestre, C. Nicolis, H. W. Rust, P. Shebalin, M. Vrac, A. Witt, and I. Zaliapin, Extreme events: Dynamics, statistics and prediction, Nonlinear Processes Geophys. **18**, 295 (2011).
[20] IPCC, *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge and New York, 2012).
[21] W. J. Ripple, C. Wolf, T. M. Newsome, J. W. Gregg, T. M. Lenton, I. Palomo, J. A. J. Eikelboom, B. E. Law, S. Huq,

P. B. Duffy, and J. Rockström, World Scientists' warning of a climate emergency 2021, BioScience **71**, 894 (2021).
[22] B. Rodgers, The climate emergency demands scientists take action and here's how, Nat. Rev. Phys. **5**, 549 (2023).
[23] K. Hasselmann, Optimal fingerprints for the detection of time-dependent climate change, J. Clim. **6**, 1957 (1993).
[24] G. C. Hegerl, H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, Detecting greenhouse-gas-induced climate change with an optimal fingerprint method, J. Clim. **9**, 2281 (1996).
[25] K. Hasselmann, Multi-pattern fingerprint method for detection and attribution of climate change, Clim. Dyn. **13**, 601 (1997).
[26] M. Allen and S. Tett, Checking for model consistency in optimal fingerprinting, Clim. Dyn. **15**, 419 (1999).
[27] M. R. Allen and P. A. Stott, Estimating signal amplitudes in optimal fingerprinting Part I: Theory, Clim. Dyn. **21**, 477 (2003).
[28] A. Hannart, A. Ribes, and P. Naveau, Optimal fingerprinting under multiple sources of uncertainty, Geophys. Res. Lett. **41**, 1261 (2014).
[29] D. Stone, M. R. Allen, F. Selten, M. Kliphuis, and P. A. Stott, The detection and attribution of climate change using an ensemble of opportunity, J. Clim. **20**, 504 (2007).
[30] G. Hegerl and F. Zwiers, Use of models in detection and attribution of climate change, Wiley Interdiscip. Rev. **2**, 570 (2011).
[31] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Francisco, California, 2009), Example for Explaning away.
[32] A. Hannart, J. Pearl, F. E. L. Otto, P. Naveau, and M. Ghil, Causal counterfactual theory for the attribution of weather and climate-related events, Bull. Am. Meteorol. Soc. **97**, 99 (2016).
[33] D. J. Wuebbles, D. W. Fahey, and K. A. Hibbard, *Climate Science Special Report: Fourth National Climate Assessment* (U.S. Global Change Research Program, Washington, DC, 2017), Vol. I.
[34] D. Ruelle, General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium, Phys. Lett. A **245**, 220 (1998).
[35] D. Ruelle, A review of linear response theory for general differentiable dynamical systems, Nonlinearity **22**, 855 (2009).
[36] M. Hairer and A. J. Majda, A simple framework to justify linear response theory, Nonlinearity **23**, 909 (2010).
[37] H. Held and T. Kleinen, Detection of climate system bifurcations by degenerate fingerprinting, Geophys. Res. Lett. **31**, L23207 (2004).
[38] T. M. Lenton, V. N. Livina, V. Dakos, E. H. van Nes, and M. Scheffer, Early warning of climate tipping points from critical slowing down: Comparing methods to improve robustness, Phil. Trans. R. Soc. A **370**, 1185 (2012).
[39] C. Boettner and N. Boers, Critical slowing down in dynamical systems driven by nonstationary correlated noise, Phys. Rev. Res. **4**, 013230 (2022).
[40] D. Ruelle, Nonequilibrium statistical mechanics near equilibrium: Computing higher-order terms, Nonlinearity **11**, 5 (1998).
[41] V. Lucarini, Evidence of dispersion relations for the nonlinear response of the Lorenz 63 system, J. Stat. Phys. **134**, 381 (2009).
[42] M. Ghil, Climate stability for a Sellers-type model, J. Atmos. Sci. **33**, 3 (1976).
[43] K. Fraedrich, H. Jansen, E. Kirk, U. Luksch, and F. Lunkeit, The planet simulator: Towards a user friendly model, Meteorol. Z. **14**, 299 (2005).
[44] V. Lucarini, K. Fraedrich, and F. Lunkeit, Thermodynamic analysis of snowball Earth hysteresis experiment: Efficiency, entropy production and irreversibility, Q. J. R. Meteorol. Soc. **136**, 2 (2010).
[45] V. Lucarini, K. Fraedrich, and F. Lunkeit, Thermodynamics of climate change: Generalized sensitivities, Atmos. Chem. Phys. **10**, 9729 (2010).
[46] O. Mehling, K. Bellomo, M. Angeloni, C. Pasquero, and J. von Hardenberg, High-latitude precipitation as a driver of multicentennial variability of the AMOC in a climate model of intermediate complexity, Clim. Dyn. **61**, 1519 (2023).
[47] K. Hasselmann, Stochastic climate models Part I. Theory, Tellus **28**, 473 (1976).
[48] P. Imkeller and J. S. von Storch, *Stochastic Climate Models* (Birkhauser, Basel, 2001).
[49] V. Lucarini and M. D. Chekroun, Theoretical tools for understanding the climate crisis from Hasselmann's programme and beyond, Nat. Rev. Phys. **5**, 744 (2023).
[50] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden, A mathematical framework for stochastic climate models, Commun. Pure Appl. Math. **54**, 891 (2001).
[51] A. J. Majda and J. Harlim, Physics constrained nonlinear regression models for time series, Nonlinearity **26**, 201 (2013).
[52] D. Kondrashov, M. D. Chekroun, and M. Ghil, Data-driven non-Markovian closure models, Physica D (Amsterdam) **297**, 33 (2015).
[53] J. Berner, U. Achatz, L. Batté, L. Bengtsson, A. de la Cámara, H. M. Christensen, M. Colangeli, D. R. B. Coleman, D. Crommelin, S. I. Dolaptchiev, and C. L. Franzke, Stochastic parameterization: Toward a new view of weather and climate models, Bull. Am. Meteorol. Soc. **98**, 565 (2017).
[54] M. Santos Gutiérrez, V. Lucarini, M. D. Chekroun, and M. Ghil, Reduced-order models for coupled dynamical systems: Data-driven methods and the Koopman operator, Chaos **31**, 053116 (2021).
[55] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
[56] G. A. Pavliotis, *Stochastic Processes and Applications* (Springer, New York, 2014), Vol. 60.
[57] D. Kondrashov, M. Chekroun, and P. Berloff, Multiscale Stuart-Landau emulators: Application to wind-driven ocean gyres, Fluids **3**, 21 (2018).
[58] B. Saltzman, *Dynamical Paleoclimatology: Generalized Theory of Global Climate Change* (Academic Press, New York, 2001).
[59] M. Santos Gutiérrez and V. Lucarini, On some aspects of the response to stochastic and deterministic forcings, J. Phys. A **55**, 425002 (2022).

[60] H. Risken, *The Fokker-Planck Equation*, 2nd ed. (Springer, New York, 1989).

[61] S. Bojinski, M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp, The concept of essential climate variables in support of climate research, applications, and policy, Bull. Am. Meteorol. Soc. **95**, 1431 (2014).

[62] V. Eyring *et al.*, Earth system model evaluation tool (esmvaltool) v2.0—an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in cmip, Geosci. Model Dev. **13**, 3383 (2020).

[63] R. Kubo, The fluctuation-dissipation theorem, Rep. Prog. Phys. **29**, 255 (1966).

[64] U. M. B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, Fluctuation–dissipation: Response theory in statistical physics, Phys. Rep. **461**, 111 (2008).

[65] C. E. Leith, Climate response and fluctuation dissipation, J. Atmos. Sci. **32**, 2022 (1975).

[66] A. Majda, R. V. Abramov, and M. J. Grote, *Information Theory and Stochastics for Multiscale Nonlinear Systems* (American Mathematical Society, Providence, 2005), Vol. 25.

[67] V. Lucarini, Response operators for Markov processes in a finite state space: Radius of convergence and link to the response theory for axiom a systems, J. Stat. Phys. **162**, 312 (2016).

[68] M. Santos Gutiérrez and V. Lucarini, Response and sensitivity using Markov chains, J. Stat. Phys. **179**, 1572 (2020).

[69] M. D. Chekroun, E. Simonnet, and M. Ghil, Stochastic climate dynamics: Random attractors and time-dependent invariant measures, Physica D (Amsterdam) **240**, 1685 (2011).

[70] M. D. Chekroun, M. Ghil, and J. D. Neelin, Pullback attractor crisis in a delay differential ENSO model, in *Advances in Nonlinear Geosciences*, edited by A. Tsonis (Springer, New York, 2018), pp. 1–33.

[71] M. D. Chekroun, I. Koren, H. Liu, and H. Liu, Generic generation of noise-driven chaos in stochastic time delay systems: Bridging the gap with high-end simulations, Sci. Adv. **8**, eabq7137 (2022).

[72] T. Tél, T. Bódai, G. Drótos, T. Haszpra, M. Herein, B. Kaszás, and M. Vincze, The theory of parallel climate realizations, J. Stat. Phys. **179**, 1496 (2020).

[73] F. Ragone, V. Lucarini, and F. Lunkeit, A new framework for climate sensitivity and prediction: A modelling perspective, Clim. Dyn. **46**, 1459 (2016).

[74] V. Lucarini, F. Ragone, and F. Lunkeit, Predicting climate change using response theory: Global averages and spatial patterns, J. Stat. Phys. **166**, 1036 (2017).

[75] V. Lembo, V. Lucarini, and F. Ragone, Beyond forcing scenarios: Predicting climate change through response operators in a coupled general circulation model, Sci. Rep. **10**, 8668 (2020).

[76] V. Lucarini, Response theory for equilibrium and non-equilibrium statistical mechanics: Causality and generalized Kramers-Kronig relations, J. Stat. Phys. **131**, 543 (2008).

[77] V. Lucarini and M. Colangeli, Beyond the linear fluctuation-dissipation theorem: The role of causality, J. Stat. Mech. (2012) P05013.

[78] M. O. Franz and B. Schökopf, A unifying view of Wiener and Volterra theory and polynomial kernel regression, Neural Comput. **18**, 3097 (2006).

[79] S. Orcioni, Improving the approximation ability of Volterra series identified with a cross-correlation method, Nonlinear Dyn. **78**, 2861 (2014).

[80] A. Gritsun and V. Lucarini, Fluctuations, response, and resonances in a simple atmospheric model, Physica D (Amsterdam) **349**, 62 (2017).

[81] T. Bodai, V. Lembo, S. Aneesh, M. Ishuzu, M. O. Franz, and E.-S. Chung, Modest geoengineering side effects predicted by an emulator, Research Square preprint, 2023, 10.21203/rs.3.rs-3308863/v3.

[82] J. Wray and G. G. R. Green, Calculation of the Volterra kernels of non-linear dynamic systems using an artificial neural network, Biol. Cybern. **71**, 187 (1994).

[83] M. Chekroun, A. Tantet, H. Dijkstra, and J. D. Neelin, Ruelle-Pollicott resonances of stochastic systems in reduced state space. Part I: Theory, J. Stat. Phys. **179**, 1366 (2020).

[84] M. Budišić, R. Mohr, and I. Mezić, Applied Koopmanism, Chaos **22**, 047510 (2012).

[85] J. Kutz, S. Brunton, B. Brunton, and J. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*, Other Titles in Applied Mathematics (Society for Industrial and Applied Mathematics, Philadelphia, 2016).

[86] G. Froyland, D. Giannakis, B. R. Lintner, M. Pike, and J. Slawinska, Spectral analysis of climate dynamics with operator-theoretic approaches, Nat. Commun. **12**, 6570 (2021).

[87] S. Corti, F. Molteni, and T. N. Palmer, Signature of recent climate change in frequencies of natural atmospheric circulation regimes, Nature (London) **398**, 799 (1999).

[88] J. P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. **57**, 617 (1985).

[89] M. D. Chekroun, J. D. Neelin, D. Kondrashov, J. C. McWilliams, and M. Ghil, Rough parameter dependence in climate models: The role of Ruelle-Pollicott resonances, Proc. Natl. Acad. Sci. U.S.A. **111**, 1684 (2014).

[90] A. Tantet, M. Chekroun, J. Neelin, and H. Dijkstra, Ruelle–Pollicott resonances of stochastic systems in reduced state space. Part III: Application to the Cane–Zebiak model of the El Ni no–Southern oscillation, J. Stat. Phys. **179**, 1449 (2020).

[91] A. Tantet, V. Lucarini, and H. A. Dijkstra, Resonances in a chaotic attractor crisis of the Lorenz flow, J. Stat. Phys. **170**, 584 (2018).

[92] P. Ashwin and A. S. von der Heydt, Extreme sensitivity and climate tipping points, J. Stat. Phys. **179**, 1531 (2019).

[93] M. Scheffer, S. R. Carpenter, T. M. Lenton, J. Bascompte, W. Brock, V. Dakos, J. Van de Koppel, I. A. Van de Leemput, S. A. Levin, E. H. Van Nes *et al.*, Anticipating critical transitions, Science **338**, 344 (2012).

[94] H. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Clarendon Press, New York, 1971).

[95] Y. Li, K. Chen, J. Yan, and X. Zhang, Uncertainty in optimal fingerprinting is underestimated, Environ. Res. Lett. **16**, 084043 (2021).

[96] R. McKitrick, Checking for model consistency in optimal fingerprinting: A comment, Clim. Dyn. **58**, 405 (2022).

[97] H. Chen, S. X. Chen, and M. Mu, A statistical review on the optimal fingerprinting approach in climate change studies, Clim. Dyn. **62**, 1439 (2024).
[98] B. Shifter, J. von Hardenberg, H. Borth, and Frank, PLASIMLSG, jhardenberg/PLASIM: Plasim-LSG TNG, Zenodo (2020), 10.5281/zenodo.4041461.
[99] R. V. Abramov and A. J. Majda, Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems, Nonlinearity **20**, 2793 (2007).
[100] G. R. North, R. E. Bell, and J. W. Hardin, Fluctuation dissipation in a general circulation model, Clim. Dyn. **8**, 259 (1993).
[101] I. Cionni, G. Visconti, and F. Sassi, Fluctuation dissipation theorem in a general circulation model, Geophys. Res. Lett. **31**, L09206 (2004).
[102] P. L. Langen and V. A. Alexeev, Estimating $2 \times CO_2$ warming in an aquaplanet GCM using the fluctuation-dissipation theorem, Geophys. Res. Lett. **32**, L23708 (2005).
[103] A. Gritsun and G. Branstator, Climate response using a three-dimensional operator based on the fluctuation–dissipation theorem, J. Atmos. Sci. **64**, 2558 (2007).
[104] V. Lucarini and S. Sarno, A statistical mechanical approach for the computation of the climatic response to general forcings, Nonlinear Processes Geophys. **18**, 7 (2011).
[105] G. L. Torres Mendonca, J. Pongratz, and C. H. Reick, Identification of linear response functions from arbitrary perturbation experiments in the presence of noise—part 1: Method development and toy model demonstration, Nonlinear Processes Geophys. **28**, 501 (2021).
[106] R. Bastiaansen, H. A. Dijkstra, and A. S. v. d. Heydt, Projections of the transient state-dependency of climate feedbacks, Geophys. Res. Lett. **48**, e2021GL094670 (2021).
[107] V. Lucarini, Revising and extending the linear response theory for statistical mechanical systems: Evaluating observables as predictors and predictands, J. Stat. Phys. **173**, 1698 (2018).
[108] T. Bódai, V. Lucarini, F. Lunkeit, and R. Boschi, Global instability in the Ghil-Sellers model, Clim. Dyn. **44**, 3361 (2015).
[109] V. Lembo, F. Lunkeit, and V. Lucarini, Thediato (v1.0)—a new diagnostic tool for water, energy and entropy budgets in climate models, Geosci. Model Dev. **12**, 3805 (2019).
[110] M. Roscher, F. Stordal, and H. Svensen, The effect of global warming and global cooling on the distribution of the latest permian climate zones, Palaeogeogr. Palaeoclimatol. Palaeoecol. **309**, 186 (2011).
[111] E. Hertwig, F. Lunkeit, and K. Fraedrich, Low-frequency climate variability of an aquaplanet, Theor. Appl. Climatol. **121**, 459 (2015).
[112] K. Fraedrich and F. Lunkeit, Diagnosing the entropy budget of a climate model, Tellus A **60**, 921 (2008).
[113] R. Boschi, V. Lucarini, and S. Pascale, Bistability of the climate around the habitable zone: A thermodynamic investigation, Icarus **226**, 1724 (2013).
[114] A. Paradise, E. Macdonald, K. Menou, C. Lee, and B. L. Fan, ExoPlaSim: Extending the planet simulator for exoplanets, Mon. Not. R. Astron. Soc. **511**, 3272 (2022).
[115] A. Tantet, V. Lucarini, F. Lunkeit, and H. A. Dijkstra, Crisis of the chaotic attractor of a climate model: A transfer operator approach, Nonlinearity **31**, 2221 (2018).
[116] F. Ragone, J. Wouters, and F. Bouchet, Computation of extreme heat waves in climate models using a large deviation algorithm, Proc. Natl. Acad. Sci. U.S.A. **115**, 24 (2018).
[117] J. Wouters, R. K. H. Schiemann, and L. C. Shaffrey, Rare event simulation of extreme european winter rainfall in an intermediate complexity climate model, J. Adv. Model. Earth Syst. **15**, e2022 MS003537 (2023).
[118] G. Margazoglou, T. Grafke, A. Laio, and V. Lucarini, Dynamical landscape and multistability of a climate model, Proc. R. Soc. A **477**, 20210019 (2021).
[119] A. S. von der Heydt *et al.*, Lessons on climate sensitivity from past climate changes, Curr. Clim. Change Rep. **2**, 148 (2016).
[120] S. Drijfhout, G. J. van Oldenborgh, and A. Cimatoribus, Is a decline of AMOC causing the warming hole above the North Atlantic in observed and modeled warming patterns?, J. Clim. **25**, 8373 (2012).
[121] H. L. Bryden, W. E. Johns, B. A. King, G. McCarthy, E. L. McDonagh, B. I. Moat, and D. A. Smeed, Reduction in ocean heat transport at 26 °N since 2008 cools the Eastern Subpolar Gyre of the North Atlantic Ocean, J. Clim. **33**, 1677 (2020).
[122] T. Kuhlbrodt, A. Griesel, M. Montoya, A. Levermann, M. Hofmann, and S. Rahmstorf, On the driving processes of the Atlantic meridional overturning circulation, Rev. Geophys. 45, RG2001 (2007).
[123] R. M. van Westen, M. Kliphuis, and H. A. Dijkstra, Physics-based early warning signal shows that AMOC is on tipping course, Sci. Adv. **10**, eadk1189 (2024).
[124] V. Lucarini and M. D. Chekroun, Supplementary data, Detecting and attributing change in climate and complex systems: foundations, Green's functions, and nonlinear fingerprints, 10.6084/m9.figshare.27375723 (2024).

End Matter

Appendix A: Green's function and response—Near equilibrium, the standard form of the fluctuation-dissipation theorem [56,63,99] allows for predicting forced fluctuations in terms of readily accessible and intuitive correlations of observables in the unperturbed system. This form of the FDT has been applied to predict climate models' response to changes in the solar irradiance [100] and greenhouse gas concentration [101,102] and to study the impact of localized heating anomalies [103]. Yet, this approach can lead to potentially large errors [80]. One can compute the climate response via LRT, bypassing the FDT, for models of extremely diverse complexity [73–75,104–106]. We frame climate change by expanding the statistical state ρ_ϵ^t and solving the time-dependent FPE associated with the SDE (3) as

$$\rho_\epsilon^t(\boldsymbol{x}) = \rho_0(\boldsymbol{x}) + \sum_{p=1}^{M} \epsilon_p \rho_p^t(\boldsymbol{x}) + h.o.t., \qquad \text{(A1)}$$

where $\rho_0(\boldsymbol{x})$ is the reference unperturbed climate described by the stationary probability density associated with Eq. (2). Such an expansion is the starting point of virtually any linear response formula; see Refs. [67,68] for a discussion on the radius of convergence. The expected value of Ψ at time t is given by

$$\langle\Psi\rangle_{\rho_\epsilon^t} = \int \mathrm{d}\boldsymbol{x}\rho_\epsilon^t(\boldsymbol{x})\Psi(\boldsymbol{x}) \approx \int \mathrm{d}\boldsymbol{x}\rho_0(\boldsymbol{x})\Psi(\boldsymbol{x}) + \sum_{p=1}^{M}\epsilon_p \int \mathrm{d}\boldsymbol{x}\rho_p^t(\boldsymbol{x})\Psi(\boldsymbol{x}). \quad \text{(A2)}$$

Under natural assumptions valid for a wide class of stochastic systems [36], the "linear response" approximates the ensemble anomaly, i.e., $\langle\Psi\rangle_{\rho_\epsilon^t} - \langle\Psi\rangle_0 = \delta[\Psi](t)$ by $\delta^{(1)}[\Psi](t)$ given by Eq. (4). Green's functions $\mathcal{G}_\Psi^p$ are [59]

$$\mathcal{G}_\Psi^p(t) = \Theta(t)\int \mathrm{d}\boldsymbol{x}\rho_0(\boldsymbol{x})\left(e^{t\mathcal{K}_\Sigma}\Psi(\boldsymbol{x})[\mathcal{L}_p\log(\rho_0)](\boldsymbol{x})\right), \quad \text{(A3)}$$

where $\Theta(t)$ is the Heaviside distribution ensuring causality [35,74,107]. The operators $\mathcal{K}_\Sigma$ and $\mathcal{L}_p$ are [59]

$$\mathcal{K}_\Sigma\Psi = \mathbf{F}\cdot\nabla\Psi + \frac{1}{2}\Sigma\Sigma^T : \nabla^2\Psi, \qquad \mathcal{L}_p\rho = -\nabla\cdot(\mathbf{G}_p\rho), \qquad p = 1,\ldots,M. \quad \text{(A4)}$$

In Eq. (A4), $\mathcal{K}_\Sigma$ denotes the Kolmogorov operator associated with Eq. (2), ":" denotes the Hadamard product, and $e^{t\mathcal{K}_\Sigma}$ in Eq. (A3) is the corresponding Markov semigroup [83]. Instead, $\mathcal{L}_p$ indicates the correction to the Kolmogorov operator due to the perturbation included in Eq. (3). These formulas provide a general version of the FDT, as Green's functions are lagged correlations of the observables $\Phi = \mathcal{L}_p\log(\rho_0)$ and Ψ.

As made very clear in Ruelle's derivation of response formulas for nonequilibrium systems [34], Eq. (4) combined with the causality of $\mathcal{G}_\Psi^p$ implies that the effect of the perturbation comes strictly after the forcing is applied, thus defining a time-ordered causal link and flow of information. This is excellently aligned with the concept of interventions and counterfactual reality needed to define causality in Pearl's sense [31].

Appendix B: The Ghil-Sellers model—The G-S energy balance model [42] captures the essence of Earth's system radiative budget. It accounts for the absorption and reflection of solar radiation, emission of infrared radiation back to space, and large-scale heat transport from tropical regions toward the poles [1]. The G-S model has been instrumental for understanding why Earth's climate is metastable [13,108] and describes the evolution of the zonally averaged surface temperature $T_S(x,t)$, where $x = 2\phi/\pi$—the normalized latitude ϕ—lies in $[-1,1]$, and t is time,

$$c(x)\partial_t T_S = \left(\frac{2}{\pi}\right)^2\frac{1}{\cos\left(\frac{\pi x}{2}\right)}\partial_x\left(\cos\left(\frac{\pi x}{2}\right)\kappa(x,T_S)\partial_x T_S\right) + \mu Q(x)(1-\alpha(x,T_S)) - \sigma T_S^4(1-m\tanh(c_3T_S^6)), \quad \text{(B1)}$$

with von Neumann boundary conditions $\partial_x T_S(-1,t) = \partial_x T_S(1,t) = 0$. Here $c(x)$ is the effective heat capacity of the atmosphere, land, and ocean per unit surface area at x. The first term on the right-hand side (rhs) represents the meridional heat transport as a diffusive law, where $\kappa(x,T_S)$ incorporates the effects of sensible and latent heat transport. The solar forcing (second term on the rhs) is modulated by the solar constant μ, the irradiance Q, and the albedo α. The long wave emission (third term on the rhs) follows the Stefan-Boltzmann law, modified by the greenhouse effect, whose intensity is modulated by the constant m (Earth's grayness). Further details on the G-S model are reported in [108], which is our reference for all constants and tabulated functions.

Numerics We implement the G-S model by discretizing the latitude in $d = 37$ increments of 5° and by using a time step of one day. We discretize the spatial derivative operators via standard centered differences. To mimic the effects of the unresolved degrees of freedom and account for natural variability, we add a white noise forcing to Eq. (B1), with spatial correlation matrix given by $\eta_0\mathbf{I}_d$ with $\eta_0 \geq 0$. We use the Euler-Maruyama scheme for time integration. We choose as reference state the warm climate established with the present-day solar constant $\mu = 1$ and $\eta_0 = 0$ [see Fig. 1(a) in [108]]. Our ensemble runs are performed using $\eta_0 = 0.2$ (reference, strong noise case) and $\eta_0 = 0.1$ (weak noise case).

Climate change experiments We simultaneously perform (a) an increase in m, which mimics the increase in $[CO_2]$, and (b) a reduction of the incoming radiation in the region $[25\,°\mathrm{N}, 45\,°\mathrm{N}]$, mimicking the effect of aerosol injection in the atmosphere in the low-to-mid latitudes of the Northern Hemisphere. The forcing (a) is realized by increasing linearly over 100 years m in Eq. (B1) from $m = m_0 = 0.5$ to $m = m_0 + \delta m$, where $\delta m = 0.01$, and keeping the reached value of m constant afterward. This protocol corresponds roughly to a classical IPCC $[CO_2]$ stabilization scenario. The forcing (b) is realized by multiplying in the region $[25\,°\mathrm{N}, 45\,°\mathrm{N}]\mu$ times a factor ν that decreases linearly in 50 years from $\nu_0 = 1$ to $\nu = 1 - \delta\nu$, $\delta\nu = 0.012$, and then letting the perturbation fade away with an exponential law with decay time of 20 years. This corresponds to a typical IPCC technology transition scenario, where aerosol emissions first peak and then decay [15].

Appendix C: The climate model PLASIM-LSG—PLASIM is an open-source low-resolution climate model with a graphical user interface and comprehensive diagnostics suites [43,109]. PLASIM has been widely used to simulate past climate conditions [44,110], test specific atmospheric processes [111], look into the climate via nonequilibrium thermodynamics [45,112], and study circulation regimes of interest for exoplanetary research [113,114]. PLASIM has also been used as a test bed to investigate the climate with statistical mechanical tools. It has been shown that one can compute climate change projections via response theory [73,74], study tipping points through a functional analytical angle [115], investigate weather extremes using rare event algorithms [116,117], and analyze the climate metastability using field-theoretical approaches [118].

The atmospheric dynamics is modeled using the multilayer primitive equations, while moisture is included by transport of water vapor. We use a T21 spectral resolution, corresponding to ≈5.6° resolution in latitude and longitude), and ten vertical levels. The land surface scheme uses five diffusive layers for the temperature and a bucket model for the soil hydrology. Here, following [46], we consider as fully coupled ocean component the large-scale geostrophic (LSG) circulation model, which is based on the primitive equations and on the thermodynamics of saline water. The version used here has an effective ≈3.5° resolution in latitude and longitude and 22 vertical layers. The model is supplemented by well-tested parametrizations for unresolved processes, undergoes periodic forcing due to the orbital annual cycle, and has homogeneous atmospheric $[CO_2]$, which can be changed following a chosen temporal protocol, as done here.

Climate change experiments We consider a 1% increase of $[CO_2]$ starting from a reference value of 360 ppm, with the system at steady state, up to doubling (which occurs after ≈70 yr). $[CO_2]$ is then kept constant at 720 ppm until the system reaches a newly established steady state. The initial conditions for the $N = 40$ climate change runs are chosen from a 4000-yr-long control run with reference $[CO_2]$, with the jth ensemble member being initialized on the first day of year $1 + 100(j - 1)$ of the control run. While $[CO_2]$ grows exponentially up to doubling, the actual radiative forcing acting on the system grows linearly with time, see Refs. [73,74].

Despite its relative simplicity, PLASIM has a fairly realistic response to CO_2 forcing, featuring a climate sensitivity [119] of ≈3 K and an amplified warming over land in the Northern Hemisphere and in the high-latitude belt (polar amplification). A strong reduction of the warming occurs in the North Atlantic, see Fig. 2(b). The so-called "warming hole" or "cold blob" [75,120,121] is due to the weakening of the Atlantic meridional overturning circulation (AMOC); see Fig. 3. The AMOC is responsible for the ocean meridional heat transport [122] and is a tipping element [8] that is approaching criticality [12,123].

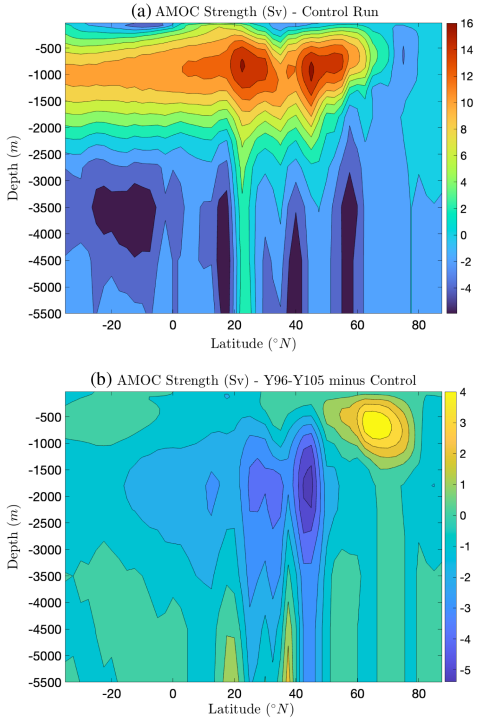

FIG. 3. PLASIM AMOC stream function showing northward transport at surface and return flow in depth (in $Sv = 10^6$ m^3 s^{-1}). (a) Climatology for reference $[CO_2]$. (b) Ensemble mean of the decadal anomaly 30 years after the end of the 1% yearly $[CO_2]$ increase ramp.

Appendix D: Computing Green's functions—Both for the G-S and the PLASIM-LSG models, we follow the strategy proposed in [73,74] for estimating Green's functions $\mathcal{G}^p_\Psi$. We first select M statistically independent initial conditions drawn from the steady-state unperturbed system. In each run, we choose as time modulation of pth forcing in Eq. (3), $g^p(t) = \Theta(t)$, which amounts to an instantaneous switch of this forcing only. By taking the average over the M runs, we estimate $\delta^{(1)}[\Psi](t)$, and [see Eq. (4)] we derive $\mathcal{G}^p_\Psi(t) \approx 1/\epsilon_p \mathrm{d}(\delta^{(1)}[\Psi](t))/\mathrm{d}t$.

For the G-S model, Green's functions describing the T_S response at the 37 latitudes due to changing m are computed by instantaneously increasing $m = m_0 \to m_0 + \delta m$. To obtain Green's functions for the second forcing, we instantaneously apply $\nu = \nu_0 = 1 \to 1 - \delta\nu$. For PLASIM-LSG, we compute Green's functions for the T_S anomalies at the 2048 grid points due to $[CO_2]$ forcing by instantaneously doubling $[CO_2]$ from 360 to 720 ppm; see also [73–75].