Internal vs Forced Variability Metrics for Geophysical Flows Using Information Theory

Aakash Sane *

School of Engineering, Brown University, Providence, Rhode Island, USA

Baylor Fox-Kemper

Department of Earth, Environment, and Planetary Sciences, Brown University, Providence, Rhode Island, USA

David S. Ullman

Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island, USA

* *Corresponding author*: Aakash Sane, aakash_sane@brown.edu

ABSTRACT

We propose a metric for measuring internal and forced variability in ensemble atmosphere, ocean, or climate models using information theory: Shannon entropy and mutual information. This metric differs from the standard ensemble-variance approaches. 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 as well as a change of units or scale. Finally, we use an example featuring a highly skewed probability distribution (Arctic sea surface temperature) to show that the new metric is robust even with a sharp nonlinear cutoff (the freezing point). We apply these two metrics to quantify internal vs forced variability in (1) idealized Gaussian data, (2) an initial condition ensemble of a realistic coastal ocean model, (3) the Community Earth System Model large ensemble. Each case illustrates the advantages of the proposed metric over variance-based metrics. Furthermore, in the coastal ocean model, the new metric is adapted to further quantify the impact of different boundary forcing choices to aid in prioritizing model improvements–i.e., comparing different choices of extrinsic forcing. The 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.

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. 2020). 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 from model's chaotic sensitivity to the initial conditions), 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.

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 (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).

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 together, which we explain in section 2. We will apply our metric on three sets of data in section 3: 1. Idealized Gaussian arrays with specified correlation 2. Ensemble output of a regional coastal model (OSOM) (Sane et al. 2020) where output is mostly non Gaussian. 3. Community Earth System Model Large ensemble (Deser et al. 2020). In section 4, we use individual components of our metric to compare simulations from differing forcings applied to OSOM.

2. 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. $\log_2(1/p_i)$ measures the information of the i^{th} outcome as proposed by Hartley (1928) and is also a measure of uncertainty (Cover 1999). The quantum $\log_2(1/p_i)$ measures the information gained by knowing that the i^{th} outcome has happened. Another interpretation for $\log_2(1/p_i)$ is that it measures reduction in uncertainty by knowing the i^{th} event has occurred, or 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 falling into the i^{th} out 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 1 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. p_i and p_j is the marginal probability of i^{th} and j^{th} outcome of x and y respectively. 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 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 by knowing y or vice versa. It 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 (for 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, for 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 non linear interactions (Correa and Lindstrom 2013) and discontinuous or intermittent visited states. We will present the effect of correlation and outliers in section 3 a.

3. Proposed metric

Let $f(t)^i$ denote ensemble output as a function of time t and i is the index of ensemble members. For ocean model simulations, the output signal $f(t)^i$ consists of trend due to external forcing given by $g(t) = \langle f(t)^i \rangle$ where angled brackets represent mean over the ensemble members and the remaining part is the intrinsic variability due to chaos given by $\eta(t)^i$. Typically, g would be estimated as the mean of the ensemble, but in some cases might be a median or a known response to forcing–another advantage of the information theory approach is that it is not required that g be the mean or that $\eta(t)^i$ have no mean value. Thus they are related as $f(t)^i = g(t) + \eta(t)^i$, where $g(t)$ represents the forced component of the total variability and the model spread $\eta(t)^i$ can be attributed to the internal chaos (or model distinctions in multi-model ensembles) (e.g., Waldman et al. 2018; Llovel et al. 2018).

Usually we are limited in the number of ensemble members due to computational costs so we perform a *jugaad* with f and g in order to use all the members at once to sample widely and obtain an accurate probability distribution. Let f be a single array obtained by combining (concatenating) all members $f(t)^i, \forall\, i \in N$, where N is the total number of ensemble members. Let g be the array obtained after copying the model mean $g(t)$ N times. Hence if there are M data points in each member, f and g are of length $N \times M$ each. This makes them of equal length and the metric can be applied to the whole ensemble output. Using f and g, we propose the following metric γ

$$\gamma = 1 - \frac{I(f;g)}{H(f)}, \tag{3}$$

$H(f)$ is the Shannon entropy of f and $I(f;g)$ is mutual information between f and g. Note that g is due to extrinsic factors and f is obtained after internal chaos adds to 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 culmination 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 $H(f) - I(f;g)$ determines 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 capture the entropy in the noisy binary tree, to each of the $2^{I(f;g)}$ microstate 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 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 $f \to g$, then $I(f;g) \to H(f) = H(g)$ from (2). This makes $\gamma = 0$ when there is no intrinsic chaos. When intrinsic chaos fully dominates the ensemble output,

[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.

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 analogy for a climate system component is a noisy channel as given in Leung and North (1990), where the governing equations of ocean (atmosphere) modeling are represented by a channel. 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.

a. Idealized Gaussian Arrays

We test our metric, γ, from (3) 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. To illustrate, we compare our metric γ with the

[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.

traditional metric of intrinsic versus extrinsic variability γ_{std}:

$$\gamma_{std} = \frac{\sigma_i}{\left(\sigma_i^2 + \sigma_g^2\right)^{1/2}}, \tag{4}$$

where subscript std stands for standard deviation, σ_i is the temporal mean of standard deviation of model spread, and σ_g is the temporal standard deviation of the model mean (not a function of time). 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 each of 10 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. As the individual members have same correlation coefficient between them, every member also has a constant correlation (different from correlation between any two ensemble members) with the mean member. We apply γ and γ_{std} on this synthetic ensemble by varying the prescribed correlation coefficient from 0 to 1. Figure 2 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 2: 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 2 shows that γ is insensitive to outliers while γ_{std} is not. γ is not sensitive because 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). Increasing the number of ensemble members does not change the result qualitatively for both the experiments.

b. 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. 2020). 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. (2020) gives more details about the model.

Using OSOM, an ensemble of simulations have been performed using perturbed initial conditions for the months July - August of 2006. The data during the initial predictability window (20 days) has been ignored and the rest 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. 2020). We examine whether the modeled temperature and salinity at each grid point follow normal distributions by evaluating the skewness and kurtosis of the model mean at each grid point. Figure 3 shows skewness and kurtosis for sea surface salinity and temperature as well 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, (4) 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 (3) is suitable for such data as it takes into account higher moments and does not rely on Gaussian distributions.

Figure 4 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 from topography intermittently. 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 2.

c. 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 CM3 model (Deser et al. 2020). Variability of sea surface temperature (Figures 5) as well as sea surface salinity (6) were estimated using both γ and γ_{std} (upper left and upper right). The skewness and excess kurtosis of the model 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.

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 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, and 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 2.

4. Quantifying extrinsic factors

In section 3 we proposed our metric and showcased its applications. 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 changes in the extrinsic forcings may be required for coastal applications where systems vary predominantly due to external forcings (note that intrinsic variability is less than half by both metrics in Figure 4). 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. 2020). 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.

We show results in Figures 7 and 8. Entropy has been plotted with respect to time. In Figure 7, Shannon entropy is plotted for spatial quantities. For example, for surface salinity, all the surface values have been considered to find Shannon entropy. Figure 8 displays mutual information. Mutual information has been calculated in the following way: for surface temperature comparison between FF and MR, surface grid point values for FF simulation were ordered and put in a single row array. We compared this array with a similar array made of surface grid point values of MR surface temperature. We consistently followed the same method for all the forcings and variables. It is user's choice to choose the type of domain, here we have chosen the same domain of OSOM as shown in figure 4. 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.

5. 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 2. Approximately, the traditional metric 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 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 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.

As can be seen in Figures 4, 5, and 6, information theory metrics show different patterns as compared to variance. 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 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, 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 7. 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 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 8 shows that zero wind (ZW) simulations are markedly different than the rest in terms of *mutual information* (i.e., they do not covary), although very similar in terms of amount of variability (Shannon entropy, Figure 7), 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 is forecast out to the full 20 day predictability range (weather forecasts only good out 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.

6. 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 and NSF 1350795. 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, doi: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)**.

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.

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.

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.

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, 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**.

Sane, A., B. Fox-Kemper, D. Ullman, C. Kincaid, and L. Rothstein, 2020: Consistent predictability of the ocean state ocean model (osom) using information theory and flushing timescales. *Earth and Space Science Open Archive*, 34, doi:10.1002/essoar.10504826.1, URL https://doi.org/10.1002/essoar.10504826.1.

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.

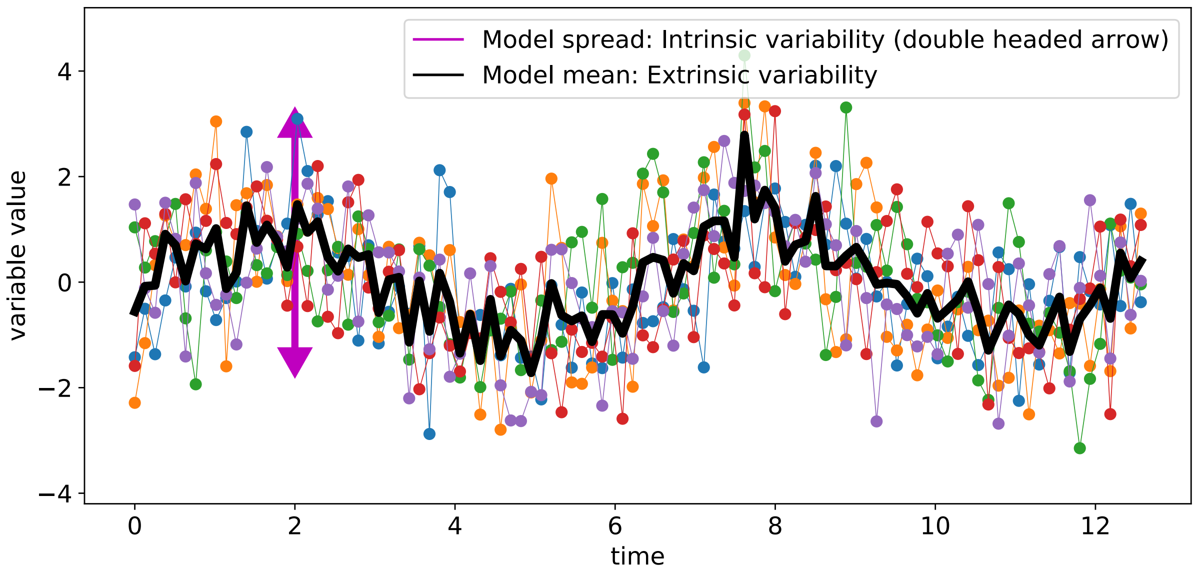
Yettella, V., J. B. Weiss, J. E. Kay, and A. G. Pendergrass, 2018: An ensemble covariance framework for quantifying forced climate variability and its time of emergence. *Journal of Climate*, **31 (10)**, 4117–4133.

LIST OF TABLES

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. (2020). MR: mean rives; ZR: zero rivers; ZW: zero wind.

LIST OF FIGURES



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

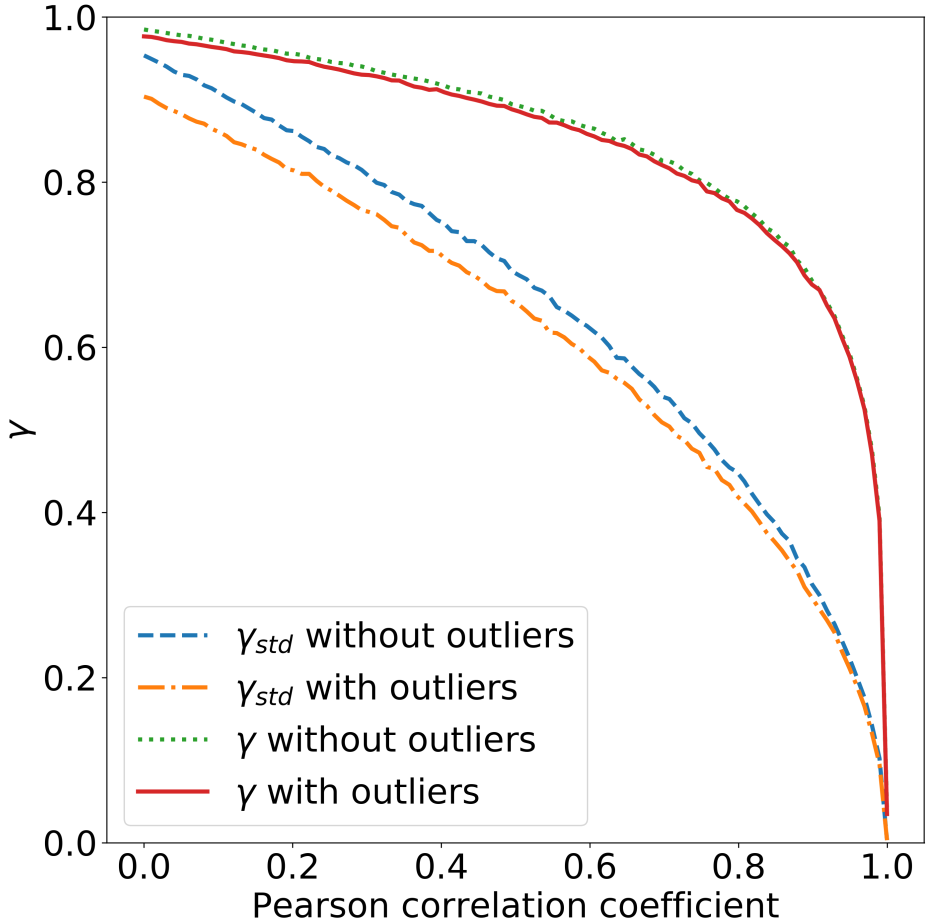
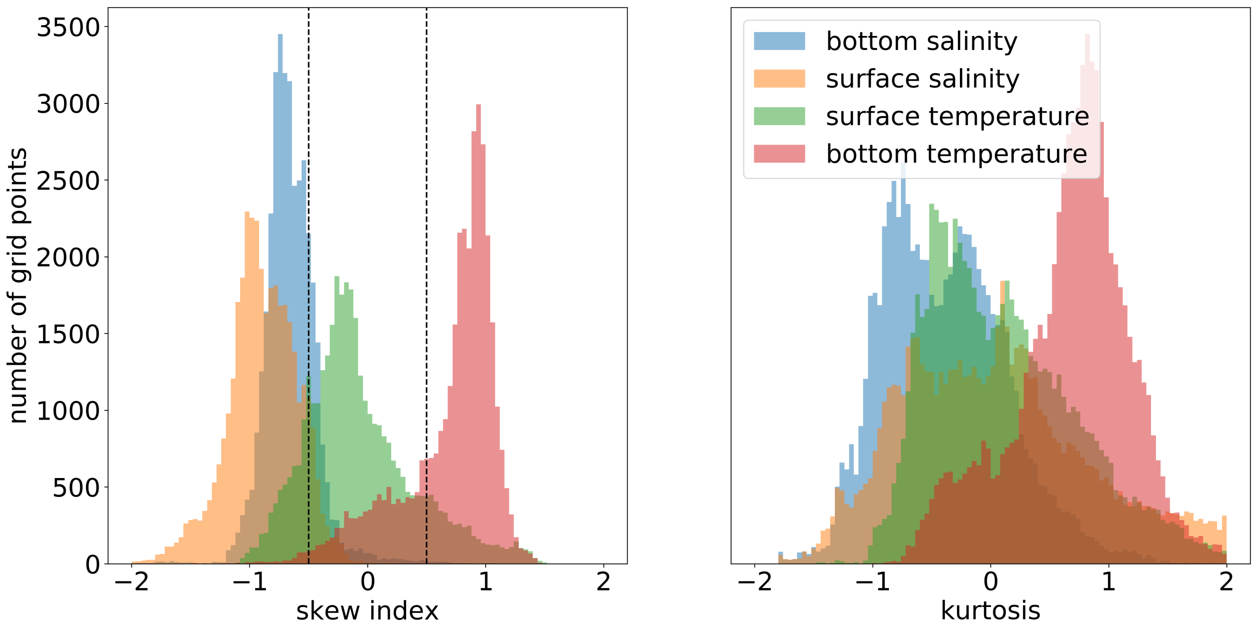


FIG. 2. Information theory metric of intrinsic vs. extrinsic variability γ as a function of correlation coefficient in idealized Gaussian correlated arrays. The horizontal axis is the correlation coefficient between mean member and ensemble members. The vertical axis shows the information theory metric γ from (3) and the traditional metric γ_{std} from (4). 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.



524 FIG. 3. Grid point wise kurtosis for OSOM output. Kurtosis is not closer to zero within (-0.5, 0.5) suggesting
525 the data distribution is non Gaussian.

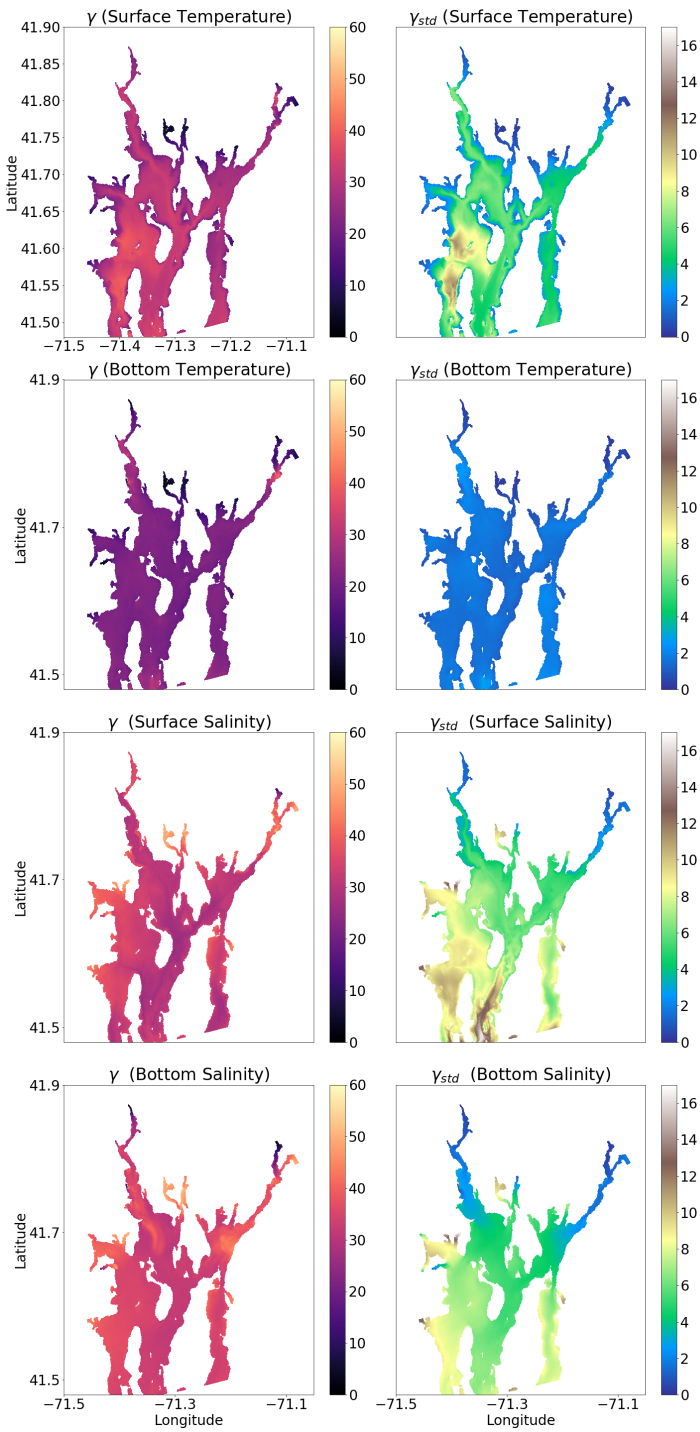


FIG. 4. Metrics γ vs γ_{std} for OSOM output. Both metrics show different contribution of intrinsic variability to total variability. γ is more uniform throughout the domain than γ_{std}. Colormaps for γ and γ_{std} are different to highlight the different ranges each of them have. γ_{std} for bottom 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%).

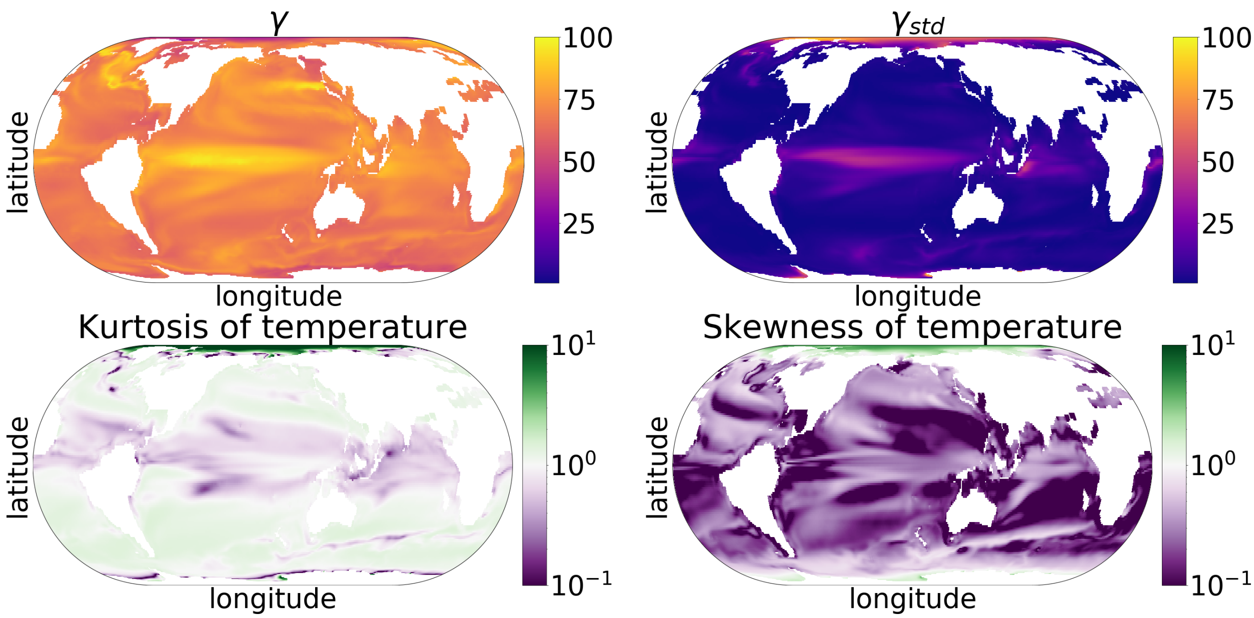


FIG. 5. Top: Intrinsic to total variability percentage for sea surface temperature. Bottom: Excess kurtosis and skewness of the model mean of temperature at each grid point. Values closer to zero (within 0.5 of zero, purple shades) are considered approximately Gaussian. The deviation of model mean away 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.

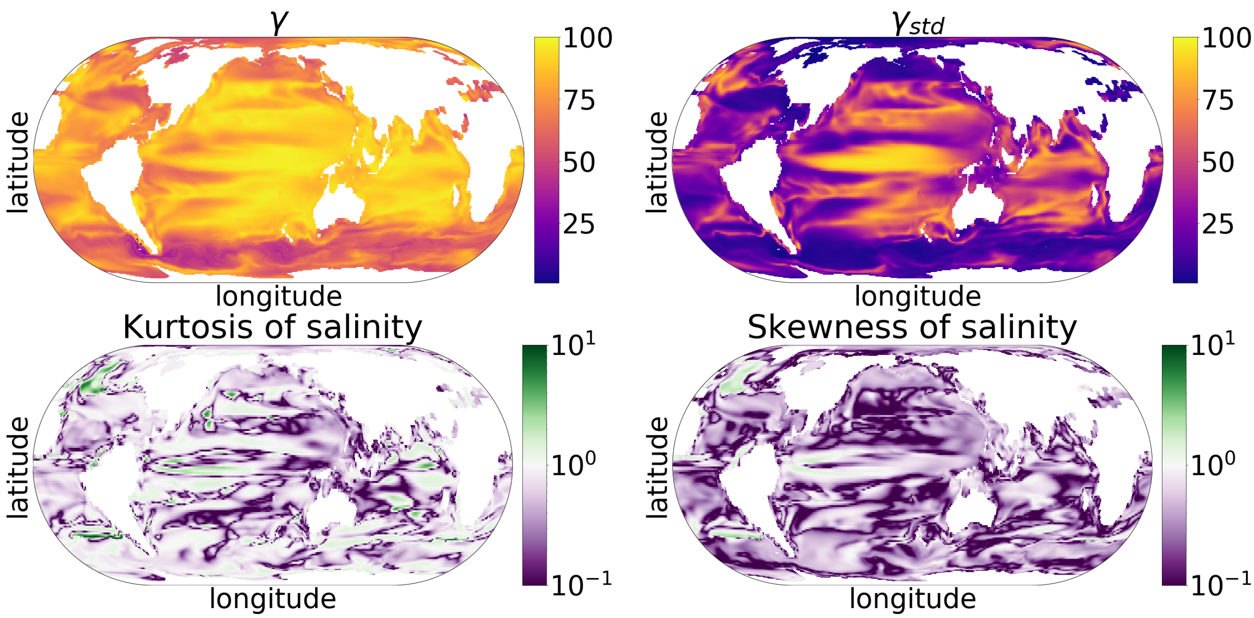
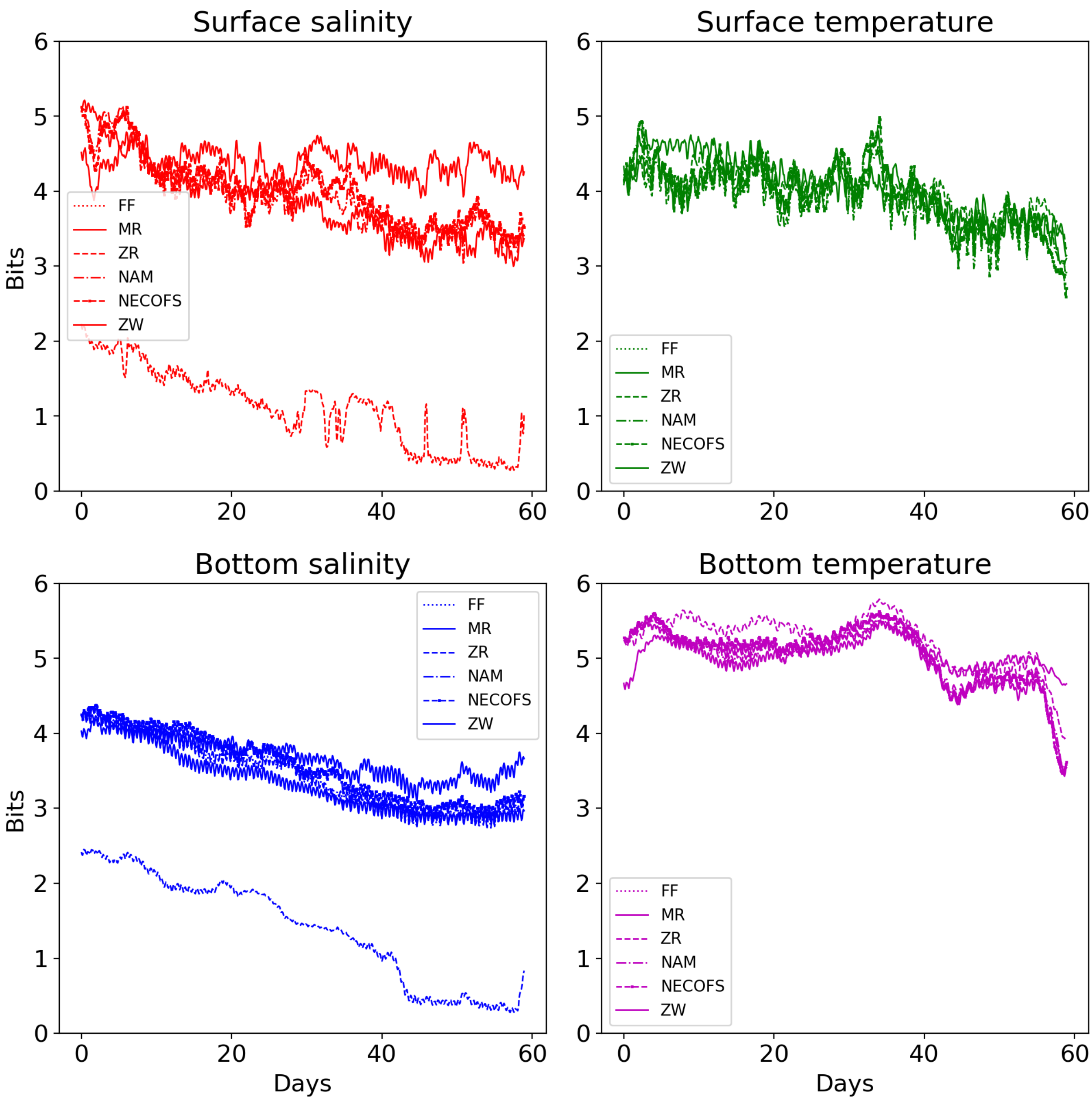
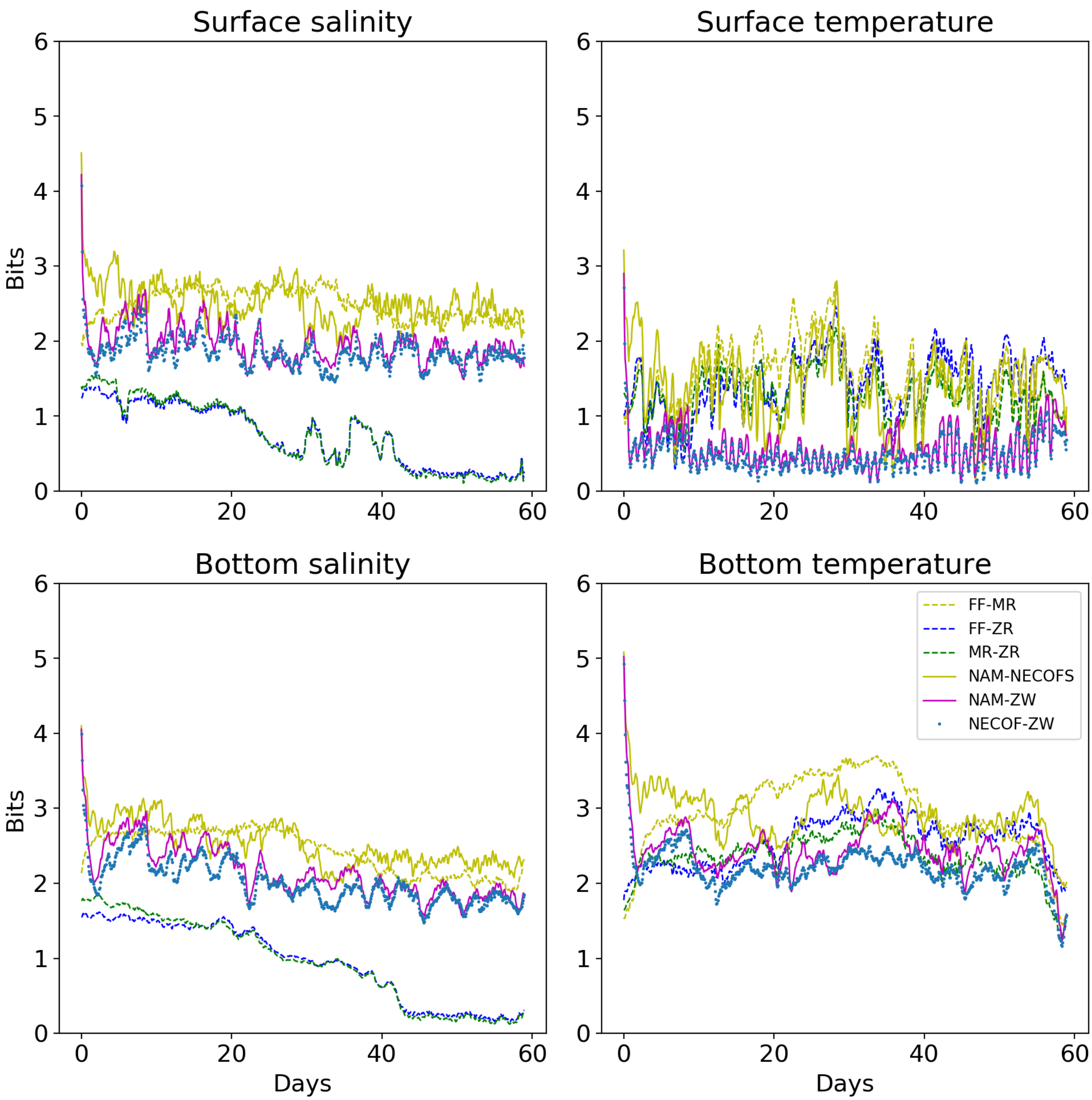


Fig. 6. Top: Intrinsic to total variability percentage for sea surface salinity. Bottom: Kurtosis and skewness of the model mean of salinity at each grid point. Values closer to zero (within 0.5 of zero, purple shades) are considered approximately Gaussian.



538 FIG. 7. 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.



542 Fig. 8. Mutual information applied to simulations from different forcings. Higher mutual information implies
543 higher similarity in terms of variability. For example NAM-NECOFS values are higher than NAM-ZW implying
544 that NAM and NECOFS are significantly different than having no wind.