Internal vs Forced Variability Metrics for Geophysical Flows Using Information Theory

Aakash Sane[1,1], Baylor Fox-Kemper[1,1,1], David Ullman[2,2,2], and Aakash Sane[1]

[1]Brown University
[2]University of Rhode Island

December 1, 2022

Abstract

We demonstrate the use of information theory metrics, Shannon entropy and mutual information, for measuring internal and forced variability in ensemble atmosphere, ocean, or climate models. This metric delineates intrinsic and extrinsic variability reliably in a wider range of circumstances. Information entropy quantifies variability by the size of the visited probability distribution, as opposed to variance that measures only its second moment. Shannon entropy and mutual information manage correlated fields, apply to any data, and are insensitive to outliers and a change of units or scale. In the first part of this article, we use climate model ensembles to illustrate an example featuring a highly skewed probability distribution (Arctic sea surface temperature) to show that the new metric is robust even under sharp nonlinear behavior (freezing point). We apply these two metrics to quantify internal vs forced variability in (1) idealized Gaussian and uniformly distributed data, (2) an initial condition ensemble of a realistic coastal ocean model (OSOM), (3) the GFDL-ESM2M climate model large ensemble. Each case illustrates the advantages of information theory metrics over variance-based metrics. Our chosen metric can be applied to any ensemble of models where intrinsic and extrinsic factors compete to control variability and can be applied regardless of if the ensemble spread is Gaussian. In the second part of this article, mutual information and Shannon entropy are used to quantify the impact of different boundary forcing in a coastal ocean model. Information theory is useful as it enables ranking the potential impacts of improving boundary and forcing conditions across multiple predicted variables with different dimensions.

Internal vs Forced Variability Metrics for Geophysical Flows Using Information Theory

Aakash Sane, [a] Baylor Fox-Kemper,[b] David S. Ullman,[c]

[a] *School of Engineering, Brown University, Providence, Rhode Island, USA*

[b] *Department of Earth, Environment, and Planetary Sciences, Brown University, Providence, Rhode Island, USA*

[c] *Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island, USA*

Aakash Sane's current affiliation: Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey

Corresponding author: Aakash Sane, aakash.sane@princeton.edu

ABSTRACT: We demonstrate the use of information theory metrics, Shannon entropy and mutual information, for measuring internal and forced variability in ensemble atmosphere, ocean, or climate models. This metric delineates intrinsic and extrinsic variability reliably in a wider range of circumstances. Information entropy quantifies variability by the size of the visited probability distribution, as opposed to variance that measures only its second moment. Shannon entropy and mutual information manage correlated fields, apply to any data, and are insensitive to outliers and a change of units or scale. In the first part of this article, we use climate model ensembles to illustrate an example featuring a highly skewed probability distribution (Arctic sea surface temperature) to show that the new metric is robust even under sharp nonlinear behavior (freezing point). We apply these two metrics to quantify internal vs forced variability in (1) idealized Gaussian and uniformly distributed data, (2) an initial condition ensemble of a realistic coastal ocean model (OSOM), (3) the GFDL-ESM2M climate model large ensemble. Each case illustrates the advantages of information theory metrics over variance-based metrics. Our chosen metric can be applied to any ensemble of models where intrinsic and extrinsic factors compete to control variability and can be applied regardless of if the ensemble spread is Gaussian. In the second part of this article, mutual information and Shannon entropy are used to quantify the impact of different boundary forcing in a coastal ocean model. Information theory is useful as it enables ranking the potential impacts of improving boundary and forcing conditions across multiple predicted variables with different dimensions.

Plain Language Summary

It is important in climate and environmental modeling to distinguish variability that is caused by external forces versus variability that arises from within the system being modeled itself. In this paper, we study an ensemble of coastal ocean models, that are forced with tides, winds, and offshore and atmospheric conditions and an ensemble of climate model simulations that are forced by greenhouse gases and solar warming. Here, we propose to use information theory–a way to count the number of physical states visited by a system under study–to quantify the amount of variability in these models that results from the external forcing versus from the internal forcing. In this way, we can prioritize improvements or inclusion of the different forcings based on how large the model response to them is.

1. Introduction

In an ocean or climate model, it is pertinent to understand the cause of variability as it leads to implications for predictability, prioritization of data collections for assimilation, and provides an understanding of the dynamics at play in different regions. In a coastal model, variability can arise from extrinsic factors such as wind forcing, solar and thermal forcing, tides, rivers, evaporation and precipitation, or it can be due to internal chaos inherent to the governing fluid equations (Sane et al. 2021). In a climate model, modes of variability such as El Niño, the North Atlantic Oscillation, or the Southern Annular Mode, can conceal or delay the emergence of attributable anthropogenic climate change signals (Milinski et al. 2019). In high-resolution ocean models, internal chaos or intrinsic variability can also be due to eddies (Leroux et al. 2018; Llovel et al. 2018). Accurately quantifying the relative contribution of external and internal factors can help in elucidating the causes responsible for the observed variability in models, help to identify key observable metrics, and help quantify concepts such as the time of emergence of climate signals (Hawkins and Sutton 2012).

Numerous methods exist in the literature to quantify intrinsic and extrinsic variability using models or observations (e.g., Frankcombe et al. 2015; Schurer et al. 2013; Liang et al. 2020). Model ensembles–i.e., a set of simulations sharing the same forcing–naturally vary because each ensemble member follows the same governing equation (with same external forcings) with identical or similar parameterizations but differ due to intrinsic chaos. Two types of model ensembles are common: initial condition ensembles (where the same model is used repeatedly with perturbed initial conditions and intrinsic variability occurs via chaos), and multi-model ensembles (where a variety of models differing in numerics and parameterizations are used to simulate change under the same forcing–in this case "intrinsic" variability also includes aspects of model formulations). Most of the discussion here will focus on initial condition ensembles, but the metrics proposed can be adapted to both cases.

To help visualize variability, a generic output from an ocean or atmospheric model is shown in Figure 1. Each color represents a different ensemble member and the black solid line is the mean of those members. The black solid line is the signal mostly due to extrinsic factors (aside from finite ensemble size limits) and the model spread (schematized by the double-headed magenta arrow in Figure 1) can be considered due to intrinsic variability or internal chaos.

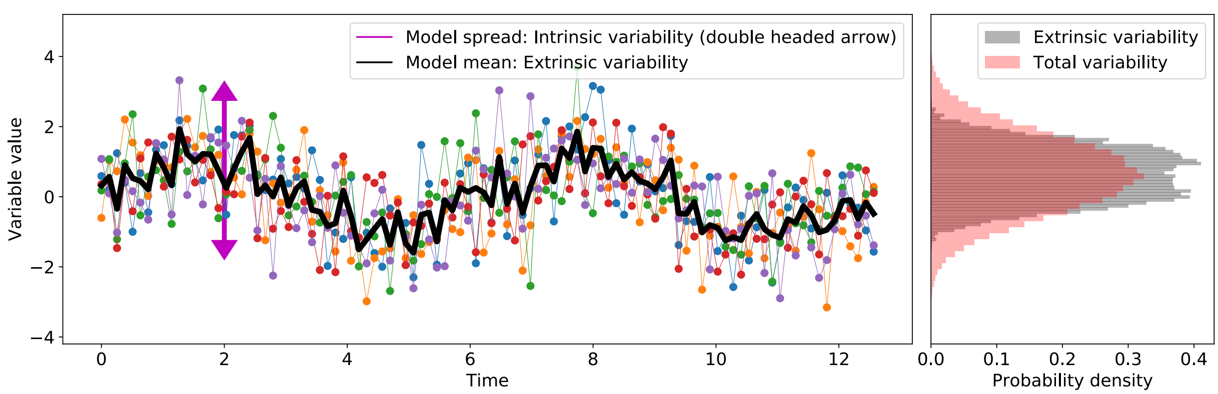

FIG. 1. A sketch of a typical ocean or climate model output for an arbitrary variable. Each ensemble is shown in different color and the mean of the ensemble is shown as black line. The ensemble mean can be considered to be the trend set by external forcings. The model spread shown by double headed magenta arrow indicates the model chaos.

One method of quantifying intrinsic and extrinsic variability is to look at variances (second central statistical moment) of model spread and model mean (Leroux et al. 2018; Llovel et al. 2018; Waldman et al. 2018; Yettella et al. 2018). Variance is sufficient to constrain all metrics of variability about the mean when distributions are Gaussian and uncorrelated, but a single statistical moment usually measures only part of a more complex variability. Many climatological variables show non-Gaussian distributions (e.g., Franzke et al. 2020). In fact, generalized variance might be misleading (e.g., Kowal 1971). Quantification of variability should be robust to or have known dependence on changes in the units of the quantity or the scale (e.g., changing temperature from Celsius to Fahrenheit or Kelvin). Comparative metrics, such as intrinsic vs. extrinsic variability should not depend on these arbitrary choices of units at all.

Variability, in essence, is a function of the number of occurrences or frequency of occurrence (or probability p_i as a fraction over all visited system states) after appropriately binning the data (and thereby making the estimated and visited number of states finite rather than continuous). Information entropy metrics measure variability by taking into account the probability distribution of the binned data, drawing on the statistical mechanics concept of entropy in quantifying the number of microstates that a variable can occupy. The fundamental measure in information theory is the Shannon (1948) or information entropy which characterizes the amount of variability in a variable (Carcassi et al. 2019). The mutual information, another metric introduced by Shannon (1948), measures how much information one variable contains about another variable.

Information theory is applied in signal processing, computer science, statistical mechanics, quantum mechanics, etc. It is used to quantify amount of information, disorder, freedom, or lack of freedom (Brissaud 2005). The application of these abstract notions to geophysical flows can have immense practical benefit when information entropy is interpreted as a measure of variability, as entropy does not rely on any particular parametric probability distribution. Metrics from information theory are not new to climate sciences. They have been introduced in predictability studies, evaluating the skill of statistical models, as well as uncertainty studies (e.g., Leung and North 1990; Schneider and Griffies 1999; Kleeman 2002; DelSole and Tippett 2007; Majda and Gershgorin 2010; Stevenson et al. 2013) and recently in studying variability (Gomez 2020) and coastal predictability (Sane et al. 2021).

In this article we bring well-established concepts of information theory to the particular application of measuring intrinsic and extrinsic variability for ensemble model runs within atmospheric and oceanographic modeling. Our metric uses Shannon entropy and mutual information. We indirectly employ conditional entropy, which depends on Shannon entropy and mutual information. To keep the metric intuitive, we have used Shannon entropy and mutual information and not cast it using conditional entropy.

There are two parts to this article. In Part 1, we apply our metric to three sets of data: 1. Idealized Gaussian and uniformly distributed arrays with specified correlation 2. Ensemble output of a regional coastal model (OSOM) (Sane et al. 2021) where most variables are non-Gaussian. Ensemble data for the duration of July-August of 2006 has been used. 3. The GFDL-ESM2M Large Ensemble (Rodgers et al. 2015; Deser et al. 2020), hereby referred to as GFDL-LE. This dataset contains historical and RCP 8.5 simulation data. All the monthly mean data from 1950 to 2100 have been used in the analysis.

In Part 2, we use OSOM to demonstrate the use of Shannon entropy and mutual information in evaluating the effects of altered boundary forcings. In coastal and estuarine systems, it is relevant to know which forcings are dominant which could potentially lead to prioritizing data collection to improve accuracy of the forcings. For example, is wind forcing dominant over river forcing, does using temporal averaged river runoff cause any appreciable changes in the estuarine circulation, or does change in the wind product alter circulation? These questions can be tackled by switching on and off or modifying each forcing and comparing the predicted variables using information theory.

Recent theoretical advances in understanding dynamical systems through the lens of information theory relate causality analysis and information transfer (e.g., Liang 2014). Although important, the transfer of such theoretical concepts into pragmatic research applications are few. Even basic concepts of information theory (Shannon entropy and mutual information) have been adopted in a limited capacity by the oceanic and atmospheric community to address problems arising in predictability and variability. We attempt to bridge the gap using approximate but practical framework which can be easily replicated and improved upon in the future, including causality analysis and the evolution of entropy within modeling systems like those studied here.

a. Information theory

We will introduce information theory concisely assuming the reader has no background knowledge–this section contains standard definitions. Consider a probability distribution p_i obtained after binning data into N bins. The user chooses the appropriate number of bins or bin widths for the range of data. Shannon (1948) identified the average information content in N possible outcomes, equally or not equally likely, as given by:

$$H = \sum_{i=1}^{N} p_i \log_2(1/p_i), \tag{1}$$

where H is the Shannon entropy with unit of bits when log is base 2 and p_i is the probability of the i^{th} outcome. The factor $\log_2(1/p_i)$ measures the information of the i^{th} outcome as proposed by Hartley (1928) and is also a measure of uncertainty (Cover 1999), as it measures the information gained by knowing that the i^{th} outcome has happened or equivalently that the variable falls in the i^{th} bin. The term information does not mean knowledge but it means the amount of uncertainty shown by a variable or the freedom that a variable has in visiting different combinations of the N bins. Shannon (1948) found Equation 1 to provide the average information (or uncertainty) for all events in a record. For the entire set of elements, a highly probable event has less uncertainty associated with it and low probability event has high uncertainty associated with it. The prefactor p_i is thus used to weight the information over all possibilities. One way to interpret the need for the prefactor p_i is that in repeated experiments the events with higher probability will occur more often, hence they should contribute more to a quantification of variability than infrequent events.

Stone (2015) gives an intuitive way of understanding Shannon entropy using a binary tree. A binary tree is a tree chart which starts with one node and splits to two nodes at each node. At each node you can take a left or right turn to proceed and if there are say 3 levels in the tree, then 8 (i.e. 2^3) outcomes or possible destinations exist. If a binary tree has N equally probable outcomes then the set of instructions required to reach the correct destination is given by $h = (N)(1/N)\log_2(N) = \log_2(N)$. The *uncertainty* about reaching the correct destination will be removed by providing $\log_2(N)$ *bits* of information. In other words, if entropy is h then 2^h states are possible.

A second metric from Shannon (1948) which is also extensively used is now known as *mutual information*. The mutual information between two signals x and y denoted by $I(X;Y)$ is (Cover 1999)

$$I = \sum_{j=1}^{N}\sum_{i=1}^{N} p_{ij}\log_2\left(\frac{p_{ij}}{p_i p_j}\right), \tag{2}$$

where p_{ij} is joint probability of i^{th} outcome of x and j^{th} outcome of y. The marginal probability of i^{th} and j^{th} outcomes of x and y respectively are p_i and p_j. The addend within the summations can be expanded to $p_{ij}\left(\log_2(p_{ij}) - \log_2(p_i) - \log_2(p_j)\right)$. I can be interpreted as the extra information in entropy of marginal distributions of x and y over the joint distribution. Mutual information is symmetric between x and y and is the measure of how much information they share. For example, if the distributions are statistically independent, then $p_{ij} = p_i p_j$ and thus $I = 0$. If the two records x and y are identical, then $p_{ij} = p_i = p_j$ and $I = H$. I is the average reduction in uncertainty in x from knowing y or vice versa and denotes how much information is transmitted between the two variables.

In the context of ocean modeling (or in general climate modeling) entropy is used to measure variability in a model output or available data. This is in tandem with interpretation of Shannon entropy in physical sciences as given in Carcassi et al. (2019). When calculating the Shannon entropy we are concerned about the possible states (e.g. the various bins in a histogram) the variable can (and does) go into while the variable value and its dimensions are of lesser importance. Entropy metrics measure variability in *bits* (when logarithm is of base 2) and hence changing the scale, e.g. switching from Celsius to Fahrenheit for temperature, does not change the value of variability (under equivalent binning). Mutual information and entropy are both dimensionally agnostic. They are also not sensitive to outliers (due to the weighting prefactor) and can capture nonlinear interactions (Watanabe 1960; Correa and Lindstrom 2013) and discontinuous distributions including intermittently visited states. We will present the effect of correlation and outliers by examples of idealized random vectors.

The following methods and results sections are divided into the two parts of the overall paper objectives. Parts A of both sections relate to evaluating intrinsic and extrinsic variability in ensemble models. Parts B describe the usage of Shannon entropy and mutual information on coastal regional modeling data to understand and compare effects of using different boundary conditions.

2. Methods

a. Part A: Intrinsic and Extrinsic variability for ensemble data

We perform analysis on each grid point at the ocean surface or ocean bottom. Let a variable from the ensemble be given by $f(n,t,x,y)$ where f is the variable, n denotes the index of the ensemble member and goes from 1 to N, t is the time index and goes from t_1 to t_M, x, y represents the spatial grid point at the surface or bottom. At a particular grid point $f(n,t,x,y)$ is $f(n,t)$. The total number of ensemble members is N and each member has M time steps. To get the signal due to extrinsic forcings, the "differencing" approach (Frankcombe et al. 2015) has been followed to estimate forced response. This approach involves averaging over the ensemble members to derive the *ensemble mean*. The ensemble mean is given by:

$$g(t) = \frac{1}{N}\sum_{n=1}^{n=N} f(n,t) \tag{3}$$

$g(t)$ is a single time varying signal for each grid point obtained by averaging across the ensemble members. There are potential problems with assuming the ensemble mean represents extrinsic variability only, such as if models are differently sensitive to the forcing signal based on the model's equilibrium sensitivity as elaborated in Frankcombe et al. (2015). For a first order approximation, we will assume the ensemble mean is the best estimate of the forced response. Once $g(t)$ is obtained, the intrinsic variability can be estimated by subtracting the ensemble mean $g(t)$ from each ensemble member. Ensemble signal, forced response and intrinsic variability are then related by:

$$f(n,t) = g(t) + \eta(n,t), \tag{4}$$

where $\eta(n,t)$ is the intrinsic variability or noise which differs from ensemble member to ensemble member. Note that the decomposition above takes place at each grid point. In Figure 1 a, $f(n,t)$ are shown by multi-colored ensemble members. $g(t)$ is shown by thick black line. As seen Figure 1 b, $g(t)$ has a probability distribution shown in gray color and subsequently has first, second, and possibly important higher statistical moments. The gray colored density histogram shows variability due to extrinsic factors and the pink colored density histogram shows total variability given by extrinsic and intrinsic factors.

1) Detrending and evaluating entropies

Analysis has been done with and without detrending the data to understand its impact. For detrending, a quadratic fit using least squares was found for the ensemble mean at each grid point and subtracted from all ensemble members and ensemble mean at the same grid point to get detrended data (e.g. Frankcombe et al. 2015). Detrending will remove some non-stationarity from the data but will also remove some part of the extrinsic variability. By this method, our aim is not to determine the forced response but to estimate the degree of *variability* contributed by the forced response (extrinsic response) and intrinsic variability originating from intrinsic chaos. The ensemble mean $g(t)$ was found at each grid point after detrending. For the non-detrended case, the raw ensemble simulation data has been used to evaluate $g(t)$ and $\eta(n,t)$.

Usually we are limited in the number of ensemble members due to computational costs so we perform a *jugaad* in order to use *all* the ensemble members at once to evaluate information entropies. All the ensemble members given by $f(n,t)$ are rearranged into a single row vector f as:

$$f = [f(1,t_1), f(1,t_2), ...f(1,t_M), f(2,t_1), f(2,t_2), ...f(N-1,t_M), f(N,t_1), ...f(N,t_M)], \quad (5)$$

and g is row vector obtained by arranging N copies of $g(t)$ in the following fashion:

$$g = [\underbrace{g(t_1), g(t_2), ...g(t_M)}_{1}, \underbrace{g(t_1), g(t_2), ...g(t_M)}_{2}, ... \underbrace{g(t_1), g(t_2), ...g(t_M)}_{N}] \quad (6)$$

This enables wide sampling and obtains an accurate probability distribution for f (assuming approximate stationarity, or enforcing stationarity by detrending), and enables g to be of the same size as f and having the same probability distribution as that of $g(t)$. The information statistics we get at each grid point are time invariant since the complete time series is considered. It is the user's choice to choose either the complete time series or a section of it for analysis. We have chosen the whole time series, as this is a sufficient demonstration of the value of information theory metrics. A time-evolving analysis raises additional issues about causality and shifting probabilities distributions of climate states that are not the focus here (Liang 2013; DelSole and Tippett 2018). By using the whole time-series, we are treating all variability as drawn from the same distribution, and seek only to associate internal (associated with each ensemble member) and external (associated with the ensemble mean) sources of variability following Leroux et al. (2018). The time-series f and g are both expressed as row vectors of the same size, $N \times M$. This step is crucial as vectors having same number of elements are necessary to evaluate joint probability distribution. This enables us to calculate mutual information between f and g.

Calculating the Shannon entropy of f and mutual information between f and g is not a trivial task. In fact optimal binning for precise measurement of information entropies is a research topic in itself. Multiple techniques exist such as equidistant partitioning, equi-probable partitioning, k nearest neighbor, usage of B-spline curves for binning to name a few (e.g. see Hacine-Gharbi et al. 2012; Kowalski et al. 2012; Knuth 2019). For a comprehensive review of the methods for estimating probability distribution see Papana and Kugiumtzis (2008). We have used equidistant partitioning throughout this article. For the case of GFDL-LE data, there were 1812 time steps available as monthly averages ranging from the year 1950 to 2100. As per Rice's rule, 25 bins are needed for the GFDL-LE data. The bin width, δw, was calculated by dividing the range of data (maximum minus the minimum value) at the grid point with the least spread. The same bin width was used for all the grid points for Shannon entropy and mutual information. Equal bin width was used for the two variables in the joint probability and marginal probability calculation for mutual information. Maintaining the same bin width and range for all the grid points is crucial because information entropy strongly depends on the precision with which data is binned.

2) Proposed metric

Using f and g, we propose the following metric γ, which has the same intent as metrics in (Leroux et al. 2018) to quantify the fraction of variability that is intrinsic, i.e., the typical amount that is unique to an ensemble member or statistical instance, but unlike (Leroux et al. 2018) this metric is built from standard information theory quantities:

$$\gamma = 1 - \frac{I(f;g)}{H(f)}. \tag{7}$$

$H(f)$ is the Shannon entropy of f, and $I(f;g)$ is mutual information between f and g. $I(f;g)$ calculates the contribution of extrinsic signal g to the whole ensemble. $H(f)$ is the total variability in the ensemble output which is the result of extrinsic and intrinsic factors. The metric γ gives the *ratio of intrinsic variability to total variability*.

$H(f)$ and $I(f;g)$ are related through conditional entropy by $H(f) = I(f;g) + H(f|g)$ (Cover 1999). $H(f|g)$ is the conditional entropy[1], i.e., average uncertainty about the value of f after g is known. It is the uncertainty in f that is not attributed to g but is attributed to noise η. Hence $H(f) - I(f;g)$ estimates variability due to intrinsic chaos, and γ gives the fraction of the variability due to intrinsic chaos.

Returning to the binary tree analogy, $I(f;g)$ would be the set of instructions sent by a source to reach one among $2^{H(f)}$ possible destinations in the presence of noise having $H(f|g)$ entropy. To

[1]Conditional entropy $H(X|Y)$ is defined by $H(X|Y) = \sum p(x|y) \log_2 p(x|y)$ (Cover 1999). It is not necessary to calculate conditional entropy to arrive at γ, but understanding is aided by the expected relation between entropy and mutual information.

capture the entropy in the noisy binary tree, to each of the $2^{I(f;g)}$ micro state possibilities noise $\left(2^{H(f|g)}\right)$ gets multiplied and the relation becomes $2^{H(f)} = 2^{I(f;g)}2^{H(f|g)}$.

$I(f;g)$ takes into account any correlation or information shared between f and g. This is vital because even though the model spread η is being treated similarly to noise added to the mean signal, it might be that model spread depends on the mean signal. A simple example is if the model spread is relative (e.g., 10% of the mean signal), rather than absolute (e.g., 2 units), then there is information about the model spread contained in the ensemble mean signal. This situation is sometimes called multiplicative noise in contrast to additive noise. The nonlinear and chaotic nature of fluid mechanics often leads the mean flow to amplify the chaotic signal (e.g., eddies) and thereby result in altered variability statistics. When $f \rightarrow g$, then $I(f;g) \rightarrow H(f) = H(g)$ from (2). This makes $\gamma = 0$ when there is no intrinsic variability or chaos. When intrinsic chaos fully dominates the ensemble output, i.e. f and g are fully decorrelated, then $I(f;g) = 0$ yielding $\gamma = 1$. We see that γ satisfies the extremes of zero noise as well as total chaos.

Another analogue for a climate system component is a noisy communication channel as given in Leung and North (1990), where the governing equations of ocean (atmosphere) modeling are taken to communicate from forcing to response. The extrinsic forcings are inputs to the channel, the intrinsic chaos is the noise created because of channel's inherent mechanisms while the outputs are the ensemble members. A noiseless channel will give γ as zero and completely noisy channel where output is independent of input will yield γ as 1.

A seemingly enticing and simpler alternative is $\gamma = 1 - \frac{H(g)}{H(f)}$, i.e. just the difference between ensemble entropy and mean entropy as a ratio with the ensemble entropy. However, this formulation is incorrect because $H(g)$ does not quantify the contribution of extrinsic factors to the variability in the ensemble, it only quantifies the variability of the mean. Relatedly, $H(f) - H(g)$ does not correctly manage mutual information between the ensemble members and their mean in estimating the intrinsic variability.

Recently, another alternative was proposed by Gomez (2020): using Shannon entropy directly as a measure of intrinsic variability. They propose using Shannon entropy of model spread $\eta(n,t)$ at each time step normalized by the logarithm of the number of bins utilized. Their metric has a lower limit of 0 and an upper limit of 1, where 0 denotes zero noise and hence zero intrinsic variability and 1 denotes complete intrinsic variability. Again, this metric is similar to γ in building upon information theory, but γ takes into account the variability of the ensemble mean, correlations between the ensemble mean and the intrinsic variability, and it is time invariant. A time dependent version of γ can be made using running time windows instead of the whole time series, but care in quantifying or controlling for lack of stationarity is needed in this interpretation (DelSole and Tippett 2018). The Gomez (2020) metric uses the spread of the ensemble members similar to measuring Shannon entropy whereas γ utilizes, in an abstract sense, the set of instructions required to choose a destination for the particular variable among the possible model states.

3) Variance based metric

A variance based metric as given in (Leroux et al. 2018) has been utilized to compare to our information based metric. The variance based metric measures intrinsic and extrinsic variability using the second moment, variance. It involves calculation of the following terms σ_g and σ_η given by:

$$\sigma_g^2 = \frac{1}{M}\sum_{t=1}^{t=M}\left(g(t) - \overline{g(t)}\right)^2, \tag{8}$$

$$\sigma_\eta^2(t) = \frac{1}{N}\sum_{n=1}^{N}\eta(n,t)^2, \tag{9}$$

where the overbar denotes temporal averaging. The total variability has been estimated as $\left(\sigma_g^2 + \overline{\sigma_\eta^2(t)}\right)^{1/2}$. The forced variability σ_g is equivalent to $I(f;g)$, and total variability $\left(\sigma_g^2 + \overline{\sigma_\eta^2(t)}\right)^{1/2}$ is equivalent to $H(f)$. Hence, γ is compared with γ_{std} given by

$$\gamma_{std} = \frac{\left(\overline{\sigma_\eta^2(t)}\right)^{1/2}}{\left(\sigma_g^2 + \overline{\sigma_\eta^2(t)}\right)^{1/2}} \tag{10}$$

b. Part B

1) Impact of changes in boundary forcings in coastal models

Here instead of using the new metric γ, we use its components: Shannon entropy and mutual information individually to compare variability between different simulations. Quantifying differences because of modifications in the extrinsic forcings may be required for coastal applications

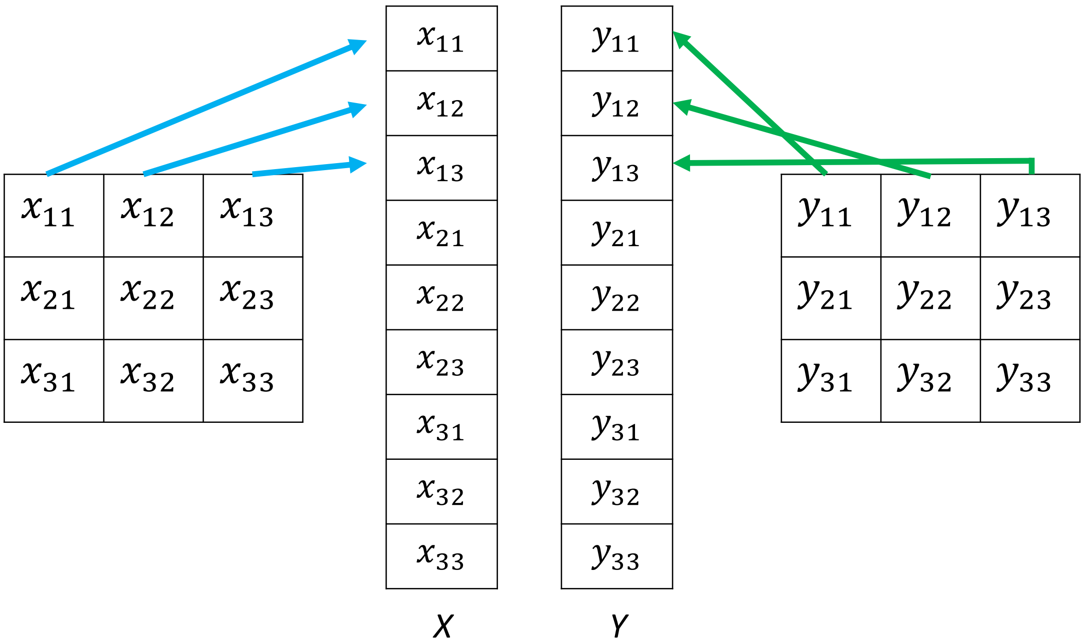

FIG. 2. Flattening process for comparing two dimensional fields using Shannon entropy and mutual information. As the flattened arrays $x_1, x_2, \ldots$ and $y_1, y_2 \ldots$ might not have linear dependence on each other, using linear dependence measure such as Pearson correlation will yield incorrect results. Mutual information measures nonlinear correlations and hence captures all linear and non-linear dependence.

where systems vary predominantly due to external forcings. For these forcing significance experiments, OSOM was run after modifying the external forcings (Table 1). OSOM is forced by tides, river runoff, atmospheric winds and air-sea fluxes, etc. (Full details of the model can be found in Sane et al. 2021). For this comparison, we quantify the effects of altering forcing on 4 modeled variables: sea surface temperature and salinity, and bottom temperature and salinity. Four altered forcing sets were utilized, beyond set (1) Full set of atmospheric forcings using the North American Mesoscale (NAM) analyses, a data-assimilating, high resolution (12 km) meteorological simulation (`https://www.ncei.noaa.gov/data/north-american-mesoscale-model/access/historical/analysis`) denoted as FF. FF stands for full forcing. (2) Full set of atmospheric forcings but using the Northeast Coastal Ocean Forecast System (NECOFS) winds (Beardsley and Chen 2014) instead of NAM, denoted as NECOFS. (3) River flows are replaced with their monthly-averaged flow, other forcing as in FF (4) River flows set to zero, other forcing as in FF. (5) Wind forcing set to zero, other forcing as in FF. These forcings have been tabulated in Table 1. The aim is to quantify the effect on total variability by removing or altering one of many processes which might contribute.

Forcing Set	Wind forcing	River forcing
FF	NAM	As Observed
NECOFS	NECOFS	As Observed
MR	NAM	Time-averaged
ZR	NAM	Zero river input
ZW	Zero winds	As Observed

TABLE 1. Different types of forcing combinations employed to test their effect on variability. FF stands for full forcing: winds, tides, rivers, etc. For more details see Sane et al. (2021). MR: mean rives; ZR: zero rivers; ZW: zero wind.

To evaluate Shannon entropy, the spatial output at a particular instant of time was rearranged into a row vector by a process called 'flattening' as shown in Figure 2. Land mask points were removed. A variable x which is a two-dimensional variable was converted to one-dimension (flattened) by concatenation. Shannon entropy was found out for the flattened variable at each time step to obtain time varying entropy of the surface or bottom variable.

Mutual information was applied between the flattened row vectors. Our focus is towards a pragmatic approach on using information theory for simulation comparisons, as opposed to an equation for the evolution of Shannon entropy and mutual information with respect to time (see Liang and Kleeman 2005). Relative comparison between mutual information values is what we seek. For example, if mutual information of surface salinity between FF and MR is higher than between FF and ZR, this implies the penalty for using time-averaged river runoff is not as severe as using zero river runoff. Replacing FF with MR will give better results than ZR. Small errors in river runoff flow rates won't cause appreciable changes to surface salinity than using zero rivers.

3. Results

a. Part A

1) Idealized Gaussian Arrays

We test our metric, γ, equation (7) on synthetic data consisting of idealized arrays of Gaussian data: $\mathcal{N}(0,1)$. For a normal Gaussian distribution Shannon entropy depends[2] only on the standard deviation σ i.e. $H = \log_2 \left(2\pi e \sigma^2\right)$. The variability in a Gaussian distribution can be increased or decreased by changing its standard deviation. Our goal is to compare γ and γ_{std}. We set out our numerical experiment as follows: we create 10 arrays, each having 10,000 elements drawn from a Gaussian distribution. Any two arrays from those 10 have a prescribed linear Pearson correlation coefficient from 0 to 1.

Thus, the 10 arrays covary linearly with a specified correlation coefficient. These 10 arrays represent ensemble members from climate simulations. The mean of 10 members gives us the synthetic forced variability signal as would be determined from the model output; averaging over the 10 ensemble members reduces the contribution from uncorrelated variability and reaffirms the covarying component into the forced variability. We apply γ and γ_{std} on this synthetic ensemble by varying the prescribed correlation coefficient from 0 to 1. Figure 3 shows that as expected both metrics increase as the correlation decreases, i.e., as internal variability dominates forced. Both metrics behave similarly when correlation decreases, i.e. noise increases but γ is more sensitive as correlation tends to 1. This distinction is due to the logarithmic nature of Shannon entropy for Gaussian distributions–in essence, information measured in bits is not proportional to distance measured between distributions in terms of summed variance–in the examples following the consequences of this distinction will become clearer. Critically both functions are monotonic with correlation, however so relative comparisons (more intrinsic fraction in this region vs. that region) are preserved.

A second related experiment was derived from the first is also shown in Figure 3: adding outliers outside of the Gaussian distribution. 50 out of 10000 elements of each individual member were artificially corrupted (values were set to a constant value of 5) to test the sensitivity of both the metrics. Figure 3 shows that γ is insensitive to outliers while γ_{std} is not. γ is not sensitive because

[2] $H = \log_2 2\pi e \sigma^2$ is the Shannon entropy of a Gaussian distribution when probability density is continuous with σ as standard deviation. The Shannon entropy of a discrete probability distribution differs, which is inconsequential here but the reader is encouraged to read Jaynes (1962). Consistently here discretely sampled and binned probability distributions are obtained directly from data without any further parameterization.

outliers occur less frequently and hence do not affect the probability distribution much, especially with the prefactor in (1) and (2). Hence information theory metrics are robust in comparison to using standard deviation (or variance). If the outliers (extreme events) occur at higher frequencies, information metrics will naturally start sensing them even if they are discontinuous from the typical conditions (e.g., multimodal distributions). The above process was repeated for 100 ensemble members each sampled from Gaussian distributions. Increasing the number of ensemble members does not change the result qualitatively for both the experiments. The results for 10 member Gaussian ensemble is shown in Figure 3 a and 100 member in Figure 3 b.

Additionally, a set of experiment was done by using uniformly distributed data $U(-1,1)$. The prescribed correlated vectors were created using the procedure outlined in Demirtas (2014). 10 and 100 ensemble members were created and γ and γ_{std} was found between the members and their mean. Results are shown in Figure 3 c, d respectively. The outlier had a value of 1.5. In all the cases, γ was less sensitive to outliers than γ_{std}.

2) Regional coastal model output

In this section we show the results of applying γ and γ_{std} on realistic simulation data from the Ocean State Ocean Model, hereafter OSOM (Sane et al. 2021). OSOM uses the Regional Ocean Modeling System (ROMS) (Shchepetkin and McWilliams 2005) to model Narragansett Bay and surrounding coastal oceanic regions and waterways. OSOM's primary purpose is for understanding and predictive modeling and forecasting of the estuarine state and climate of this Rhode Island body. Sane et al. (2021) gives more details about the model.

Using OSOM, an ensemble of simulations have been performed using perturbed initial (ocean) conditions under the same atmospheric and tidal forcing for the months July - August of 2006. This ensemble consists of 10 members. The data during the first predictability window (20 days) that is sensitive to initial conditions has been ignored and the remaining simulation has been used to look at variability within the "climate projection" of the model beyond when forecasts sensitive to initial conditions are possible (see the related application of information theory to assess predictability in Sane et al. 2021). We examine whether the modeled temperature and salinity at each grid point follow normal distributions by evaluating the skewness and kurtosis of the ensemble mean at each grid point. Figure 4 shows skewness and kurtosis for sea surface salinity and temperature as well

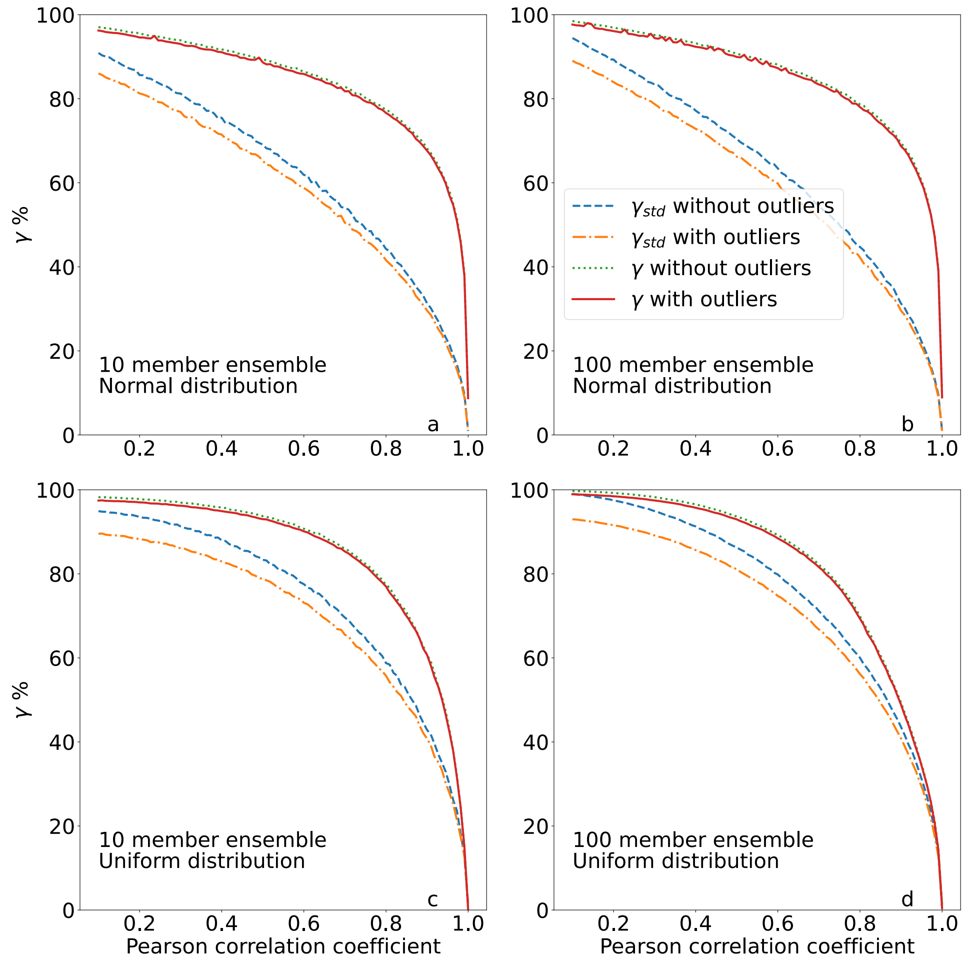

FIG. 3. Information theory metric of intrinsic vs. extrinsic variability γ as a function of correlation coefficient in idealized Gaussian correlated arrays (a and b) and idealized uniformly distributed arrays (c and d). The horizontal axis is the correlation coefficient between mean member and ensemble members. The vertical axis shows the information theory metric γ from (7) and the traditional metric γ_{std} from Equation (10). A second related experiment adding (50 out of 10,000) "corrupted" outliers to each individual member is also shown. The information theory metric γ does not change for these outliers which shows its robustness while γ_{std} is highly sensitive. Results are similar for Gaussian distribution members and uniformly distributed members. γ is more sensitive towards linear correlation of 1. This is due to the logarithmic nature of γ.

as bottom salinity and temperature for the Narragansett Bay region. The horizontal axis shows skewness and excess kurtosis, which are the third and fourth statistical moments respectively, normalized by powers of the standard deviation to dimensionless ratio and in the case of excess kurtosis a constant value of 3 is subtracted. For Gaussian distributions, skewness and excess kurtosis both should be close to zero. The vertical axis denotes the number of occurrences at a grid point. Observe that the majority of grid point values are away from zero. These variables are considerably non-Gaussian in OSOM. Thus, Equation (10) is at a disadvantage, because the prevalence of higher statistical moments implies that the variance does not contain a complete description of the variability. The information theory metric (7) is suitable for such data as it takes into account higher moments and does not rely on Gaussian distributions.

Figure 5 shows the ratio of intrinsic variability to total variability applied on every grid point for OSOM. γ is displayed on left whereas γ_{std} is shown on right for comparison. The features highlighted by both metrics are qualitatively different. The contribution of intrinsic chaos to total variability is more uniform using the γ metric than using γ_{std}. The intrinsic chaos displayed using γ_{std} might be misleading because the probability distributions are non-Gaussian. Furthermore, where the γ metric highlights internal variability tends to agree in similar dynamical locations–all river mouths show high surface salinity intrinsic variability. While surface temperature intrinsic variability is higher in more open regions of the Bay where eddies form intermittently due to varying topography. Also note that the ranges are quite different between γ and γ_{std}, but this is to be expected from the different rate of increase with correlation seen in Figure 3.

3) Community Earth System Model large ensemble

A complementary experiment was performed by using γ to evaluate internal vs. forced variability in the global climate simulation output for climate change scenario RCP8.5 using the (randomly selected among the models compared) GFDL-LE model. All the 40 members from the ensemble were utilized. Variability of sea surface temperature (Figures 6) as well as sea surface salinity (Figures 7) were estimated using both γ and γ_{std} (upper left and upper right). Similar results were obtained for the detrended data for temperature (Figures 8) and salinity (Figures 9) The skewness and excess kurtosis of the ensemble mean were also plotted to find the deviation of variables away from Gaussian distributions (lower). Regions shaded in purple have low values of excess kurtosis and skewness and might be considered Gaussian. The detrended data shows a higher percentage of intrinsic variability than non-detrended data which suggests that detrending

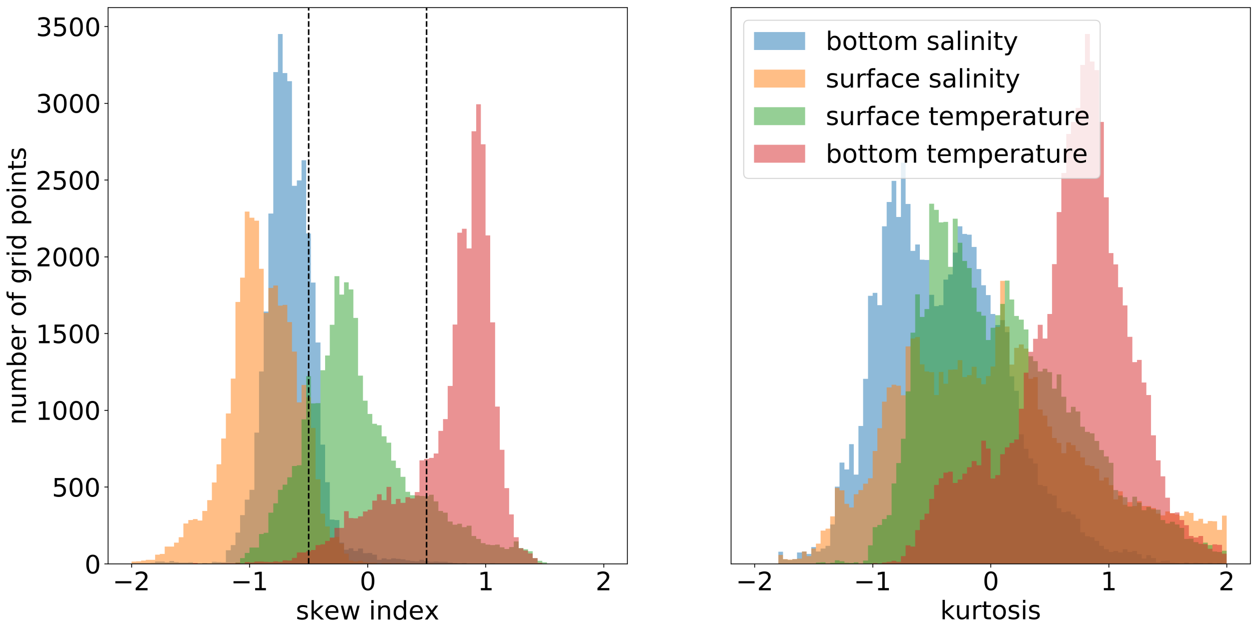

FIG. 4. Grid point wise kurtosis for OSOM output. Kurtosis is not closer to zero within (-0.5, 0.5) suggesting the data distribution is non Gaussian.

has removed some proportion of extrinsic variability, presumably the climate change signal present in these simulations.

Note in particular the Arctic sea surface temperatures, which have a highly skewed and excessive kurtosis distribution due to the freezing point of seawater. The standard metric (γ_{std}) deems this region to be among the most intrinsically variable in the world, while the information theory metric has it as a low intrinsic variability region. It is clear that a Gaussian metric should not be applied to this region due to the skewness and excess kurtosis, and in this case the inference is opposite using the two metrics. In the equatorial Pacific where Gaussian statistics are more reliable, the two metrics agree that internal variability is high.

A less drastic failure occurs from the modest excess kurtosis in extra-tropical temperatures and in a few isolated regions in surface salinity. These regions are also non-Gaussian, but also are not heavily skewed (i.e., they are more long-tailed and intermittent than Gaussian). These regions differ in relative estimation of intrinsic versus total variability. It is also the case that the γ metric is closer to one in most regions than γ_{std}, which is to be expected when the correlation coefficients are low from Figure 3.

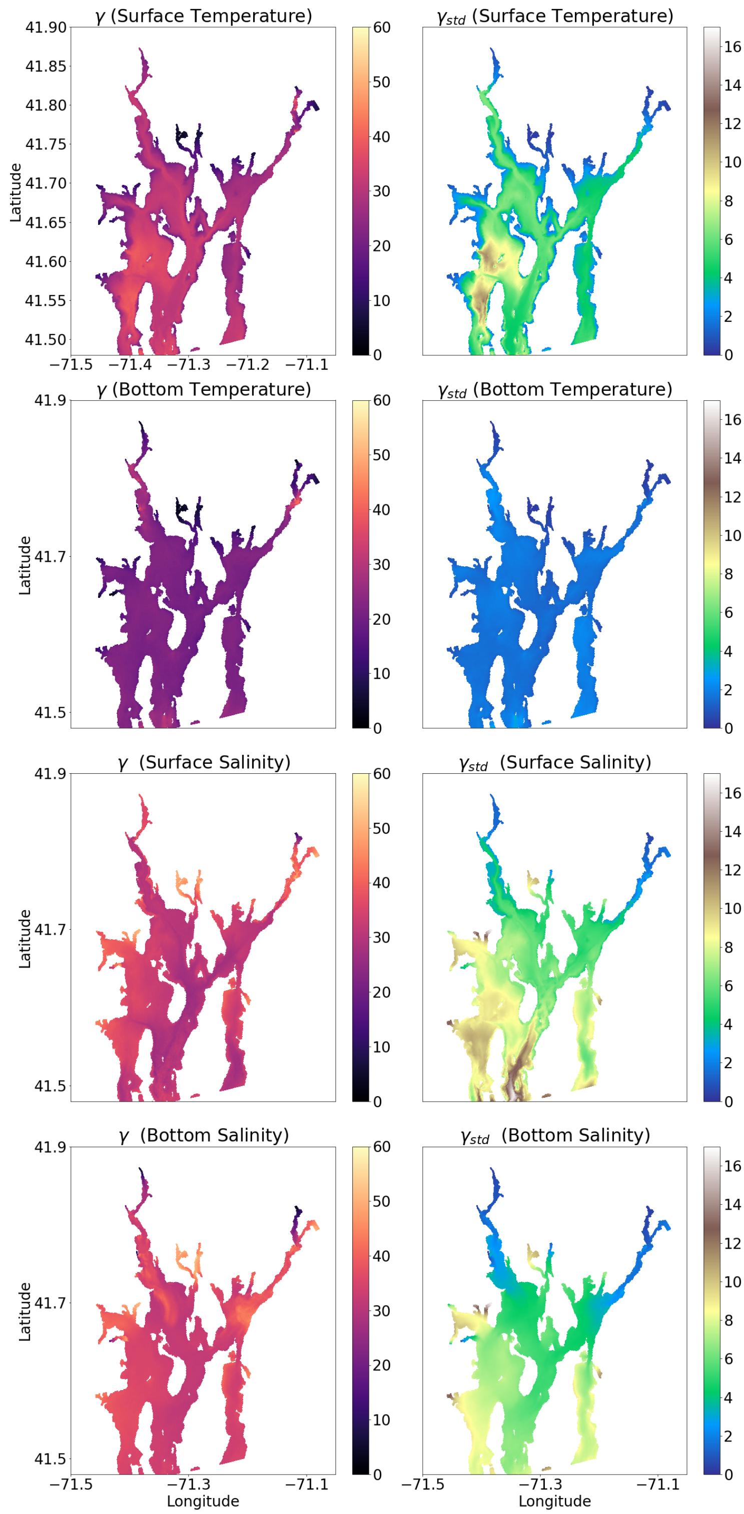

FIG. 5. Metrics γ vs γ_{std} for OSOM output. Both metrics show different contribution of intrinsic variability to total variability. γ is more uniform throughout the domain than γ_{std}. Colormaps for γ and γ_{std} are different to highlight the different ranges each of them have. γ_{std} for bottom temperature has maximum value of 5%, and pattern is almost uniform except at the river sources where values are on the lower side (less than 1%).

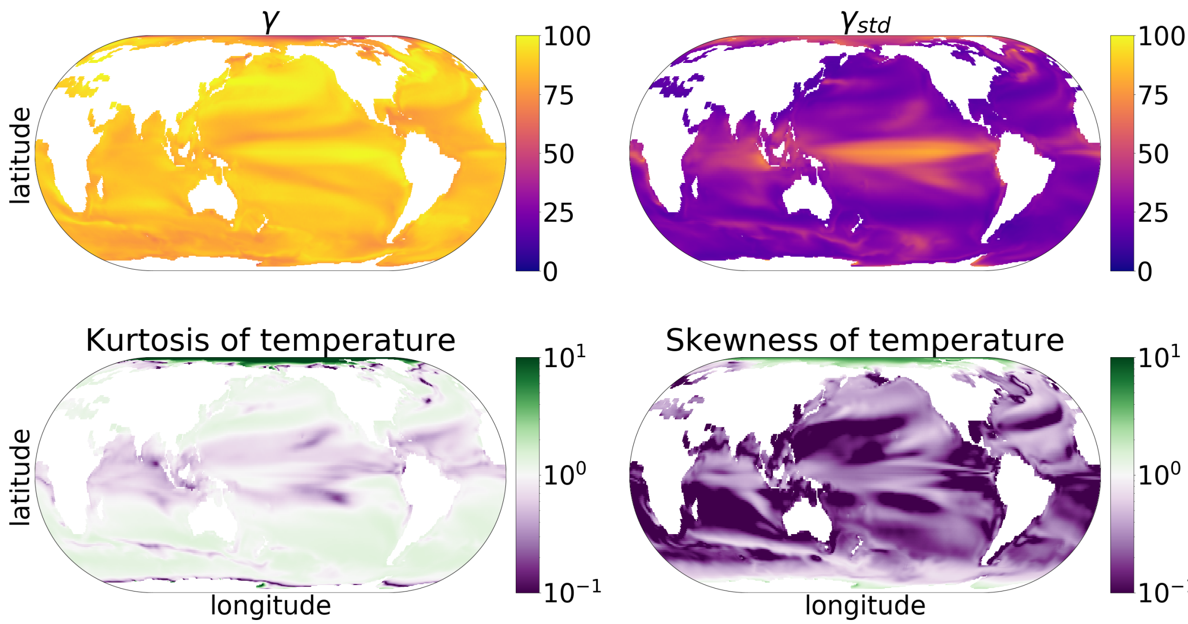

FIG. 6. Top: Intrinsic to total variability percentage for sea surface temperature. Bottom: Excess kurtosis and skewness of the ensemble mean of temperature at each grid point. Values closer to zero (within 0.5 of zero, purple shades) are considered approximately Gaussian. The deviation of ensemble mean away from non normality implies that the ensemble members are also non normal. The Arctic regions have the most skewness and excess kurtosis implying non-Gaussian distributions.

b. Part B

1) Impact due to changes in boundary conditions in coastal models:

We show results of the coastal model analysis under different forcing in Figures 10 and 11. Entropy has been plotted with respect to time to aid understanding. In Figure 10, Shannon entropy is plotted for spatial quantities. For example, for surface salinity, all the surface values have been considered to find Shannon entropy using the flattening approach. Figure 11 displays mutual information. It is user's choice to choose the type of domain, here we have chosen the same domain of OSOM as shown in Figure 5. If Shannon entropy is more or less equal for two forcings, it implies they similarly affect variability. Mutual information should be compared for two pairs of forcings. Greater mutual information implies the two pairs share more *bits* of information, suggesting one of the forcing in that pair can be replaced with the other without significantly affecting variability.

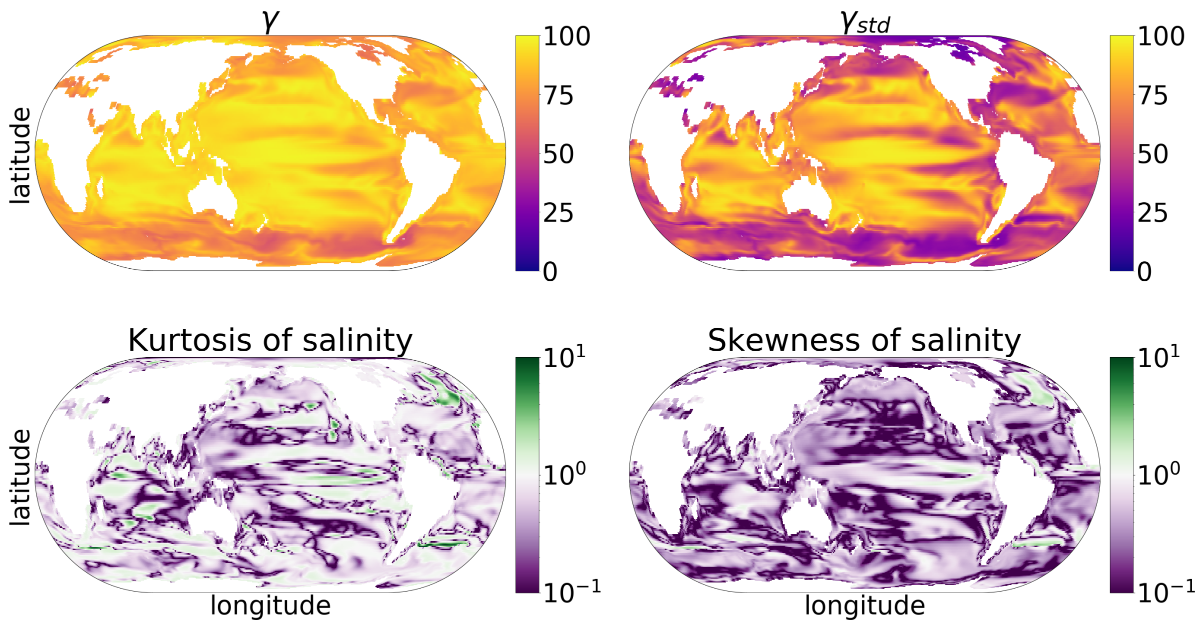

FIG. 7. Top: Intrinsic to total variability percentage for sea surface salinity. Bottom: Kurtosis and skewness of the ensemble mean of salinity at each grid point. Values closer to zero (within 0.5 of zero, purple shades) are considered approximately Gaussian.

4. Discussion

Our numerical experiments performed using γ on idealized Gaussian arrays show that γ is monotonic and decreases as the linear Pearson correlation coefficient increases. Thus aside from the qualitative differences the new metric finds when the data are non-Gaussian, the ranges of intrinsic versus total variability are quite different between γ and γ_{std}. This is to be expected from the different rates of increase with correlation seen in Figure 3. The traditional metric (γ_{std}) falls approximately linearly as the correlation coefficient increases, so that a correlation coefficient of 0.5 gives a γ_{std} just above 0.5. The new metric γ agrees with γ_{std} that correlation of 0 implies $\gamma = 1$, and correlation of 1 implies $\gamma = 0$, but for a correlation of 0.5 is closer to $\gamma = 0.9$. Only very near correlation coefficients of 1 does γ fall below 0.5. If roughly linear dependence on correlation coefficient is desired, γ can be raised to a power–γ^3 resembles γ_{std} and γ^6 resembles the correlation coefficient. These higher powers do not lose the ability to apply to non-Gaussian data nor become non-monotonic, but they will lose their interpretation as a ratio of bits of information entropy, and

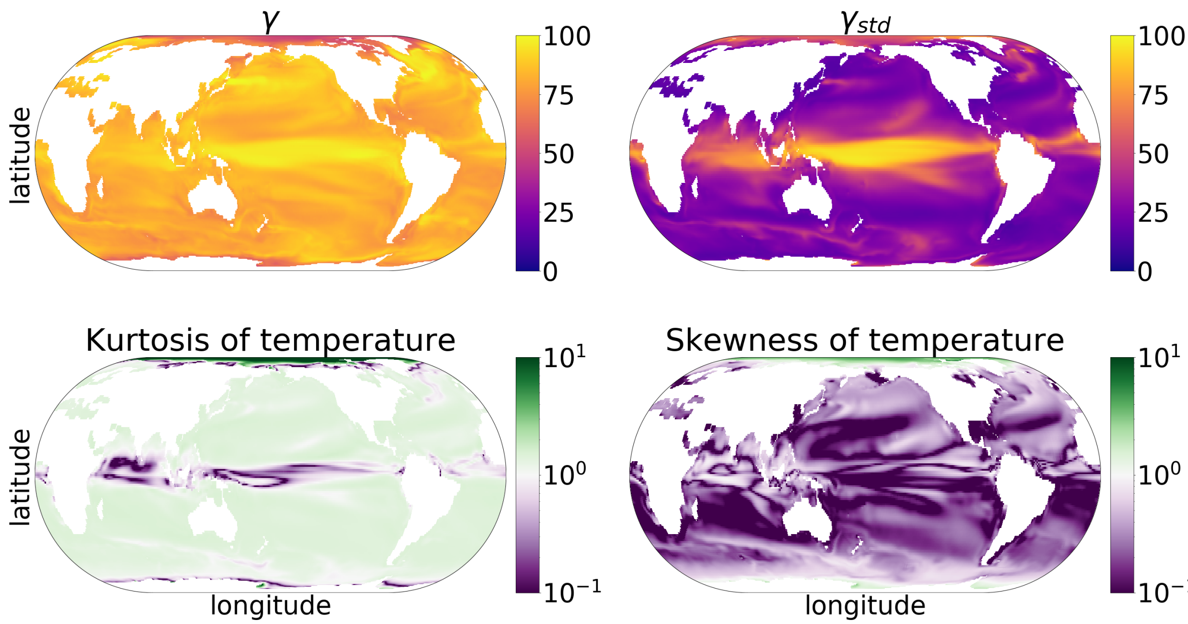

FIG. 8. Top: Intrinsic to total variability percentage for detrended sea surface temperature. Bottom: Excess kurtosis and skewness of the ensemble mean of temperature at each grid point. Values closer to zero (within 0.5 of zero, purple shades) are considered approximately Gaussian. The deviation of ensemble mean away from non normality implies that the ensemble members are also non normal. The Arctic regions have the most skewness and excess kurtosis implying non-Gaussian distributions.

instead reflect ratios of bits cubed of information entropy, etc. An alternative is to take γ_{std} raised to a different power: $\gamma_{std}^{1/3}$ is roughly similar to γ.

To check for sensitivity due to our binning choice, the endpoints of each bin were shifted by $\delta w/2$ and the results were compared. In theory, such a shift should not meaningfully affect the outcome, so this comparison gives a sense of how sensitive the results are to binning choices. For temperature, the raw GFDL-LE data (without detrending) gave an error of 4.7% in γ (see next section for definition) and detrended data gave an error of 11%. For salinity, the error in γ for raw data was 1.7% and for detrended data was 2%. Similar analysis for ROMS-OSOM coastal ensemble data gave negligible error for shifting the bin endpoints. Different binning strategies will be left to be explored for future research.

As can be seen in Figures 5, 6, and 7, information theory metrics show different patterns when compared to variance. Information theory metrics, especially mutual information, account for

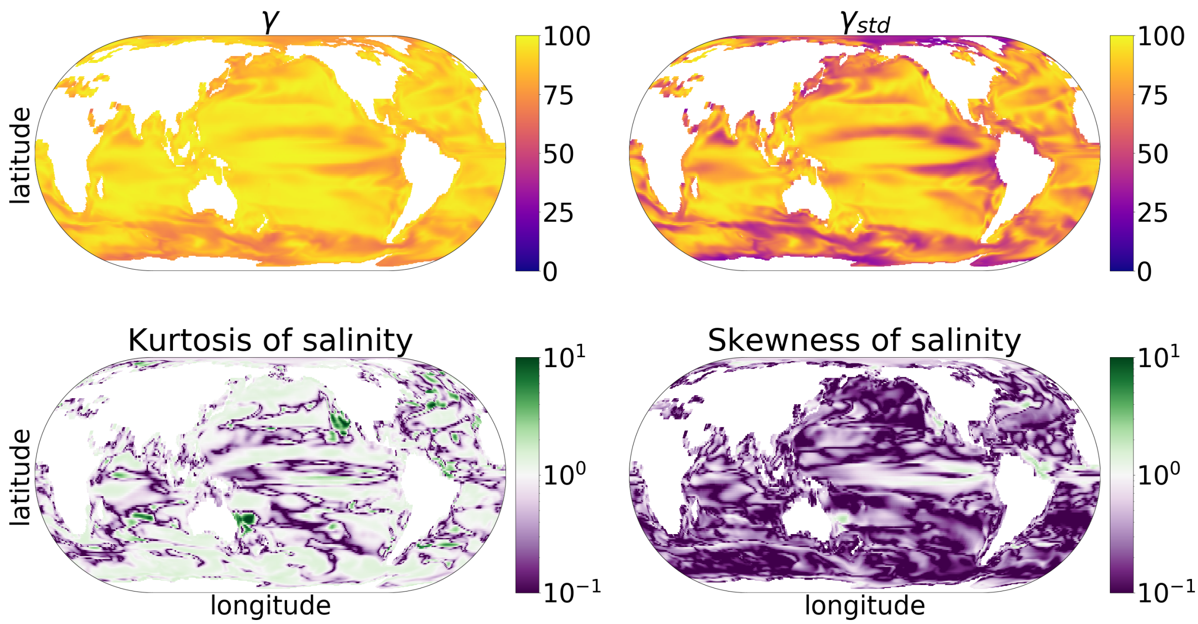

FIG. 9. Top: Intrinsic to total variability percentage for detrended sea surface salinity. Bottom: Kurtosis and skewness of the ensemble mean of salinity at each grid point. Values closer to zero (within 0.5 of zero, purple shades) are considered approximately Gaussian.

all non-linear shared information between the ensemble members and the mean including linear correlation, and this is one reason for the differences. We have argued that non-Gaussian statistics are another (which is not wholly independent of non-linear shared relationships). There are likely other aspects of differences between these metrics, but the management of these two expected aspects of geophysical fluids–nonlinear relationships and non-Gaussian distributions–justify the introduction of the new metric.

For the regional coastal model OSOM, forcings differ as to how they affect different variables. As might be expected, river runoff is more important for salinity than for temperature. However, for July-August, replacing rivers with the monthly-mean river flow gives nearly the same result (in terms of variability) as fully time-varying rivers. For the duration considered (July-August), averaging the river runoff gives similar effect for salinity as compared to giving the observed river runoff in the simulations, see Figure 10. Temperature is less sensitive to any of the forcing alterations, because although temperature and salinity are passive tracers they have different sources and sinks. Switching the wind product from NAM to NECOFS does not have any significant effect

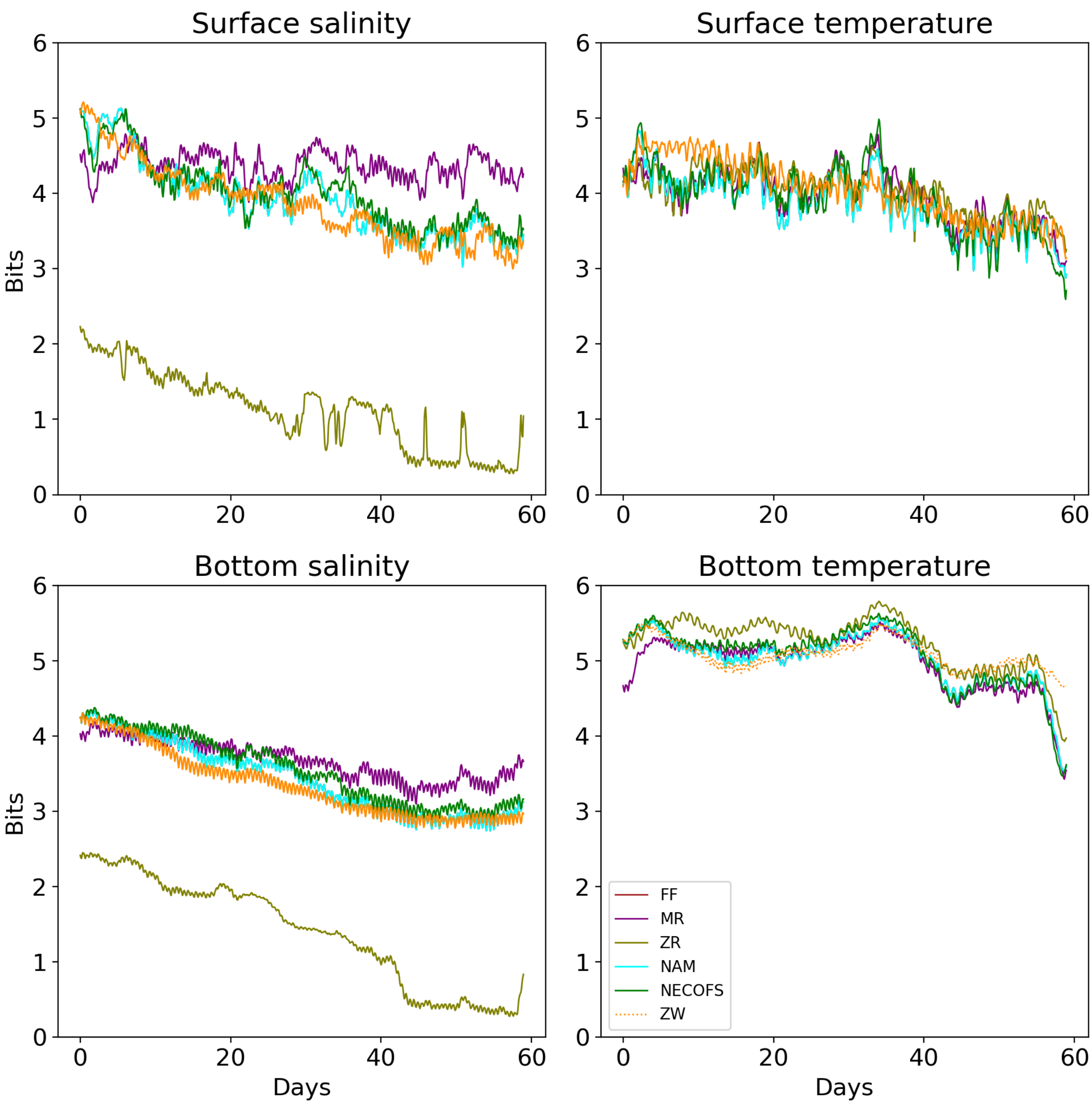

FIG. 10. Shannon entropy applied to temperature and salinity. Replacing fully time varying rivers with monthly-mean river flow gives almost the same result for salinity. Same is true by replacing wind product with a different one. Rivers set to zero affects salinity but not temperature. Winds are important in terms of variability but different wind products do not noticeably alter variability.

on the sources or sinks of temperature or salinity, but switching the wind off definitely affects the parameters by eliminating wind-driven mixing altogether. Figure 11 shows that zero wind (ZW) simulations are markedly different than the rest in terms of *mutual information* (i.e., they

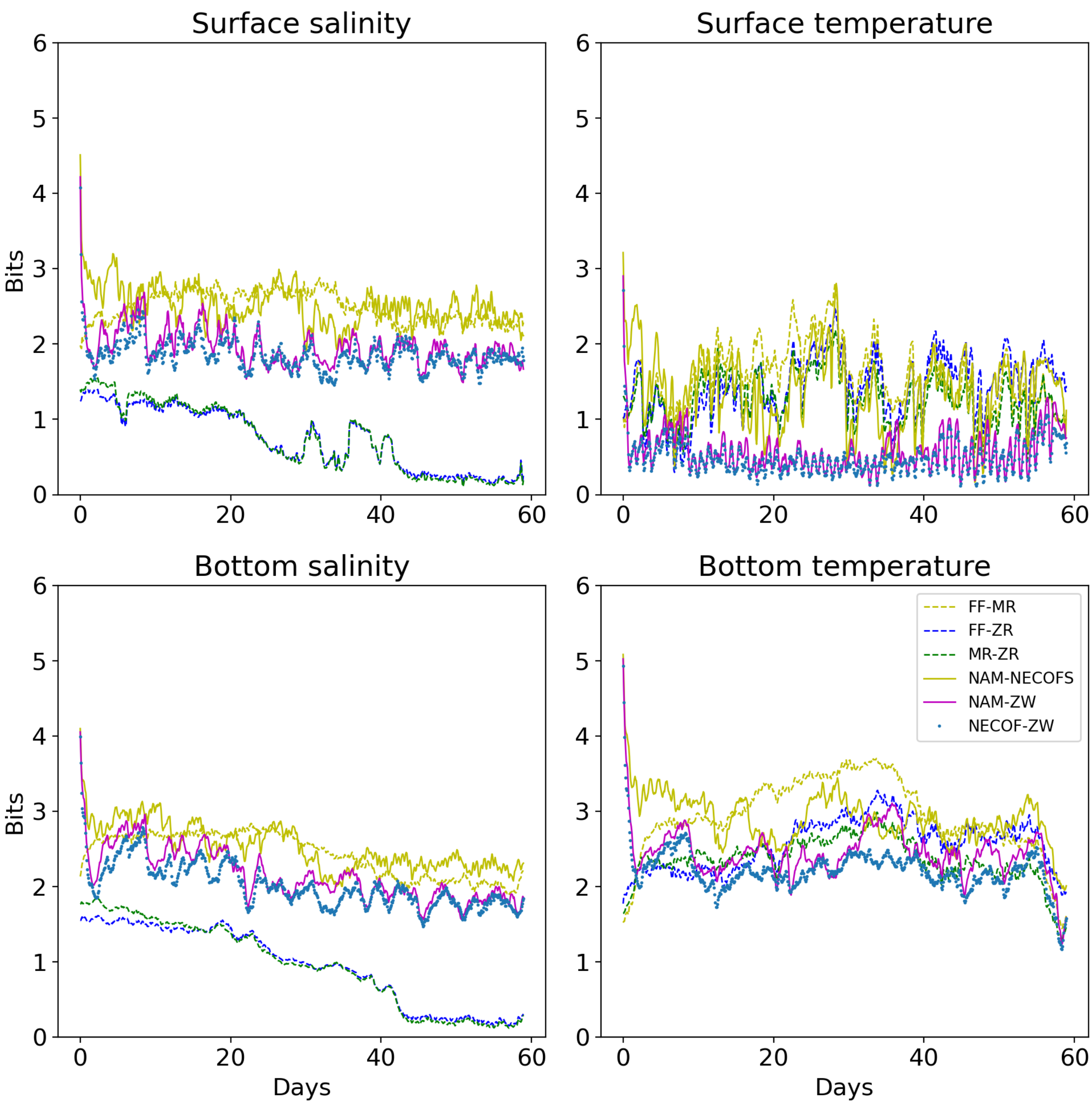

FIG. 11. Mutual information applied to simulations from different forcings. Higher mutual information implies higher similarity in terms of variability. For example NAM-NECOFS values are higher than NAM-ZW implying that NAM and NECOFS are significantly different than having no wind.

do not covary), although very similar in terms of amount of spatial variability (Shannon entropy, Figure 10), because even without winds tides, fluxes, and rivers still vary. The zero river case tends to eliminate both variability and mutual information (ZR). Please note that our simulations are for July-August, and results might be different for different season.

If we were to prioritize improvements based on Shannon entropy and mutual information, note that the two highest mutual information cases are where NAM is substituted with NECOFS and where mean rivers are substituted for varying rivers. The first observation is important from a forecast perspective, because it means that we can not easily tell the difference between different wind products, although something rather than zero winds should be used if the estuary needs to be forecasted for the full 20 day predictability range (weather forecasts are reliable for about 7 days in this region). Similarly, knowing that substituting the mean of the rivers for the fully varying rivers has little impact implies that rivers can be fixed in time for forecasts beyond where they might be predicted based on expected weather and precipitation. Finally, despite the fact that Narragansett Bay is a dominantly tidally-mixed estuary, among the sources of overall variability (i.e., sources of information entropy) considered, preserving an inflow of fresh water is key, even though that inflow can be steady. Winds do not appreciably increase information entropy of the Bay, but they are an important source of forced co-variation, and so are important for predictions but do not raise the overall level of variability.

5. Conclusion

We have proposed an information theory metric to determine contribution of intrinsic chaos and external variability to total variability in ensemble model simulations. Our metric uses Shannon entropy and mutual information and has several advantages over using only standard deviation (or variance). We have applied our metric on idealized Gaussian arrays as well as realistic coastal ocean and global climate model. We conclude that:

1. The new information theory metric is more reliable when outliers are present, because outliers get assigned less probability and because Gaussian distributions have a difficult time approximating long-tailed (i.e., outlier prone) distributions.

2. The new information theory metric is more reliable when variability is non-Gaussian because it is based on non-parametric measures of the probability distributions.

3. The new information theory metric varies monotonically with ensemble member to ensemble mean correlation, but is quantified in fraction of bits required to capture internal variability versus bits required to capture of total variability.

4. The use of the information theory metric in a coastal ocean model ensemble and a climate model ensemble qualitatively changes the focus to regions that were previously erroneously labeled as having high or low internal variability.

5. In this case, the coastal ensemble had a much smaller intrinsic (chaotic) proportion of its total variability in comparison to the climate ensemble had more intrinsic (weather, climate oscillations, etc.) as a proportion of its total. Importantly, the resolution of the models helps determine the proportion of intrinsic variability, so such comparisons are model-specific: a higher resolution coastal model might well have a larger intrinsic fraction than a coarser climate model.

Acknowledgments. The Rhode Island Coastal Ecology Assessment Innovation & Modeling grant (NSF 1655221) supported this work. BFK was also supported by ONR N00014-17-1-2963. This material is based upon work conducted at a Rhode Island NSF EPSCoR research facility Center for Computation and Visualization (Brown University), supported in part by the National Science Foundation EPSCoR Cooperative Agreement #OIA-1655221.

Data availability statement. All the data and the codes used to plot results can be downloaded via Brown University's digital archive DOI: `urlplaceholder`.

References

Beardsley, R. C., and C. Chen, 2014: Northeast coastal ocean forecast system (necofs): A multi-scale global-regional-estuarine fvcom model. *AGUFM*, **2014**, OS23C–1211.

Brissaud, J. B., 2005: The meanings of entropy. *Entropy*, **7 (1)**, 68–96, https://doi.org/10.3390/e7010068.

Carcassi, G., C. A. Aidala, and J. Barbour, 2019: Variability as a better characterization of shannon entropy. *arXiv preprint arXiv:1912.02012*.

Correa, C. D., and P. Lindstrom, 2013: The mutual information diagram for uncertainty visualization. *International Journal for Uncertainty Quantification*, **3 (3)**.

Cover, T. M., 1999: *Elements of information theory*. John Wiley & Sons.

DelSole, T., and M. K. Tippett, 2007: Predictability: Recent insights from information theory. *Reviews of Geophysics*, **45 (4)**.

DelSole, T., and M. K. Tippett, 2018: Predictability in a changing climate. *Climate Dynamics*, **51 (1)**, 531–545.

Demirtas, H., 2014: Generating bivariate uniform data with a full range of correlations and connections to bivariate binary data. *Communications in Statistics-Theory and Methods*, **43 (17)**, 3574–3579.

Deser, C., and Coauthors, 2020: Insights from earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, 1–10.

Frankcombe, L. M., M. H. England, M. E. Mann, and B. A. Steinman, 2015: Separating internal variability from the externally forced climate response. *Journal of Climate*, **28 (20)**, 8184–8202.

Franzke, C. L., and Coauthors, 2020: The structure of climate variability across scales. *Reviews of Geophysics*, **58 (2)**, e2019RG000 657.

Gomez, B. G., 2020: Intrinsic ocean variability modulated by the atmosphere in the gulf of mexico: an ensemble modelling study. Ph.D. thesis, Université Grenoble Alpes [2020-....].

Hacine-Gharbi, A., P. Ravier, R. Harba, and T. Mohamadi, 2012: Low bias histogram-based estimation of mutual information for feature selection. *Pattern recognition letters*, **33 (10)**, 1302–1308.

Hartley, R. V. L., 1928: Transmission Information. *Bell System Technical Journal*, **7 (3)**, 535–563.

Hawkins, E., and R. Sutton, 2012: Time of emergence of climate signals. *Geophysical Research Letters*, **39 (1)**.

Jaynes, E. T., 1962: Information theory and statistical mechanics. Brandies University Summer Institute Lectures in Theoretical Physics.

Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *Journal of the atmospheric sciences*, **59 (13)**, 2057–2072.

Knuth, K. H., 2019: Optimal data-based binning for histograms and histogram-based probability density models. *Digital Signal Processing*, **95**, 102 581.

Kowal, R. R., 1971: 296. note: Disadvantages of the generalized variance as a measure of variability. *Biometrics*, **27 (1)**, 213–216, URL http://www.jstor.org/stable/2528939.

Kowalski, A. M., M. T. Martin, A. Plastino, and G. Judge, 2012: On extracting probability distribution information from time series. *Entropy*, **14 (10)**, 1829–1841, https://doi.org/10.3390/e14101829.

Leroux, S., T. Penduff, L. Bessières, J.-M. Molines, J.-M. Brankart, G. Sérazin, B. Barnier, and L. Terray, 2018: Intrinsic and atmospherically forced variability of the amoc: insights from a large-ensemble ocean hindcast. *Journal of Climate*, **31 (3)**, 1183–1203.

Leung, L.-Y., and G. R. North, 1990: Information theory and climate prediction. *Journal of Climate*, **3 (1)**, 5–14.

Liang, X. S., 2013: The liang-kleeman information flow: Theory and applications. *Entropy*, **15 (1)**, 327–360.

Liang, X. S., 2014: Entropy evolution and uncertainty estimation with dynamical systems. *Entropy*, **16 (7)**, 3605–3634.

Liang, X. S., and R. Kleeman, 2005: Information transfer between dynamical system components. *Physical review letters*, **95 (24)**, 244 101.

Liang, Y.-c., and Coauthors, 2020: Quantification of the arctic sea ice-driven atmospheric circulation variability in coordinated large ensemble simulations. *Geophysical Research Letters*, **47 (1)**, e2019GL085 397.

Llovel, W., T. Penduff, B. Meyssignac, J.-m. Molines, L. Terray, L. Bessières, and B. Barnier, 2018: Contributions of atmospheric forcing and chaotic ocean variability to regional sea level trends over 1993–2015. *Geophysical Research Letters*, **45 (24)**, 13–405.

Majda, A. J., and B. Gershgorin, 2010: Quantifying uncertainty in climate change science through empirical information theory. *Proceedings of the National Academy of Sciences*, **107 (34)**, 14 958–14 963.

Milinski, S., N. Maher, and D. Olonscheck, 2019: How large does a large ensemble need to be. *Earth Syst. Dynam. Discuss., 2019, 1–19, doi: 10.5194/esd-2019*, **70**.

Papana, A., and D. Kugiumtzis, 2008: Evaluation of mutual information estimators on nonlinear dynamic systems. *Nonlinear Phenomena in Complex Systems*, **11 (2)**, 225–232.

Rodgers, K. B., J. Lin, and T. L. Frölicher, 2015: Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an earth system model. *Biogeosciences*, **12 (11)**, 3301–3320, https://doi.org/10.5194/bg-12-3301-2015, URL https://bg.copernicus.org/articles/12/3301/2015/.

Sane, A., B. Fox-Kemper, D. S. Ullman, C. Kincaid, and L. Rothstein, 2021: Consistent predictability of the ocean state ocean model (osom) using information theory and flushing timescales. *Journal of Geophysical Research: Oceans*, e2020JC016875, https://doi.org/https://doi.org/10.1029/2020JC016875, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JC016875.

Schneider, T., and S. M. Griffies, 1999: A conceptual framework for predictability studies. *Journal of climate*, **12 (10)**, 3133–3155.

Schurer, A. P., G. C. Hegerl, M. E. Mann, S. F. Tett, and S. J. Phipps, 2013: Separating forced from chaotic climate variability over the past millennium. *Journal of Climate*, **26 (18)**, 6954–6973.

Shannon, C., 1948: A Mathematical Theory of Communication. *Bell System Technical Journal*, **27 (April 1928)**, 379–423,623–656, URL http://math.harvard.edu/{~}ctm/home/text/others/shannon/entropy/entropy.pdf.

Shchepetkin, A. F., and J. C. McWilliams, 2005: The regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean modelling*, **9 (4)**, 347–404.

Stevenson, S., B. Rajagopalan, and B. Fox-Kemper, 2013: Generalized linear modeling of the el niño/southern oscillation with application to seasonal forecasting and climate change projections. *Journal of Geophysical Research: Oceans*, URL http://dx.doi.org/10.1002/jgrc.20260, in press.

Stone, J. V., 2015: *Information theory: a tutorial introduction*. Sebtel Press.

Waldman, R., S. Somot, M. Herrmann, F. Sevault, and P. E. Isachsen, 2018: On the chaotic variability of deep convection in the mediterranean sea. *Geophysical Research Letters*, **45 (5)**, 2433–2443.

Watanabe, S., 1960: Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, **4 (1)**, 66–82, https://doi.org/10.1147/rd.41.0066.

Yettella, V., J. B. Weiss, J. E. Kay, and A. G. Pendergrass, 2018: An ensemble covariance framework for quantifying forced climate variability and its time of emergence. *Journal of Climate*, **31 (10)**, 4117–4133.