Posted on 10 Aug 2023 — CC-BY-NC 4 — https://doi.org/10.1002/essoar.10505545.4 — This a preprint and has not been peer reviewed. Data may be preliminary.

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

August 10, 2023

Abstract

We demonstrate the use of information theory metrics, Shannon entropy and mutual information, for measuring internal and forced variability in general circulation coastal and global ocean models. These metrics have been applied on spatially and temporally averaged data. A combined metric reliably delineates intrinsic and extrinsic variability in a wider range of circumstances than previous approaches based on variance ratios that therefore assume Gaussian distributions. Shannon entropy and mutual information manage correlated fields, apply to any distribution, and are insensitive to outliers and a change of units or scale. Different metrics are used 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. A metric based on information theory partly agrees with the traditional variance-based metric and identifies regions where non-linear correlations might exist. Mutual information and Shannon entropy are used to quantify the impact of different boundary forcings in a coastal ocean model ensemble. Information theory enables ranking the potential impacts of improving boundary and forcing conditions across multiple predicted variables with different dimensions. The climate model ensemble application shows how information theory metrics are robust even in a highly skewed probability distribution (Arctic sea surface temperature) resulting from sharply non-linear behavior (freezing point).

Internal vs Forced Variability Metrics for General Circulation Models Using Information Theory

Aakash Sane[1*], Baylor Fox-Kemper[2], and David S. Ullman[3]

[1]School of Engineering, Brown University, Providence, Rhode Island, USA
[2]Department of Earth, Environment, and Planetary Sciences, Brown University, Providence, Rhode Island, USA
[3]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

Ocean model simulations show variability due to intrinsic chaos and external forcing (air-sea fluxes, river input, anthropogenic emissions, etc.). It is important to estimate their contributions to total variability for attribution. Using variance to estimate variability might be unreliable due to the existence of higher statistical moments. We show the use of non-parametric information theory metrics, Shannon entropy and mutual information, for measuring internal and forced variability in coastal and global ocean models. These metrics are applied on spatially and temporally averaged data. Metrics delineate variability in a wider range of circumstances than previous approaches based on variance ratios that assume Gaussian distributions. Metrics work on correlated fields, apply to any distribution, and are insensitive to outliers and a change of units or scale. Metrics are applied to (1) idealized data, (2) ensemble of a realistic coastal ocean model (OSOM), (3) GFDL-ESM2M large ensemble. The information theory metric partly agrees with the variance-based metric and possibly identifies regions of non-linear correlations. The metric detects higher intrinsic variability in the Arctic region as compared to the variance metric. The climate model ensemble application shows how information theory metrics are robust in a skewed probability distribution (Arctic sea surface temperature) resulting from sharply non-linear behavior (freezing point). In different experiments, we quantify sensitivity of OSOM to changes in forcing. Variations in the river runoff and changing the wind product do not add information (variability). Information theory enables ranking the impacts of improving boundary and forcing conditions across multiple variables with different dimensions.

Plain Language Summary

It is important in climate modeling to distinguish variability caused by external forces versus variability that arises within the system to estimate causes in a particular variation. Disturbances from the atmosphere such as wind, solar heating and cooling, anthropogenic emissions are external disturbances and variations due to swirls are internal chaotic disturbances. We use information theory - a way to quantify the amount of variability in these models. Here, we study multiple runs of a coastal ocean model and an ocean climate model. We found that it matters a lot how you measure the internal and external variability. Making fewer assumptions about the statistics of variability proved more robust, especially in the Arctic in global model and at depth in an estuary. For the global model, we found internal chaos to dominate temperature variations in the Arctic in contrast to variations in salinity. In a different set of experiments, the coastal model was run by slightly changing the wind, averaging the river input instead of the full river flow, etc. We found that we cannot neglect river input. Different winds had the same impact. These experiments reveal the importance of uncertainty in forcing conditions to help us design a forecasting system.

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 to elucidate the causes responsible for observed variability in models, help to identify key observable metrics, and help quantify concepts such as the time of emergence of climate signals (Hawkins & 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); Y.-C. Liang et al. (2020)). 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). Initial condition ensembles are a set of simulations sharing the same forcing and the same governing equations and identical parameterizations, but they still diverge from one another because slightly different initial conditions evolve into substantially different conditions later in the simulations due to intrinsic chaos–most geophysical fluid dynamics models and climate models are intrinsically chaotic. Most of the discussion here will focus on initial condition ensembles, but the metrics proposed can be adapted to both types of ensembles.

To help visualize variability, a generic idealized 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 solid black line is the signal due mainly to extrinsic factors (aside from the limits of the finite ensemble size) and the spread of the model (schematized by the double-headed magenta arrow in Figure 1) can be considered due to intrinsic variability or internal chaos.

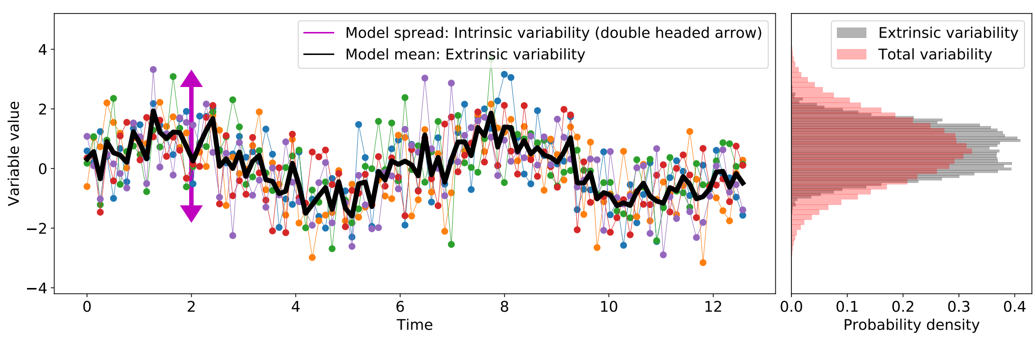

Figure 1. A sketch of a typical ocean or climate model output for an arbitrary variable. Each ensemble is shown in a different color, and the mean of the ensemble is shown as a black line. The ensemble mean can be considered to be the trend set by external forcings. The model spread shown by the double-headed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

One method of quantifying intrinsic and extrinsic variability is to look at variances (second central statistical moment) of the model spread and the mean of the model (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 distribution. 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 a 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, often estimated by a histogram formed after appropriately binning the data, which then approximates a distribution with a discrete probability p_i as a fraction over all states of the visited system. A histogram thus makes the estimated and visited number of states discrete rather than continuous. Information entropy metrics measure variability by taking into account the probability distribution of the binned data, drawing on the concept from statistical mechanics of entropy in quantifying the number of microstates that a variable can occupy. The fundamental measure in information theory is the Shannon entropy (Shannon, 1948) (a.k.a. the information entropy) that characterizes the amount of variability in a variable (Carcassi et al., 2021). Mutual information, another metric introduced by Shannon (1948), measures how much information a variable contains about another variable.

Information theory is applied in signal processing, computer science, statistical mechanics, quantum mechanics, etc. It is used to quantify the 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. Information theory metrics are not new to climate sciences. They have been introduced in predictability studies, evaluating the skill of statistical models, as well as uncertainty studies (Leung & North, 1990; Schneider & Griffies, 1999; Kleeman, 2002; DelSole & Tippett, 2007; Majda & Gershgorin, 2010; Stevenson et al., 2013) and recently in studying variability (Gomez, 2020), coastal predictability (Sane et al., 2021) and drivers of drought (Shin et al., 2023).

In the two parts of 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. We use Shannon entropy and mutual information and a particularly useful combination of the two. We indirectly employ conditional entropy, which depends on Shannon entropy and mutual information but is less intuitive so is not discussed in detail. Recent theoretical advances in understanding dynamical systems through the lens of information theory relate causality analysis and information transfer (e.g., X. S. Liang (2014)). Although important, this theory has had few concrete applications. Even the basic information theory concepts (Shannon entropy and mutual information) have enjoyed only limited adoption by the oceanic and atmospheric community, primarily arising in predictability quantification (e.g., Sane et al. (2021)). We begin to bridge the gap with a pragmatic framework which can be easily replicated and improved upon, including causality analysis and the evolution of entropy within modeling systems like those studied here.

In Part 1, we apply this intrinsic vs. extrinsic metric to three sets of data: 1) Idealized Gaussian and uniformly distributed arrays with specified correlation, 2) Initial condition ensemble output of a regional coastal model (OSOM) (Sane et al., 2021) over July-August 2006 where most variables are not Gaussian, and 3) The GFDL-ESM2M Large Ensemble (Rodgers et al., 2015; Deser et al., 2020), an climate model initial condition ensemble hereby referred to as GFDL-LE. This large ensemble dataset contains historical and future projection data following the RCP 8.5 scenario. All the GFDL-LE monthly mean data from 1950 to 2100 were used in the analysis.

In Part 2, we use OSOM to demonstrate the use of Shannon entropy and mutual information to quantify the extrinsic forcing effects of altered boundary forcing types. For example, is wind forcing dominant over river forcing, does using temporal averaged river runoff cause any appreciable changes in estuarine circulation, or does change in the wind product alter circulation? In coastal and estuarine systems, knowledge of which forcings are dominant helps prioritize data collection and refinement of the most impactful forcings.

1.1 Information theory

We 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 when 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 a low probability event has high uncertainty associated with it. Thus, the prefactor p_i is 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 the quantification of variability than infrequent events.

Stone (2015) gives an intuitive way to understand Shannon entropy using a binary tree. A binary tree is a tree chart which starts with one node and splits into two branches 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 the entropy is h then 2^h states are possible.

A second metric from Shannon (1948) which is also widely used is *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 the 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 $p_{ij}(\log_2(p_{ij}) - \log_2(p_i) - \log_2(p_j))$. 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 the amount of 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 due to knowing y or vice versa and denotes how much information is transmitted between the two variables.

In the context of ocean or climate modeling, entropy can be used to measure variability in a model output or available data. This is in tandem with the interpretation of Shannon entropy in physical sciences as given in Carcassi et al. (2021). When calculating the Shannon entropy, the primary concern is counting the possible states, e.g. the various bins in a histogram, where the variable can go into while any assigned bin value or its dimensions are of lesser importance. Entropy metrics measure variability in *bits* (when the 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 & Lindstrom, 2013) and discontinuous distributions, including states visited intermittently. 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 objective of the paper. 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 the effects of using different boundary conditions.

2 Methods

2.1 Part A: Intrinsic and Extrinsic Variability for Ensemble Data

Analysis begins on each grid point on the ocean surface or ocean bottom. Let a variable in 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. The total number of members of the ensemble is N and each member has M time steps. At a particular grid point $f(n,t,x,y)$ is $f(n,t)$. To obtain the signal due to extrinsic forcings, the "differencing" approach (Frankcombe et al., 2015) has been followed to estimate the forced response. This approach involves averaging the members of the ensemble to derive *ensemble mean*. The ensemble mean is given by the following:

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

$g(t)$ is a single time-varying signal for each grid point obtained by averaging across the ensemble members. There are potential problems with assuming that 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) and Johnson et al. (2023). 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. The 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 that differs from one ensemble member to another. Note that the above decomposition takes place at each grid point. In Figure 1a, $f(n,t)$ are shown by multi-colored ensemble members. $g(t)$ is shown by a thick black line. As seen in Figure 1b, $g(t)$ has a probability distribution shown in gray and subsequently has the first, second and possibly important higher statistical moments. The gray density histogram shows variability due to extrinsic factors, and the pink density histogram shows total variability given by extrinsic and intrinsic factors.

2.1.1 Evaluating entropies

The ensemble simulation data has been used without detrending to evaluate $g(t)$ and $\eta(n,t)$. Detrending will remove some nonstationarity from the data, but will also remove some part of the extrinsic variability. Our aim is not to determine the forced response but to estimate the degree of *variability* contributed by the forced response (extrinsic response) and the intrinsic variability originating from the intrinsic chaos. Metrics have been calculated at each grid point by treating them independently.

Usually we are limited in the number of ensemble members due to computational costs, so we concatenate into 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:

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

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

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

This enables wide sampling and obtains an accurate probability distribution for f (assuming approximate stationarity, or enforcing stationarity by detrending), and allows 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 because this is a sufficient demonstration of the value of information theory metrics. A time-evolving analysis raises additional issues about causality and the shifting probabilities distributions of climate states that are not the focus here (X. S. Liang, 2013; DelSole & Tippett, 2018). By using the whole time series, we treat 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 the same number of elements are necessary to evaluate joint probability distribution. This enables us to calculate the mutual information between f and g.

Calculating the Shannon entropy of f and the mutual information between f and g is a difficult task that necessitates careful consideration. Optimal binning for precise measurement of information entropies is a research topic in itself, and various techniques have been proposed, such as equidistant partitioning, equiprobable partitioning, k nearest neighbor, usage of B-spline curves for binning to name a few (Hacine-Gharbi et al., 2012; Kowalski et al., 2012; Knuth, 2019). A comprehensive review of these methods can be found in Papana and Kugiumtzis (2008). Although the histogram binning technique is one of the most commonly used techniques (for example Campuzano et al. (2018); Pothapakula et al. (2019); Shin et al. (2023)), it introduces uncertainty. There are several techniques to estimate this uncertainty, such as the one proposed in Knuth et al. (2005). In this article, we use histograms with equidistant partitioning where constant optimal bin widths are determined using the Freedman-Diaconis rule (Freedman & Diaconis, 1981; Knuth, 2019) at each grid point to get a discrete probability distribution. The same bin width was used for the marginal and joint probability distributions. Two approaches were used to estimate the sensitivity of the metric to binning: varying the bin width around the optimal value and bootstrapping over the ensemble members. The metrics were found to be more sensitive to changes in the bin widths than to bootstrapping. Therefore, to estimate uncertainty, if the width of the bin was found to be δw, then it was varied from $0.5\delta w$ to $1.5\delta w$ to obtain a reasonable estimate of uncertainty. Sweeping across the number of bins was performed also in (Sane et al., 2021) to get an estimate of predictability time-scale.

2.1.2 Information theory based 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 *ratio of intrinsic variability to total variability.* When $f \to g$, then $I(f;g) \to H(f) = H(g)$ 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 and total chaos.

Related quantities appear in other applications. The quantity $I(f;g)/H(f)$ is defined as "uncertainty coefficient" (Eshima, 2020). It is the ratio of entropy of f explained by g. $H(f)$ 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 $H(X|Y) = \sum p(x|y)\log_2 p(x|y)$ (Cover, 1999). It is not necessary to calculate the conditional entropy to arrive at γ. $H(X|Y)$ gives the average uncertainty about the value of f after g is known, or just the uncertainty in f that is not attributed to g but is attributed to 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.

$I(f;g)$ takes into account any correlation or information shared between f and g. This is vital because even though the spread of the model η is treated similarly to the noise added to the mean signal, it might be that the spread of the model depends on the mean signal. A simple example is that if the model spread is relative (e.g., 10% of the mean signal, or *multiplicative noise*), rather than absolute (e.g., 2 units, or *additive noise*), then there is information about the model spread contained in the ensemble mean signal. The nonlinear and chaotic nature of fluids often leads the mean flow to amplify the chaotic signal (e.g., eddies) and thereby result in altered variability statistics that can be represented as multiplicative noise.

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 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 $2^{H(f)} = 2^{I(f;g)}2^{H(f|g)}$. Another analog of a component of the climate system 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 a completely noisy channel where the output is independent of the input will give γ as 1.

A seemingly enticing and simpler alternative is $\gamma = 1 - \frac{H(g)}{H(f)}$, i.e. just the difference between the entropy of the ensemble and the mean entropy as a ratio with the entropy of the ensemble. 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 the mutual information between the ensemble members and their mean in estimating intrinsic variability.

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, the 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 & 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.

2.1.3 Variance based metric

A variance based metric as given in (Leroux et al., 2018) has been utilized to compare with 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 the temporal averaging. 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 the total variability $\left(\sigma_g^2 + \overline{\sigma_\eta^2(t)}\right)^{1/2}$ is equivalent to $H(f)$. Therefore, γ is compared to γ_{std} given by

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

2.2 Part B: Information Entropy and Boundary Forcing

2.2.1 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

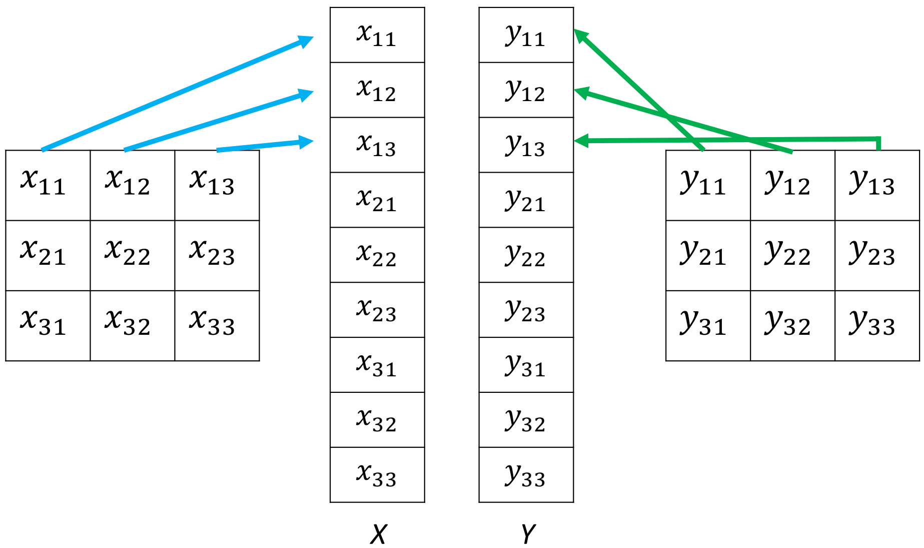

Figure 2. Flattening process for comparing two-dimensional fields using Shannon entropy and mutual information. As the flattened arrays $x_1, x_2, ...$ and $y_1, y_2...$ may not have linear dependence on each other, using linear dependence measures such as Pearson's correlation might produce incorrect results. Mutual information measures nonlinear correlations and hence captures all linear and nonlinear dependence.

be required for coastal applications 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, air-sea fluxes, etc. All model details 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. One control and four altered forcing sets were utilized,

1. (Control) Full atmospheric forcing 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 FF. FF stands for full forcing.
2. Full set of atmospheric forcing, but using the winds of the Northeast Coastal Ocean Forecast System (NECOFS) winds (Beardsley & 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 forcing sets have been tabulated in Table 1. The aim is to quantify the effect on total variability by removing or altering one of many processes that might contribute.

To evaluate spatial Shannon entropy, the spatial output at a particular instant in 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-dimensional array (flattened) by concatenation. Shannon

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

Table 1. Different types of forcing combinations were used 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.

entropy was found for the flattened variable at each time step to obtain a time-varying entropy of each surface or bottom variable.

Mutual information was applied between the flattened row vectors. Our focus is on a pragmatic approach to using information theory for relative comparisons among simulations, rather than an equation for the evolution of Shannon entropy and mutual information with respect to time (see X. S. Liang and Kleeman (2005)). For example, if mutual information on surface salinity between FF and MR is higher than between FF and ZR, this implies that the penalty for using time-averaged river runoff is not as severe as using zero river runoff. The replacement of FF with MR will give more similar results to FF than replacing FF with ZR will. We can interpret this to indicate that small errors in river runoff flow rates will not cause appreciable changes to surface salinity while using zero rivers will strongly impact the solution.

3 Results

3.1 Part A: Intrinsic and Extrinsic Variability Results for Ensemble Data

3.1.1 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[1] only on the standard deviation σ. 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 correlation coefficient between 0 and 1.

Thus, the 10 arrays are linearly correlated 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, that is, as internal variability dominates forced.

[1] $H = \log_2 2\pi e\sigma^2$ 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). Throughout this article, discretely sampled and binned probability distributions are obtained directly from the data without any further parameterization

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, relative comparisons (more intrinsic fraction in one region vs. a different region) are preserved.

A second related experiment was derived from the first and 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 metrics. Figure 3 shows that γ is insensitive to outliers while γ_{std} is not. γ is not sensitive because outliers occur less frequently and therefore do not greatly affect the probability distribution, 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 experiments. The results for a Gaussian ensemble of 10 members are shown in Figure 3 a and 100 members in Figure 3 b.

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

3.1.2 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 & McWilliams, 2005) to model Narragansett Bay and the surrounding coastal oceanic regions and waterways. OSOM's primary purpose is to understand 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 has been performed using perturbed initial (ocean) conditions under the same atmospheric and tidal forcing for the months July and August of 2006. This ensemble consists of 10 members. Data during the first predictability window (20 days) where results are still linked directly to the initial conditions have been ignored and the remaining simulation has been used to examine variability within the "climate projection" of the model beyond when forecasts or predictions that rely deterministically on initial conditions are possible. During this phase the different ensemble members visit different possible futures within the envelope of the projected "climate" (see the related application of information theory to assess predictability in Sane et al. (2021)). The modeled temperature and salinity at each grid point typically do not follow Gaussian distributions as the skewness and kurtosis each grid point shown in Figure 4 for salinity and temperature of the sea surface and bottom 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

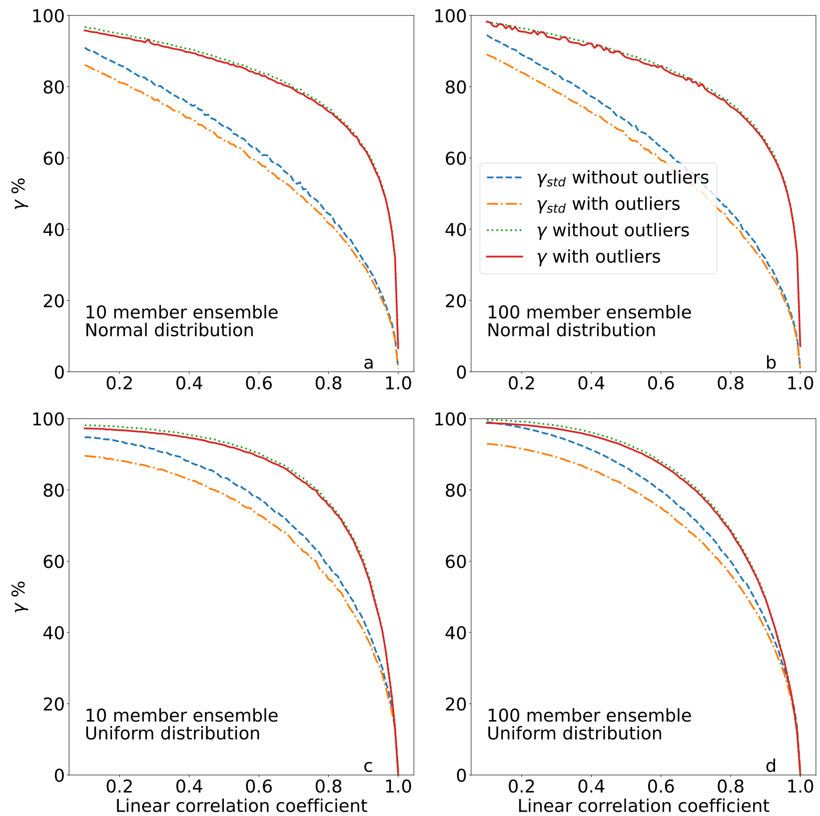

Figure 3. Information theory metric of intrinsic vs. extrinsic variability γ as a function of the correlation coefficient in idealized Gaussian correlated arrays (a and b) and idealized uniformly distributed arrays (c and d). The horizontal axis is the correlation coefficient between the mean member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

case of excess kurtosis a constant value of 3 is subtracted. For Gaussian distributions, both skewness and excess kurtosis 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 and thus these variables are considerably non-Gaussian in OSOM. Therefore, the variance method in 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 7 shows the ratio of intrinsic variability to total variability applied at every point in the OSOM grid. γ_{std} is displayed on left whereas γ is shown in the center for comparison. The uncertainty in γ has been plotted in the third column in Figure 7. 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, it 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. The Eastern North-South passageway of the bay shows different structure of γ than γ_{std} of salinity.

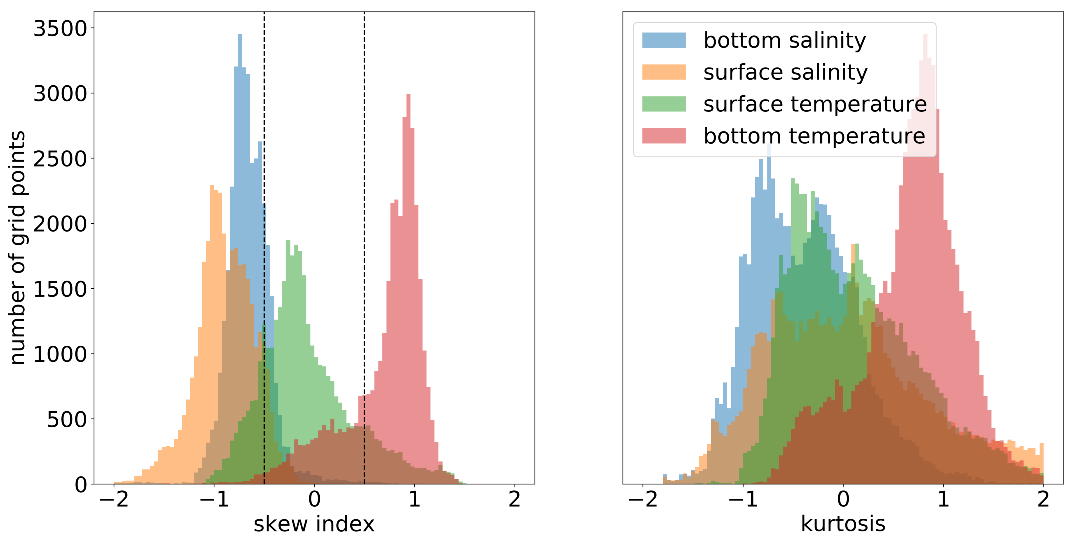

Figure 4. Grid point-wise skewness and excess kurtosis for OSOM output. Neither are close to zero, e.g., within (-0.5, 0.5), suggesting that the temperature and salinity data distribution is non-Gaussian.

3.1.3 Earth System Model Large Ensemble

A complementary experiment was performed using γ to evaluate internal versus forced variability in global climate simulation output for the RCP8.5 climate change scenario using the GFDL-LE model (randomly selected among the models compared). The 30 members of the ensemble were utilized. The variability of sea surface temperature (Figures 5) and sea surface salinity (Figures 6) were estimated using both γ and γ_{std}.

Note in particular the Arctic sea surface temperatures in Figure 5, which have a highly skewed and excessive kurtosis distribution due to the freezing point of seawater. The standard metric (γ_{std}) considers this region to be among the most intrinsically variable in the world, while the information theory metric considers it as a region of middling intrinsic variability–much lower than the equatorial regions where El Nino variability is profound. This region is also subject to intermittent and drastic swings in salinity as sea ice forms and melts, but note that the standard metric indicates low salinity variability while the information theory metric ranks it as high in Figure 6. It is clear that a Gaussian metric should not be applied to the Arctic due to the skewness and excess kurtosis, and in this case the inference is opposite using the standard and information theory metrics. In the equatorial Pacific, where Gaussian statistics are more reliable, the two metrics agree that internal variability is high.

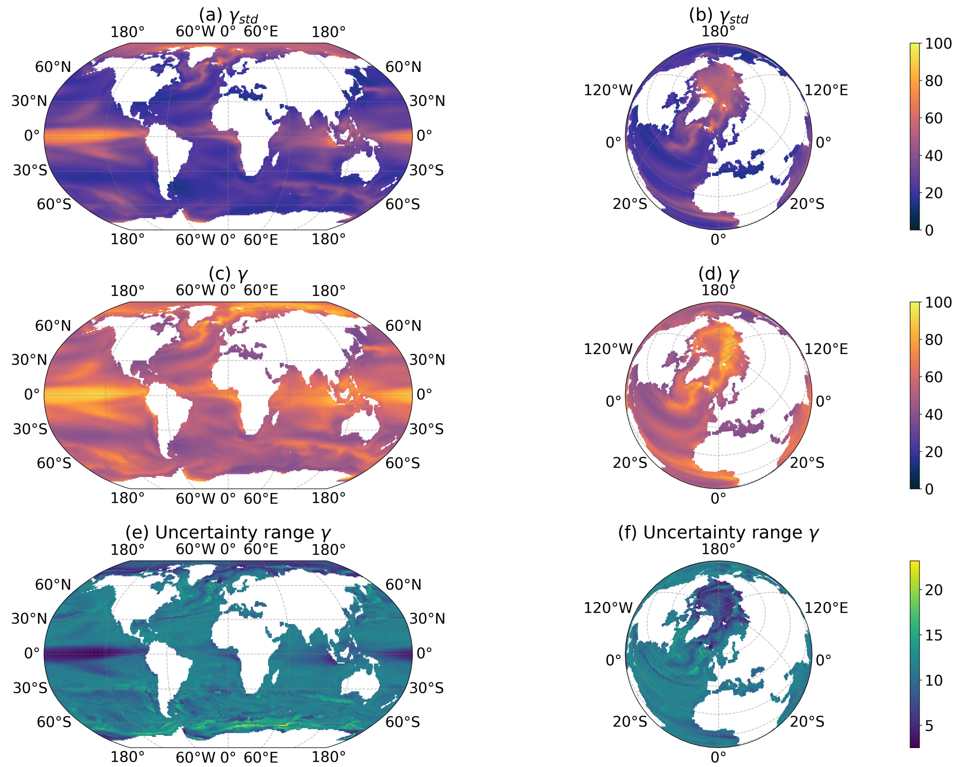

Figure 5. Intrinsic to total variability for sea surface temperature using (a, b) γ_{std} and (c, d) γ. (e, f) Uncertainty range in γ found by sweeping across the bin width as explained in the text. We can see a difference in the magnitude and pattern of the intrinsic to total variability around the Arctic region. Difference in other regions such as Mediterranean sea and Pacific equator is also visible.

A less drastic failure occurs from the modest excess kurtosis in extratropical temperatures and in a few isolated regions in surface salinity. These regions are also non-Gaussian but are also not heavily skewed (i.e., they are more long-tailed and intermittent than Gaussian). These regions differ in the 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 expected when the correlation coefficients are low in Figure 3.

3.2 Part B: Information Entropy and Boundary Forcing Results

3.2.1 Impact due to changes in boundary conditions in coastal models:

We show the results of the coastal model analysis under different forcing in Figures 8 and 9, under the same region as shown in Figure 7. The entropy has been plotted with respect to time as some variability occurs. In Figure 8, Shannon entropy is plotted for spatial quantities. For example, for surface salinity, all surface values have been considered to find the Shannon entropy using the flattening approach. If Shannon entropy is more or less equal for two forcings, it implies that they similarly affect variability. Both winds and rivers seem to have similar effects in this regard. However, Figure 9 displays mutual information which should be compared for two pairs of forcings. Greater mutual information implies that the two pairs share more

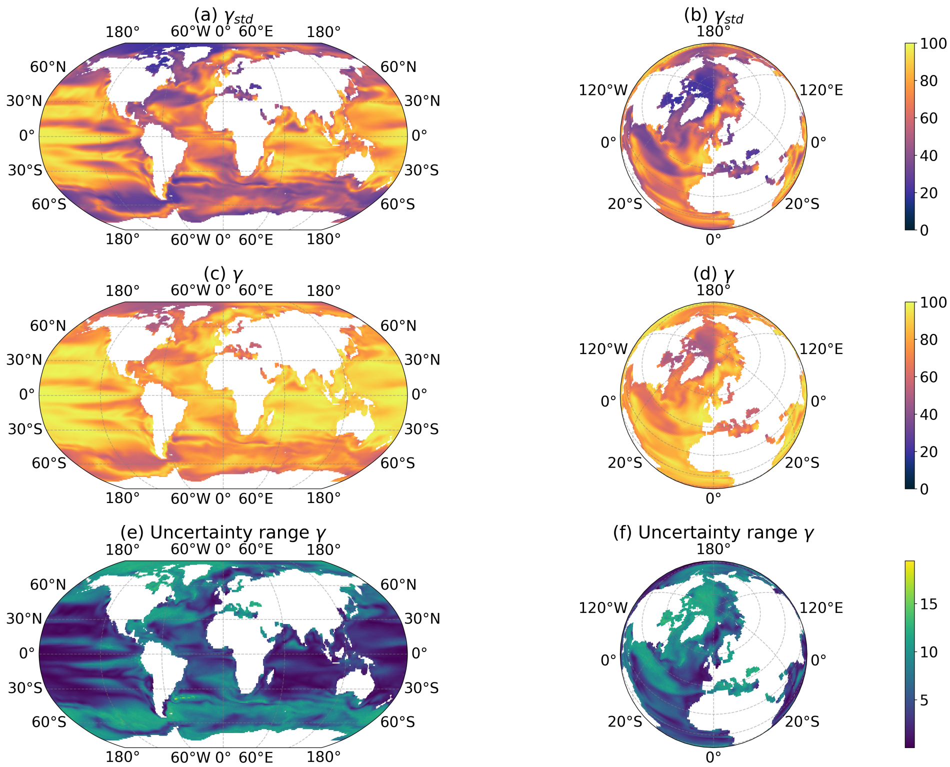

Figure 6. Intrinsic to total variability for sea surface salinity using (a, b) γ_{std} and (c, d) γ. (e, f) Uncertainty range in γ by sweeping across the bin width as explained in the text. We can see a difference in the magnitude and pattern of the intrinsic to total variability around the Arctic region. Difference in other regions such as Mediterranean sea and Pacific equator is also visible

bits of information, suggesting that one of the forcing in that pair can be replaced with the other without significantly affecting variability. For temperature dependence on wind in Figure 8, we see that only NAM and NECOFS, our two realistic forcing conditions, share much mutual information. Figure 9 shows zeroing the rivers strongly reduces the salinity variability. Futhermore, in terms of salinity impact, full rivers and mean rivers share information as do NAM and NECOFS wind forcing.

4 Discussion

Our numerical experiments performed using γ on idealized Gaussian arrays show that γ is monotonic and decreases as the linear correlation coefficient increases. Thus, apart 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 a correlation of 0 implies $\gamma = 1$, and a correlation of 1 implies $\gamma = 0$, but for a correlation of 0.5 it is closer to $\gamma = 0.9$. Only very near the correlation coefficients of 1 does γ fall below 0.5. If a roughly linear dependence on the 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

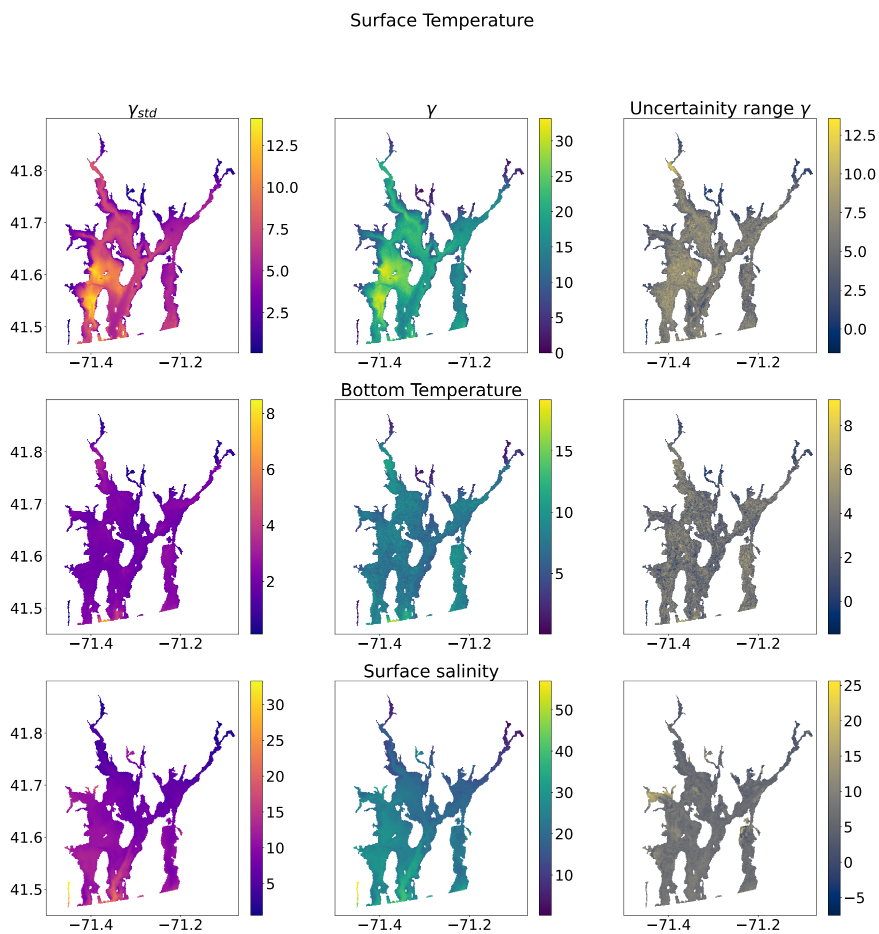

Figure 7. Metrics γ vs γ_{std} for the OSOM output. Both metrics show different contributions of intrinsic variability to total variability. γ is more uniform in the domain than γ_{std}. Right panels show the uncertainty in γ due to binning choices. The color maps for γ and γ_{std} are different to highlight their different ranges. γ_{std} for bottom temperature (not shown) has a maximum value of 5%, and the pattern is almost uniform except at the river sources where the values are on the lower side (less than 1%).

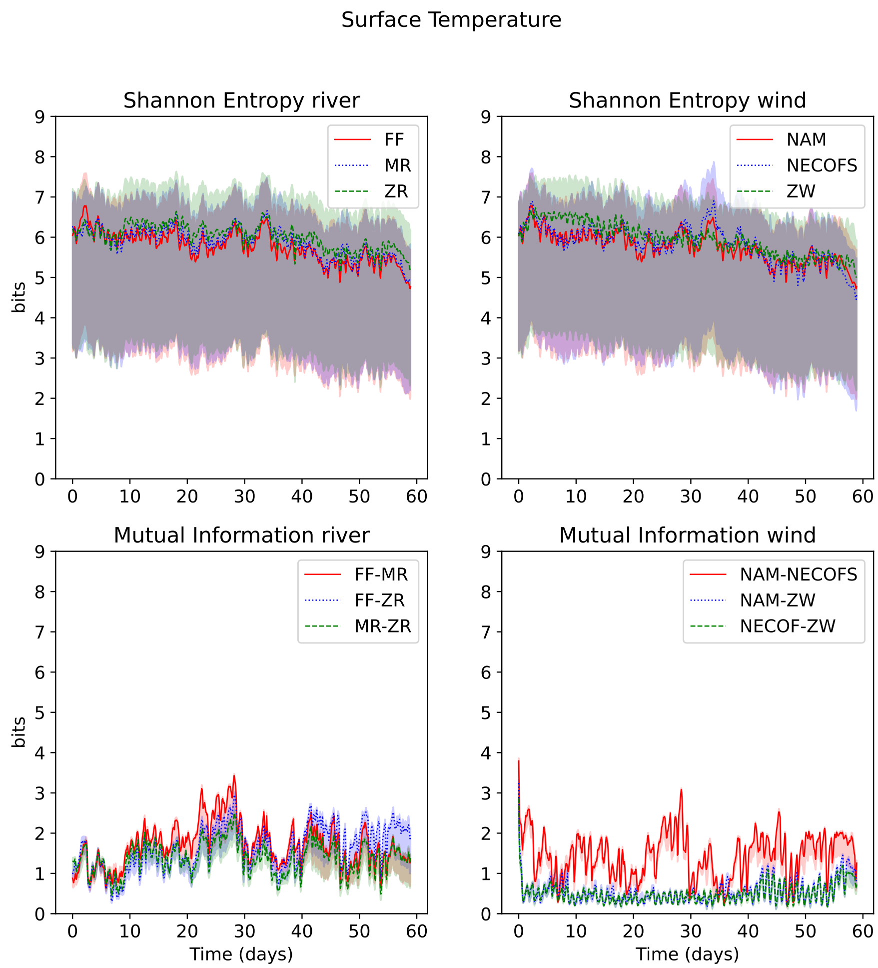

Figure 8. Shannon entropy applied to temperature and salinity. Replacing fully time-varying rivers with monthly mean river flow gives almost the same result for salinity. The same is true by replacing the wind product with a different one. Setting the river to zero affects salinity, but not temperature. Winds are important in terms of variability, but different wind products do not noticeably alter variability.

they will lose their interpretation as a ratio of bits of information entropy, and 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.

The uncertainty associated with binning is small–typically much less than the variability across the domain and the metrics are thus not overly sensitive to the binning procedure. The exploration of alternative strategies to evaluate entropies will remain a topic of future investigation and may further improve precision.

As can be seen in Figures 7, 5, and 6, information theory metrics show different patterns compared to variance metrics. Information theory metrics, especially mutual information, account for *all* non-linear shared information between the ensemble members and the mean including linear correlation, and this is one reason for

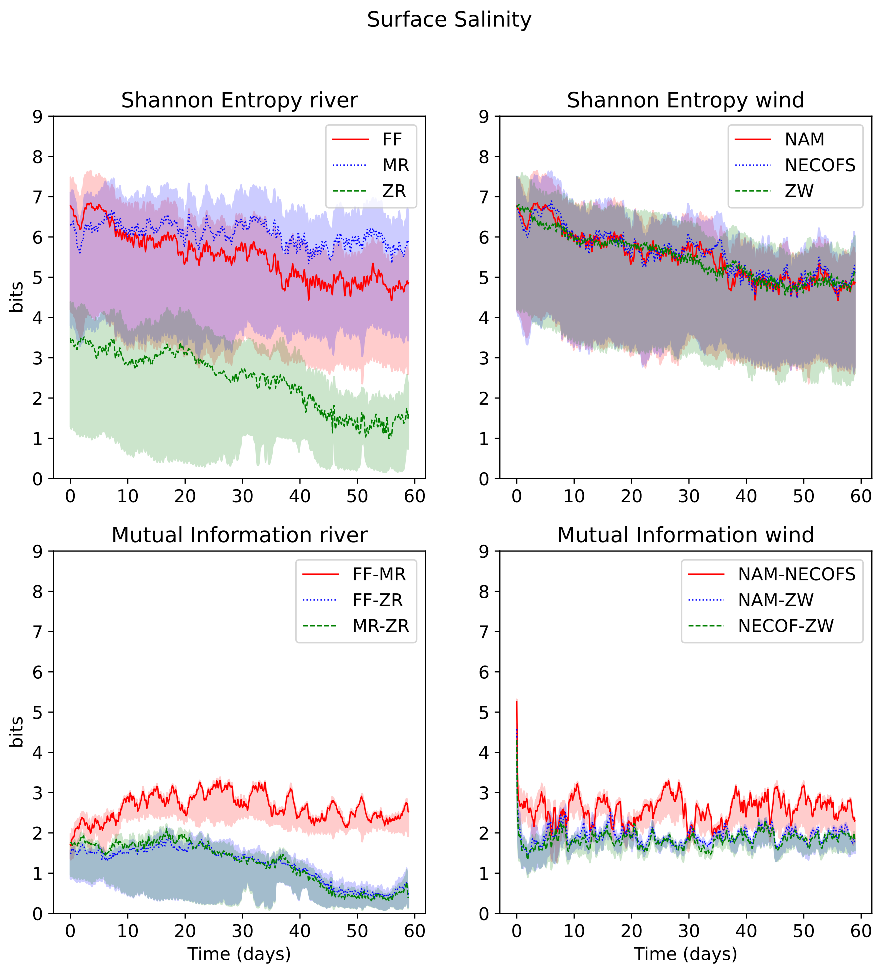

Figure 9. Mutual information applied to simulations from different forcings. Higher mutual information implies higher similarity in terms of variability. For example, NAM-NECOFS values are higher than NAM-ZW implying that NAM and NECOFS are significantly different than having no wind.

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—-justifies analyzing the data with nonparametric metrics in addition to second moment statistics.

For the regional coastal model OSOM, forcings differ in shared information and as to how they affect different variables. As might be expected, river runoff is more important for salinity than for temperature. However, for July to August, replacing rivers with the monthly mean river flow gives nearly the same result (in terms of variability) as fully time-varying rivers. Similarly, averaging the river runoff gives a similar effect for salinity compared to giving the observed river runoff in the simulations; see Figure 8. This might be due to lower river runoff during summer leading to lower variability in the flow rate hence averaging river runoff might be appropriate. We cannot conclude if there will be a similar effect in winter because the higher river runoff lead to larger variability and replacing river runoff with its mean might be unfruitful. Temperature is less sensitive to any of the forcing alterations, because it has different sources and sinks than salinity. Switching the wind product from NAM to NECOFS does not have a significant effect 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 9 shows that zero-wind (ZW) simulations are markedly different from the rest in terms of *mutual information* (i.e. they do not covary), although very similar in terms of amount of spatial variability (Shannon entropy, Figure 8), because even without winds tides, fluxes, and rivers still vary. The zero-river case tends to eliminate both variability and mutual information (ZR).

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 cannot easily tell the difference between different wind products, although something rather than zero winds should be used if the estuary needs to be forecast for the full 20 day predictability range (weather forecasts are reliable for only about 7 days in this region). Similarly, knowing that substituting the mean of the rivers for 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.

It should be noted that a major difference between coastal and global ensemble is in the way they are forced. The ocean in the coastal model has been forced by fixed atmosphere, tides, and rivers whereas the GFDL ESM2E has atmosphere which responds to the changes in the ocean. The intrinsic variability seen in the coastal ensemble is due to the ocean alone. The intrinsic variability observed in the global ensemble might not just be due to the ocean alone but might have all the possible sources present in the coupled system.

5 Conclusion

We showed usage of information theory metrics to determine contribution of intrinsic chaos and external variability to total variability in ensemble model simulations. The metric consists of Shannon entropy and mutual information and is non-parametric compared to variance. We have applied metrics on idealized Gaussian arrays, as well as realistic coastal ocean and global climate models. We conclude that:

1. The 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 information theory metric is more reliable when variability is non-Gaussian because it is based on nonparametric measures of the probability distributions and captures nonlinear correlations.
3. The new information theory metric varies monotonically with ensemble member to ensemble mean correlation, but is quantified in fractions of bits required to capture internal variability versus bits required to capture total variability.

4. The use of the information theory ratio 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. The use of Shannon entropy and mutual information can quickly focus attention on which forcing choices cause the most variability and need attention as their substitutions significantly affect the outcomes. These conclusions can be drawn regardless of the fact that the dimensions of wind, rivers, salinity, and temperature have no specified unit conversion factors.
6. In these ensemble simulations, the coastal ensemble had a much smaller intrinsic (chaotic) proportion of its total variability in comparison to the climate ensemble which had more intrinsic variability (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.
7. For the global simulations, Arctic ocean is known to be salinity dominated and temperature plays the role of a passive tracer when near the freezing point (Timmermans & Jayne, 2016), (MacKinnon et al., 2016) Information theory metric γ clearly shows high intrinsic variability in temperature at the Arctic and extrinsic forcing is low to moderate. This implies the intrinsic variability in temperature is extraneous to the dynamics of the Arctic Ocean.

Other applications of these and similar information theory metrics are likely to be revealing of new behavior and sensitivity of models.

6 Data Availability Statement

We have made the code and data available at `https://doi.org/10.5281/zenodo.7992844`. The GFDL-ESM2M Large Ensemble climate model data can be accessed from `https://www.cesm.ucar.edu/community-projects/mmlea` and has been described in (Rodgers et al., 2015) and (Deser et al., 2020).

Acknowledgments

The Rhode Island Coastal Ecology Assessment Innovation & Modeling grant (NSF 1655221) supported this work. B. Fox-Kemper 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.

References

Beardsley, R. C., & Chen, C. (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. doi: 10.3390/e7010068

Campuzano, S., De Santis, A., Pavón-Carrasco, F. J., Osete, M. L., & Qamili, E. (2018). New perspectives in the study of the earth's magnetic field and climate connection: The use of transfer entropy. *PloS One*, *13*(11), e0207270.

Carcassi, G., Aidala, C. A., & Barbour, J. (2021). Variability as a better characterization of shannon entropy. *European Journal of Physics*, *42*(4), 045102.

Correa, C. D., & Lindstrom, P. (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., & Tippett, M. K. (2007). Predictability: Recent insights from information theory. *Reviews of Geophysics*, *45*(4).

DelSole, T., & Tippett, M. K. (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., Lehner, F., Rodgers, K., Ault, T., Delworth, T., DiNezio, P., ... Ting, M. (2020). Insights from earth system model initial-condition large ensembles and future prospects [dataset]. *Nature Climate Change*, 1–10.

Eshima, N. (2020). Statistical data analysis and entropy. In (p. 13-14). Springer.

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

Franzke, C. L. E., Barbosa, S., Blender, R., Fredriksen, H.-B., Laepple, T., Lambert, F., ... Yuan, N. (2020). The structure of climate variability across scales. *Reviews of Geophysics*, *58*(2), e2019RG000657.

Freedman, D., & Diaconis, P. (1981). On the histogram as a density estimator: L 2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *57*(4), 453–476.

Gomez, B. G. (2020). *Intrinsic ocean variability modulated by the atmosphere in the gulf of mexico: an ensemble modelling study* (Unpublished doctoral dissertation). Université Grenoble Alpes [2020-....].

Hacine-Gharbi, A., Ravier, P., Harba, R., & Mohamadi, T. (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., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*(1).

Jaynes, E. T. (1962). *Information theory and statistical mechanics* (Vol. 3). 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*, 102581.

Knuth, K. H., Golera, A., Curry, C. T., Huyser, K. A., Wheeler, K. R., & Rossow, W. B. (2005). Revealing relationships among relevant climate variables with information theory. In *Eart-sun system technology conference 2005.*

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

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

Leroux, S., Penduff, T., Bessières, L., Molines, J.-M., Brankart, J.-M., Sérazin, G., ... Terray, L. (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., & North, G. R. (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., & Kleeman, R. (2005). Information transfer between dynamical system components. *Physical review letters*, *95*(24), 244101.

Liang, Y.-C., Kwon, Y.-O., Frankignoul, C., Danabasoglu, G., Yeager, S., Cherchi, A., ... Tian, T. (2020). Quantification of the arctic sea ice-driven atmospheric circulation variability in coordinated large ensemble simulations. *Geophysical Research Letters*, *47*(1), e2019GL085397.

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

MacKinnon, J. A., Nash, J. D., Alford, M. H., Lucas, A. J., Mickett, J. B., Shroyer, E. L., ... Wagner, G. L. (2016). A tale of two spicy seas. *Oceanography*, *29*(2), 50–61.

Majda, A. J., & Gershgorin, B. (2010). Quantifying uncertainty in climate change science through empirical information theory. *Proceedings of the National Academy of Sciences*, *107*(34), 14958–14963.

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

Johnson, L., Fox-Kemper, B., Li, Q., Pham, H., & Sarkar, S. (2023, May). A finite-time ensemble method for mixed layer model comparison. *Journal of Physical Oceanography*. Retrieved from https://fox-kemper.com/pubs/pdfs/JohnsonFox-Kemper23.pdf (Accepted)

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

Pothapakula, P. K., Primo, C., & Ahrens, B. (2019). Quantification of information exchange in idealized and climate system applications. *Entropy*, *21*(11), 1094.

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

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

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

Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F., & Phipps, S. J. (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. Retrieved from http://math.harvard.edu/{~}ctm/home/text/others/shannon/entropy/entropy.pdf

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

Shin, C.-S., Dirmeyer, P. A., & Huang, B. (2023). A joint approach combining correlation and mutual information to study land and ocean drivers of us droughts: Methodology. *Journal of Climate*, 1–40.

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

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

Timmermans, M.-L., & Jayne, S. R. (2016). The arctic ocean spices up. *Journal of Physical Oceanography*, *46*(4), 1277–1284.

Waldman, R., Somot, S., Herrmann, M., Sevault, F., & Isachsen, P. E. (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. doi: 10.1147/rd.41.0066

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