

## QUANTIFYING UNCERTAINTY FOR NON-GAUSSIAN ENSEMBLES IN COMPLEX SYSTEMS\*

RAFAIL V. ABRAMOV<sup>†</sup> AND ANDREW J. MAJDA<sup>†</sup>

**Abstract.** Many situations in complex systems require quantitative estimates of the lack of information in one probability distribution relative to another. In short-term climate and weather prediction, examples of these issues might involve a lack of information in the historical climate record compared with an ensemble prediction, or a lack of information in a particular Gaussian ensemble prediction strategy involving the first and second moments compared with the non-Gaussian ensemble itself. The relative entropy is a natural way to quantify this information. Here a recently developed mathematical theory for quantifying this lack of information is converted into a practical algorithmic tool. The theory involves explicit estimators obtained through convex optimization, principal predictability components, a signal/dispersion decomposition, etc. An explicit computationally feasible family of estimators is developed here for estimating the relative entropy over a large dimensional family of variables through a simple hierarchical strategy. Many facets of this computational strategy for estimating uncertainty are applied here for ensemble predictions for two “toy” climate models developed recently: the Galerkin truncation of the Burgers–Hopf equation and the Lorenz ’96 model.

**Key words.** predictability, relative entropy, ensemble predictions

**AMS subject classifications.** 82C, 65C, 86A10

**DOI.** 10.1137/S1064827503426310

**1. Introduction.** Complex systems with many spatial degrees of freedom arise in environmental science in diverse contexts such as atmosphere/ocean general circulation models (GCMs) for climate or weather prediction [7], pollution models, and models for the spread of hazardous biological, chemical, or nuclear plumes. These nonlinear models are intrinsically chaotic over many time scales with sensitive dependence on initial conditions. In this paper, such models are represented discretely as a large system of ODEs for a vector  $\vec{X} \in \mathbb{R}^N$  given by

$$(1.1) \quad \frac{d\vec{X}}{dt} = \vec{F}(\vec{X}, t),$$

with  $N \gg 1$ .

Given both the uncertainty in a deterministic initial condition,  $\vec{X}_0$ , as well as the intrinsic chaos in solutions of (1.1), it is natural to consider an ensemble of initial data characterized by a probability density,  $p_0(\vec{X})$ , satisfying  $p_0 \geq 0$ ,  $\int p_0 = 1$  and with mean given by  $\vec{X}_0$ , i.e.,

$$(1.2) \quad \int_{\mathbb{R}^N} \vec{X} p_0(\vec{X}) d\vec{X} = \vec{X}_0.$$

The idea is to utilize the ensemble of solutions of (1.1) drawn from the initial data  $p_0(\vec{X})$  to quantify the uncertainty and measure the confidence interval and predictive

---

\*Received by the editors April 23, 2003; accepted for publication (in revised form) March 25, 2004; published electronically December 22, 2004. The work of the second author was partially supported by National Science Foundation grant DMS-9972865, Office of Naval Research grant N00014-96-1-0043, and National Science Foundation CMG Research grant DMS-0222133. The work of the first author was supported as a postdoctoral fellow through these research grants.

<http://www.siam.org/journals/sisc/26-2/42631.html>

<sup>†</sup>Courant Institute of Mathematical Sciences and Center for Atmosphere/Ocean Science, New York University, New York, NY 10012 (abramov@cims.nyu.edu, jonjon@cims.nyu.edu).

power of the deterministic solution beginning at  $\vec{X}_0$ . Theoretically, this is straightforward because it is easy to establish that the initial probability density,  $p_0(\vec{X})$ , evolves to a new probability density,  $p_t(\vec{X})$ , at later times which satisfies the Liouville equation,

$$(1.3) \quad \begin{aligned} \frac{\partial}{\partial t} p_t(\vec{X}) + \operatorname{div}_{\vec{X}}(\vec{F} p_t) &= 0, \\ p_t(\vec{X}) \Big|_{t=0} &= p_0(\vec{X}). \end{aligned}$$

The Liouville equation is a linear PDE in a very large spatial dimension,  $\mathbb{R}^N$ , with  $N \gg 1$ , and is impractical to solve directly. In practice, instead a finite ensemble of individual solutions of (1.1),  $\{\vec{X}_t^r, 1 \leq r \leq R\}$ , is constructed by sampling the initial distribution,  $p_0(\vec{X})$ . Then  $p_t(\vec{X})$  is approximated by the empirical probability distribution for ensemble prediction

$$(1.4) \quad p_t^E = \frac{1}{R} \sum_{r=1}^R \delta_{\vec{X}_t^r}(\vec{X}),$$

where  $\delta_{\vec{X}_0}(\vec{X})$  is the Dirac delta measure at  $\vec{X}_0$ . The probability density in (1.4) is an explicit solution of the Liouville equation in (1.3).

The empirical ensemble prediction in (1.4) is the central probability measure of interest in this paper. How can one quantify and estimate the uncertainty in such a prediction? In the remainder of this introduction, some of the important scientific issues associated with empirical ensemble prediction are discussed, as is the potential use of information theory [6] in quantifying many aspects of uncertainty. The rest of the paper focuses on demonstrating the fashion in which a recently developed mathematical framework [16] can be converted into a practical tool for assessing uncertainty in complex systems.

The first practical issue to confront is that the empirical ensemble size,  $R$ , in (1.4) is often very small in environmental science, on the order of  $R \leq 50$  for weather prediction and short-term climate prediction in GCMs [24, 21], while, for example,  $R = 500$  for intermediate models for predicting El Niño [10, 11]. This is due to the computational cost in these extremely complex systems in an operational setting requiring real time prediction. Furthermore, the idealized ensemble distribution,  $p_0(\vec{X})$ , is also unknown in practice. Thus, in this paper we intentionally focus on the empirical distribution,  $p_t^E$ , rather than perfect predictability scenarios. In the perfect predictability scenario, the initial distribution,  $p_0(\vec{X})$  from (1.3), is known exactly and the approximation by an ensemble  $R \gg 1$  is the main emphasis. These studies in idealized systems are very useful in determining the features governing predictability in a given system [12, 9].

What are available to quantify the uncertainty from the empirical ensemble prediction  $p_t^E$  and readily evaluated are some of the moments,

$$(1.5) \quad \int_{\mathbb{R}^N} (\vec{X} - \bar{\vec{X}})^\alpha p_t^E(\vec{X}) d\vec{X} = \frac{1}{R} \sum_{r=1}^R (\vec{X}_t^r - \bar{\vec{X}}^r)^\alpha, \quad 0 \leq |\alpha| \leq 2L,$$

with  $\vec{X}^\alpha = X_1^{\alpha_1} \dots X_N^{\alpha_N}$  [7],  $|\alpha| = \sum \alpha_i$ . Of course, in practice only the low order moments, where  $L = 2$  or  $4$ , are utilized for a judicious restricted set of variables

$$(1.6) \quad \vec{X}_i, \quad 1 \leq i \leq M.$$

Some of the main important issues in quantifying the uncertainty of an ensemble prediction are listed below.

- (A) How much of a lack of information is contained in a given prediction strategy, for example, measuring only the second moments in (1.5), compared with  $p_t^E$ ?
- (B) When does the ensemble prediction distribution exhibit bimodality in some variable so that there are at least two different scenarios of significant change predicted in a given variable for times of interest? How can this be quantified in a computationally tractable fashion?
- (C) In a given dynamical system in (1.1), which subsets of variables  $X_1 \dots X_M$  are more predictable than the others? How can this be quantified?
- (D) For long-term climate prediction, when is an ensemble prediction useful at all beyond the historical climate record? How can one estimate the lack of information in the climate record beyond the ensemble prediction? What features control the variation of this predictive utility in a given dynamical system?
- (E) What features characterize the rare ensemble predictions with more than typical information beyond the climate record?

All of the above issues involve a lack of information in one probability distribution,  $\Pi(\vec{X})$ , compared with another probability distribution,  $p(\vec{X})$ , over some subset of variables,  $X_1 \dots X_M$ . In information theory, this lack of information content is quantified by the relative entropy,

$$(1.7) \quad P(p, \Pi) = \int_{\mathbb{R}^M} p \ln \left( \frac{p}{\Pi} \right)$$

(see [6]). Recently, in an important paper, Kleeman [12] has suggested the use of relative entropy to address the issue in (D) for long-term climate prediction. In this case,  $\Pi$  is the historical record climate distribution on a given set of variables, while  $p = p_t^E$  is the empirical ensemble prediction, so  $P(p_t^E, \Pi)$  should be estimated. The main issues in (D) were raised earlier by Anderson and Stern [4]. In the situation in (A), monitoring only the first and second moments of the prediction as is usually done in practice [24, 7] leads to the Gaussian prediction,  $p_G$ ; so in this case  $P(p_t^E, p_G)$  should be estimated. Recently, Roulston and Smith [22] have suggested a similar use of information theory as a “score” or predictability measure for ensemble forecasts. Below we show how to develop rigorous computationally feasible tests for bimodality to address (B) through non-Gaussian estimators for the empirical prediction ensemble. With regard to (C), for the special case when both  $p$  and  $\Pi$  are Gaussian distributions, Schneider and Griffies [23] have developed the idea of principal predictability components utilizing the entropy difference; this interesting concept has been generalized by Majda, Kleeman, and Cai [16] to one typical practical situation where  $\Pi$  is Gaussian while  $p$  is non-Gaussian with (1.7) as the more precise measure of lack of information in  $\Pi$ . Finally, to address the important issue in (E) for long-term prediction as well as the key factors that determine variability in predictive utility, Kleeman [12] has introduced and applied the concept of the signal/dispersion decomposition for the special case when both  $\Pi$  and  $p$  are Gaussian distributions. Recently, Kleeman, Majda, and Timofeyev [8] have determined the controlling factors of this facet of predictability in a simple model with statistical features of the atmosphere in a Gaussian setting. This decomposition for suitable non-Gaussian measures,  $p$  and  $\Pi$ , has been developed by Majda, Kleeman, and Cai in [16], and section 2.3 below contains a new, more

precise decomposition into a signal, dispersion, and cross-term. Both the signal and cross-term are directly and cheaply evaluated from the moments of  $p$ . Thus, they are readily available in determining variations in utility in a given dynamical system.

With the background discussion presented above, once one adopts the relative entropy in (1.7) as the theoretical tool for addressing the important issues in (A)–(E), the following practical issues arise.

- (F) How can one numerically compute or approximate the relative entropy  $P(p, \Pi)$  for a complex system with many degrees of freedom and a subset of variables with  $M \gg 1$ ?
- (G) Are there computationally feasible and mathematically rigorous strategies to estimate  $P(p, \Pi)$  from below, i.e.,  $P(p, \Pi) \geq P(p^*, \Pi)$ , where  $P(p^*, \Pi)$  can be evaluated either analytically or by a rapid numerical procedure?

The objective of this paper is to illustrate computationally feasible strategies to address the practical issues in (F) and (G) so that the rigorous mathematical theory developed recently [16] can be applied to address the important practical questions for quantifying uncertainty in (A)–(E). Section 2 contains an overview of these computational strategies for estimating (1.7) as well as a brief summary of the key theoretical results [16]. Two different toy models for statistical features of the atmosphere [15, 14, 17, 18] and the statistical properties of the climate distribution are introduced in section 3 for an interesting range of parameters. These toy models are used to illustrate, in a simple context, how the tools discussed in this paper can be applied in more practical situations.

**2. A rigorous computational framework for estimating predictive information content.** Here we develop a mathematically rigorous and computationally feasible framework for estimating the relative entropy in (1.7) which we also call the predictive utility or utility. This framework addresses the key computational issues listed in (F) and (G) in section 1 in a straightforward fashion. The material presented below summarizes the mathematical theory developed recently by Majda, Kleeman, and Cai [16] in a convenient fashion for applications; the interested reader can consult that paper for the rigorous proofs. In section 4 of this paper, we address all of the issues listed in (A)–(E) in section 1 for quantifying uncertainty by applying this computational framework to “toy” climate models.

The relative entropy in (1.7) for the probability measures  $p$  and  $\Pi$  has several attractive features listed below.

- (A)  $P(p, \Pi)$  measures the average lack of information in  $\Pi$  compared with  $p$ ;
- (B)  $P(p, \Pi) > 0$  unless  $p \equiv \Pi$ , with  $P(\Pi, \Pi) \equiv 0$ ;
- (C)  $P(p, \Pi)$  remains unchanged through an arbitrary nonlinear change of variables.

Often a Gaussian measure is utilized as an estimate of an ensemble prediction as in the application in (A) in section 1. Somewhat surprisingly, in many applications in atmospheric science [25] as well as in simplified dynamical models for geophysical flows [19], the probability distribution for the climate of a suitable collection of variables is essentially Gaussian. Thus, for all the diverse applications for estimating predictive utility listed in (A)–(E) in section 1, it is interesting to consider the important special case of studying  $P(p, \Pi)$  when  $\Pi$  is a Gaussian measure. Recall that for such a Gaussian distribution,

$$(2.1) \quad \Pi(X_1, \dots, X_M) = (2\pi)^{-M/2} (\det C)^{-1/2} \exp\left(-\frac{1}{2}((\vec{X} - \vec{X}), C^{-1}(\vec{X} - \vec{X}))\right),$$

where  $\vec{\bar{X}} \in \mathbb{R}^M$  is the mean,

$$(2.2) \quad \int X_j \Pi = \bar{X}_j, \quad 1 \leq j \leq M,$$

and  $\mathcal{C} = (C_{ij})$  is the positive definite,  $\mathcal{C} > 0$ , symmetric,  $M \times M$  covariance matrix,

$$(2.3) \quad \int (X_i - \bar{X}_i)(X_j - \bar{X}_j) \Pi = C_{ij}.$$

Since the predictive utility in (1.7) is invariant under changes of variables, for a Gaussian measure  $\Pi$ , defined above in (2.1)–(2.3), consider the change of variables

$$(2.4) \quad \vec{X} = \vec{\bar{X}} + \mathcal{C}^{1/2} \vec{Y}.$$

In (2.4),  $\mathcal{C}^{1/2}$  is the square root of the positive definite matrix  $\mathcal{C}$ ; the matrix  $\mathcal{C}^{1/2}$  is also symmetric and positive definite, commutes with  $\mathcal{C}$ , and satisfies  $\mathcal{C}^{1/2} \mathcal{C}^{-1} \mathcal{C}^{1/2} = I$ . Since  $\mathcal{C}$  is positive definite symmetric, there is an orthonormal basis,  $\{\vec{e}_{i=1}^M\}$ , called empirical orthogonal functions (EOFs) in the atmosphere/ocean community and corresponding positive eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > 0$  so that the covariance matrix  $\mathcal{C}$  is diagonalized, i.e.,

$$(2.5) \quad \mathcal{C} \vec{f} = \sum_i \lambda_i \vec{e}_i(\vec{f}, \vec{e}_i).$$

The eigenvector  $\vec{e}_1$  is called the first EOF,  $\vec{e}_2$  the second EOF, and so forth. With (2.5), the matrix  $\mathcal{C}^{1/2}$  is defined by

$$(2.6) \quad \mathcal{C}^{1/2} \vec{f} = \sum_i \lambda_i^{1/2} \vec{e}_i(\vec{f}, \vec{e}_i).$$

From (2.1) it follows that in the coordinate system defined by (2.4), the Gaussian distribution  $\Pi$  has the simplified form

$$(2.7) \quad \Pi(\vec{Y}) = (2\pi)^{-M/2} e^{-|\vec{Y}|^2/2}.$$

With (2.4), given  $p(\vec{X})$ , one gets the transformed distribution,

$$(2.8) \quad p_{\Phi}(\vec{Y}) = p(\vec{X} + \mathcal{C}^{1/2} \vec{Y}) \det(\mathcal{C}^{1/2}).$$

Thus, without loss of generality, we assume below that the Gaussian measure  $\Pi(\vec{X})$  has the canonical form

$$(2.9) \quad \Pi(\vec{X}) = (2\pi)^{-M/2} \exp\left(-\frac{1}{2} |\vec{X}|^2\right).$$

**2.1. Explicit estimators for predictive utility.** For simplicity in exposition, here we assume one-dimensional probability distributions to avoid cumbersome notation. All of the results presented below extend immediately to multivariable distributions with more complex notation. We assume a given prior distribution  $\Pi(\lambda)$ , for a scalar variable which is centered so that it has zero mean,

$$(2.10) \quad \int \lambda \Pi(\lambda) d\lambda = 0.$$

We consider the Probability Distribution Function (PDF) from an ensemble prediction,  $p(\lambda)$ . As discussed in the introduction, it is a reasonable strategy in practice to measure some moments of the ensemble prediction distribution,  $p(\lambda)$ ,

$$(2.11a) \quad \bar{\lambda} = M_1 = \int \lambda p(\lambda) d\lambda,$$

$$(2.11b) \quad M_j = \int (\lambda - \bar{\lambda})^j p(\lambda) d\lambda, \quad 2 \leq j \leq 2L.$$

For retaining statistical significance in an ensemble prediction, one usually measures only the first two or four moments in practice so that  $L = 1$  or  $L = 2$  in (2.11b).

With the moment information from the ensemble prediction in (2.11), it is natural to define  $p^*$  as the probability measure with the least bias which retains the information in (2.11). To do this, we introduce the set of probability measures which satisfy the  $2L$  constraints in (2.11),  $PM_{2L}$ . Since  $-p \ln \left( \frac{p}{\Pi} \right)$  is a concave function of  $p$ , we define  $p^*$  via the maximum entropy principle [13]

$$(2.12) \quad \mathcal{S}(p^*, \Pi) = \max_{p \in PM_{2L}} \mathcal{S}(p, \Pi),$$

where  $\mathcal{S}(p, \Pi) = -P(p, \Pi)$ . The usual Lagrange multiplier calculation yields the explicit formula for  $p^*$ ,

$$(2.13) \quad -\ln \left( \frac{p^*}{\Pi} \right) = \sum_{\substack{j=0 \\ j \neq 1}}^{2L} \alpha_j (\lambda - \bar{\lambda})^j + \alpha_1 \lambda,$$

where  $\alpha_j, 0 \leq j \leq 2L$  are the Lagrange multipliers for the  $2L$  constraints.

With (2.11) and (2.12),  $p^*$  provides a rigorous predictability estimator for the ensemble prediction

$$(2.14) \quad P(p, \Pi) \geq P(p^*, \Pi).$$

Furthermore, with (2.11) and (2.13), there is an explicit formula for the predictive utility,

$$(2.15) \quad P(p^*, \Pi) = - \sum_{j=2}^{2L} \int p^* \alpha_j (\lambda - \bar{\lambda})^j - (\alpha_0 + \bar{\lambda} \alpha_1).$$

Next, assume that we have ensemble predictability estimators  $p_{2L_1}^*, p_{2L_2}^*$ , where  $p_{2L_1}^*$  involves  $2L_1$  constraints and  $p_{2L_2}^*$  involves  $2L_2$  constraints with  $L_1 < L_2$ . By the definition in (2.12) we have

$$(2.16) \quad \mathcal{S}(p_{2L_1}^*, \Pi) \geq \mathcal{S}(p_{2L_2}^*, \Pi)$$

which yields the *Predictability Estimator Principle*,

$$(2.17) \quad P(p, \Pi) \geq P(p_{2L_2}^*, \Pi) \geq P(p_{2L_1}^*, \Pi).$$

One important advantage of the construction outlined above is that it leads to both an explicit probability distribution  $p^*(\lambda)$  and the explicit estimates for the predictive

utility in (2.14), (2.15) even for small ensembles,  $R \cong 50$ . Extracting meaningful predictive information from direct sampling in this context is extremely difficult with the small sample size. This explicit distribution also allows for both quantitative and direct visual identification of bimodal behavior in ensemble prediction from (B) in section 1.

Another important practical feature of the constrained optimization problem in (2.12) is that it can be recast as an unconstrained convex optimization of the Lagrangian function in the dual space of Lagrange multipliers (see the *SIAM Review* article [26]). This dual problem is solved readily by modified Newton methods, provided the dimension of the constraint set is not too large. In sections 3 and 4 of this paper we apply such an algorithm in low dimensions with at most five moment constraints,  $L = 2$ , and  $M = 1$  or  $2$ . For a fixed number of moment constraints,  $2L$ , the dimension of the space of Lagrange multipliers grows like  $\mathcal{O}(M^{2L})$ , so alternative computationally feasible strategies are needed for large  $M$  and are described below in section 2.2.

*The case with Gaussian  $\Pi(\vec{X})$ .* For the special case when  $\Pi(\vec{X})$  is Gaussian and the trace of second moments of  $p(\vec{X})$  coincides with those of  $\Pi$ , i.e.,

$$(2.18a) \quad \int_{\mathbb{R}^M} |\vec{X}|^2 p = \int_{\mathbb{R}^M} |\vec{X}|^2 \Pi,$$

then

$$(2.18b) \quad P(p, \Pi) = \mathcal{S}(\Pi) - \mathcal{S}(p),$$

where  $\mathcal{S}(p)$  is the Shannon entropy,

$$(2.19) \quad \mathcal{S}(p) = - \int_{\mathbb{R}^M} p \ln p.$$

The Shannon entropy measures the average lack of information in a probability measure  $p$ . In this special case  $p_{2L}^*$  is calculated alternatively through the maximum entropy principle

$$(2.20) \quad \mathcal{S}(p_{2L}^*) = \max_{p \in \mathcal{P}_{M_{2L}}} \mathcal{S}(p).$$

Following Jaynes,  $p_{2L}^*$  is the least biased probability distribution given the information in (2.11). The solution of (2.20) when only the mean and second moments are specified,  $p_2^*$ , is the Gaussian distribution,  $p_2^* = p_G^*$ , with this mean and second moments. For  $L = 2$ , so that the first four moments of a probability distribution are specified,

$$(2.21a) \quad \mathcal{S}(p_4^*) \leq \mathcal{S}(p_G^*)$$

and

$$(2.21b) \quad p_4^* = \exp \left( \sum_{m=0}^4 \alpha_m (\lambda - \bar{\lambda})^m \right),$$

where  $\alpha_m$  are the appropriate Lagrange multipliers. The third and fourth moments of  $p_4^*$  are characterized by the skewness and flatness,

$$(2.22) \quad \text{Skew} = \frac{\int (\lambda - \bar{\lambda})^3 p_4^*}{(\int (\lambda - \bar{\lambda})^2 p_4^*)^{3/2}}, \quad \text{Flat} = \frac{\int (\lambda - \bar{\lambda})^4 p_4^*}{(\int (\lambda - \bar{\lambda})^2 p_4^*)^2}.$$

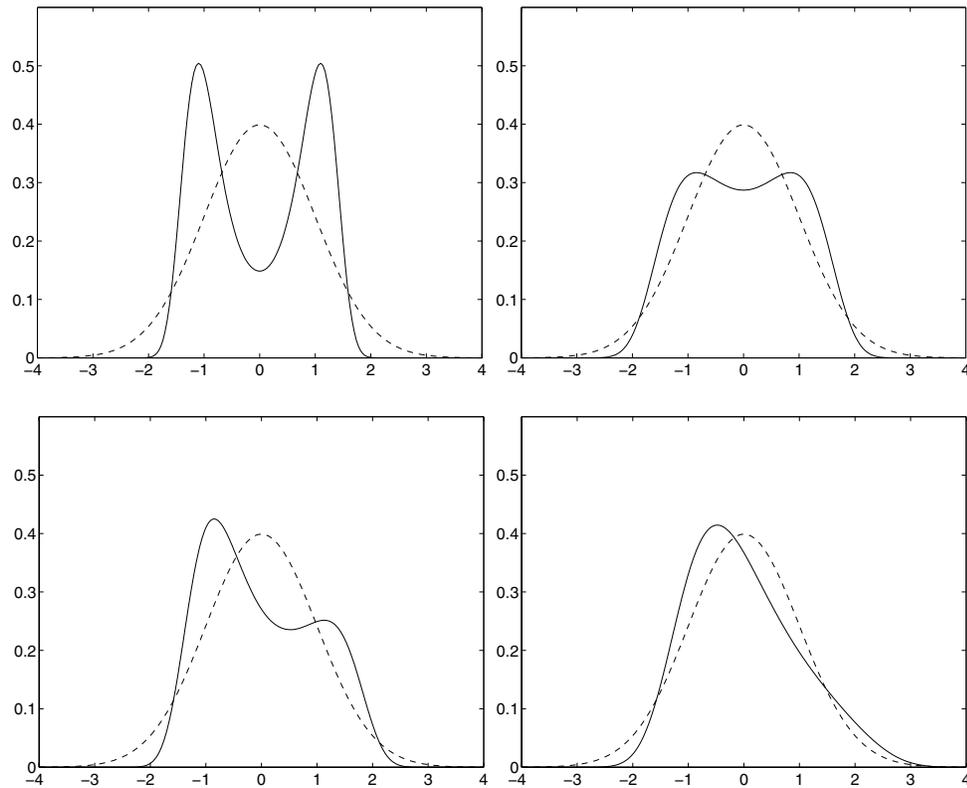


FIG. 1. The probability density functions  $p_4^*$  from (2.21) with zero mean, unit variance, various skewness and flatness (solid lines) versus Gaussians (dashed lines). Upper left—zero skewness, flatness 1.5; upper right—zero skewness, flatness 2; lower left—skewness 0.3, flatness 2; lower right—skewness 0.5, flatness 2.8.

TABLE 1

The utility  $P(p_4, p_G)$  for various skewness and flatness of  $p_4$ , both  $p_4$  and  $p_G$  have zero mean and unit variance. The utility monotonically increases with increasing skewness and/or decreasing flatness.

$P(p_4, p_G)$	Skew $_{p_4} = 0$	Skew $_{p_4} = 0.3$	Skew $_{p_4} = 0.5$
Flat $_{p_4} = 1.5$	0.2772	0.3921	0.6921
Flat $_{p_4} = 2$	$6.015 \cdot 10^{-2}$	$9.879 \cdot 10^{-2}$	0.1885
Flat $_{p_4} = 2.8$	$1.016 \cdot 10^{-3}$	$1.226 \cdot 10^{-2}$	$3.842 \cdot 10^{-2}$

In Figure 1 we graph the probability distributions  $p_4^*$  and  $p_G^*$  for zero mean, variance one, and varying skewness and flatness. The bimodal and skewed behavior of the probability distribution  $p_4^*$  compared with the Gaussian is clearly evident. This is useful for the issues in (B) in section 1. Table 1 presents the quantitative values of the utility  $P(p_4, p_G)$ , demonstrating the same trends from Figure 1.

**2.2. A hierarchical approach for estimating predictive utility with large  $M$ .** Here we consider the important special case when  $\Pi(\vec{X})$  is Gaussian, so without loss of generality,  $\Pi(\vec{X})$  has the form in (2.9). The interested reader can consult section 4 of Majda, Kleeman, and Cai [16] for a generalization when  $\Pi(\vec{X})$  is a product of

factors in some basis. The objective here is to develop a hierarchical computational strategy for estimating  $P(p, \Pi)$  when the dimension  $M$  is large.

The first step in the hierarchy is to calculate the Gaussian estimator,  $p_G^*$ , utilizing only the first and second moments of  $p$  which from section 2.1 satisfies

$$(2.23) \quad P(p, \Pi) \geq P(p_G^*, \Pi).$$

*Principal predictability components.* For estimating predictive utility, following [23], it is very natural to introduce a second change of coordinate which diagonalizes the covariance matrix of  $p$ . Since  $\text{Cov}p$  is a symmetric positive definite  $M \times M$  matrix, there exists a rotation matrix,  $\mathcal{O}$ , with  $\mathcal{O}^T = \mathcal{O}^{-1}$ , so that

$$(2.24) \quad \mathcal{O}^{-1} \text{Cov}p \mathcal{O} = D = \begin{pmatrix} \gamma_1 & 0 & \dots & 0 \\ 0 & \gamma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \gamma_M \end{pmatrix} > 0,$$

where  $D$  is the positive diagonal matrix with nonzero diagonal entries  $\gamma_i > 0$ . Consider the new variable

$$(2.25) \quad \vec{X} = \mathcal{O}\vec{Z}.$$

Since  $\mathcal{O}$  is a rotation matrix, it follows that the Gaussian measure  $\Pi(\vec{Z})$  retains the same normalized form in (2.9).

In this coordinate system, as a consequence of (2.24), the Gaussian distribution  $p_G^*(\vec{Z})$  assumes the factored form

$$(2.26) \quad p_G^*(\vec{Z}) = \prod_{i=1}^M (2\pi\gamma_i)^{-1/2} e^{-(Z_i - \bar{Z}_i)^2 / 2\gamma_i} \stackrel{\text{def}}{=} \prod_{i=1}^M p_i^*(Z_i),$$

while the distribution  $\Pi(\vec{Z})$  retains the factored form

$$(2.27) \quad \Pi(\vec{Z}) = \prod_{i=1}^M (2\pi)^{-1/2} e^{-Z_i^2 / 2} = \prod_{i=1}^M \Pi_i(Z_i).$$

In (2.26),  $\bar{Z}_i$  is the mean of  $p(\vec{Z})$  in the  $i$ th coordinate. In the  $Z$  variables, with (2.26) and (2.27), the Gaussian estimator for the predictability splits into a sum of one-dimensional principal predictability factors

$$(2.28) \quad P(p, \Pi) \geq P(p_G^*, \Pi) = \sum_{i=1}^M P(p_i^*, \Pi_i) = \sum_{i=1}^M \frac{1}{2} [\ln \gamma_i^{-1} + \gamma_i - 1] + \sum_{i=1}^M \frac{1}{2} [\bar{Z}_i^2].$$

These variables obviously can be ordered into a ranked list such as according to the highest predictability component and so on. These variables are called the *principal predictability components* and are a first practical step in understanding the issue from (C) in section 1. This procedure involves only the numerical diagonalization of a symmetric matrix, a well-known and reasonably efficient computational task.

*An identity for predictive utility.* The important ingredient in developing further non-Gaussian estimators for information content is the following identity: in any coordinates  $\vec{X}$  where  $\Pi(\vec{X}) = \prod_{j=1}^M \Pi_j(X_j)$ ,

$$(2.29) \quad P(p, \Pi) = P\left(p, \prod_{j=1}^M p_j(X_j)\right) + \sum_{j=1}^M P(p_j, \Pi_j),$$

where  $p_j(X_j)$  is the  $j$ th marginal one-dimensional probability distribution of  $p$ , i.e.,  $p_j$  is obtained from  $p$  by integrating over all variables except  $X_j$ .

*The first non-Gaussian estimator.* By using the fact that  $P(p, \prod_{j=1}^M p_j) \geq 0$  in the principal predictability components (2.24), it follows easily that for any  $L \geq 1$

$$(2.30) \quad P(p, \Pi) \geq \sum_{j=1}^M P(p_{j,2L}^*, \Pi_j) \geq P(p_G^*, \Pi)$$

with  $p_{j,2L}^*$  being the  $2L$ -dimensional estimator discussed in section 2.1 for the marginal distribution  $p_j(Z_j)$  in the  $j$ th predictability component with  $\Pi_j(Z_j)$  given in (2.26). Thus, already the sum of  $M$  one-dimensional non-Gaussian estimators in the proper coordinates improves over the Gaussian estimate of predictive utility. To further improve the estimate for predictive utility, one needs to find a computationally efficient lower bound for the first term on the right-hand side of (2.29) involving  $p$  alone.

*The second non-Gaussian estimator.* Let  $p_{2L}^*(Z_1 \dots Z_M)$  be a multivariable estimator defined by a subset of all the moments of order  $2L$ , to be specified later, which includes all the  $2L$ -moments of the one-dimensional marginals  $p_j(Z_j)$  in the principal predictability coordinates. By using (2.29) for  $p_{2L}^*$ , we have

$$(2.31) \quad P(p, \Pi) \geq P(p_{2L}^*, \Pi) = P\left(p_{2L}^*, \prod_{j=1}^M p_{j,2L}^*\right) + \sum_{j=1}^M P(p_{j,2L}^*, \Pi_j).$$

Let  $p_{j,l}(Z_j, Z_l)$  denote the two-dimensional marginal distribution obtained by integrating  $p$  over all variables except  $j$  and  $l$ . Assume for simplicity that  $M$  is odd; then it is shown in section 4 of [16] that the first term on the right-hand side of (2.31) is estimated rigorously from below by a sum of  $M - 1$  two-dimensional marginals, i.e.,

$$(2.32) \quad P\left(p_{2L}^*, \prod_{j=1}^M p_{j,2L}^*\right) \geq \sum_{j=0}^{(M-3)/2} [P(p_{2j+1,2j+3,2L}^*, p_{2j+1,2L}^* p_{2j+3,2L}^*) + P(p_{2j+2,2j+3,2L}^*, p_{2j+2,2L}^* p_{2j+3,2L}^*)].$$

Efficient evaluation of the right-hand side of (2.32) combined with (2.31) leads to an improved estimate of utility. Each of the individual terms in (2.31) can be estimated rapidly by the methods of section 2.1 by an optimization with only five additional constraints if  $L = 2$ . From the first non-Gaussian estimator in (2.30), the probability distributions  $p_{j,2L}^*(Z_j)$  and  $p_{k,2L}(Z_k)$  are already known. The estimator  $p_{j,k}^*(Z_j, Z_k)$  includes all the mixed moments of  $p$  in the  $\bar{Z}$ -coordinates up to order  $2L$ ,

$$(2.33) \quad \int_{\mathbb{R}^M} (Z_j - \bar{Z}_j)^\alpha (Z_k - \bar{Z}_k)^\beta p, \quad \begin{matrix} 2 \leq \alpha + \beta \leq 2L, \\ 1 \leq \alpha, \beta. \end{matrix}$$

Thus,  $p_{j,k,2L}^*(Z_j, Z_k)$  can be calculated rapidly by solving the optimization problem

$$(2.34) \quad P(p_{j,k,2L}^*, p_{j,2L}^* p_{k,2L}^*) = \min_{C_{2L}^*} P(p, p_{j,2L}^* p_{k,2L}^*),$$

where the constraints in (2.33) define  $C_{2L}^*$ . This involves a ten-dimensional unconstrained optimization problem for  $L = 2$ . Furthermore, these can be generated separately for any of the two-dimensional marginals needed in (2.32). This avoids the computational catastrophe for  $M \gg 1$  described in the last paragraph of section 2.1 by sampling special mixtures of mixed moments and estimating them cheaply.

*Remark.* The two non-Gaussian estimators for utility can be applied in any coordinate system  $\vec{X}$  where  $\Pi(\vec{X})$  splits into one-dimensional factors. In a given situation, some other basis rather than the principal predictability components might be more useful.

**2.3. The generalized signal/dispersion decomposition.** Here we assume that the probability distribution  $\Pi(\lambda)$  is given as the least-biased distribution generated by the moment optimization problem in (2.20), so that

$$(2.35) \quad \Pi(\lambda) = \exp \left( \sum_{n=0}^{2N} \alpha_n (\lambda - \bar{\lambda}_\Pi)^n \right),$$

where, as in (2.21),  $\alpha_n$  are the Lagrange multipliers for the moment constraints, and  $\alpha_0$  depends on  $\alpha_1 \dots \alpha_{2N}$  as

$$(2.36) \quad \alpha_0 = -\ln \int \exp \left( \sum_{n=1}^{2N} \alpha_n x^n \right) dx.$$

For the probability distribution  $p_{2L}^*$  we impose the  $2L$ -moment constraints,

$$(2.37) \quad f_i = \langle (\lambda - \bar{\lambda}_{p_{2L}^*})^i \rangle_{p_{2L}^*}, \quad i = 1, \dots, 2L.$$

As in section 2.1, we have

$$(2.38) \quad P(p_{2L}^*, \Pi) = \sum_{k=2}^{2L} \theta_k f_k - \alpha_0 - \ln \int \exp \left( \sum_{k=1}^{2L} \theta_k (\lambda - \bar{\lambda}_{p_{2L}^*})^k + \sum_{n=1}^{2N} \alpha_n (\lambda - \bar{\lambda}_\Pi)^n \right) d\lambda.$$

To simplify the integrand in (2.38), we make the linear change of variables

$$(2.39a) \quad \sum_{k=1}^{2L} \theta_k (\lambda - \bar{\lambda}_{p_{2L}^*})^k + \sum_{n=1}^{2N} \alpha_n (\lambda - \bar{\lambda}_\Pi)^n = \sum_{k=0}^{\max(2L, 2N)} \gamma_k (\lambda - \bar{\lambda}_{p_{2L}^*})^k \quad \forall \lambda,$$

$$(2.39b) \quad \gamma_k = \theta_k + \sum_{n=k}^{2N} \binom{n}{k} \alpha_n (\bar{\lambda}_{p_{2L}^*} - \bar{\lambda}_\Pi)^{n-k},$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

With this change of variables we obtain the following expression for the utility:

$$(2.40) \quad \begin{aligned} P(p_{2L}^*, \Pi) &= \sum_{k=2}^{2L} \left( \gamma_k - \sum_{n=k}^{2N} \binom{n}{k} \alpha_n (\bar{\lambda}_{p_{2L}^*} - \bar{\lambda}_\Pi)^{n-k} \right) f_k \\ &\quad + \gamma_0 - \sum_{n=1}^{2N} \alpha_n (\bar{\lambda}_{p_{2L}^*} - \bar{\lambda}_\Pi)^n - \alpha_0, \end{aligned}$$

where  $\gamma_0$  depends on  $\gamma_1 \dots \gamma_{\max(2L, 2N)}$  as in (2.36). In our constraint framework, where all information comes from measured means and moments, we can distinguish two sources of predictive utility: one coming from the difference in mean between  $p_{2L}^*$  and  $\Pi$ , which determines the distance between the PDFs, and the higher moments, which govern the shapes of the PDFs for  $p_{2L}^*$  and  $\Pi$ . Thus it is natural to split the relative entropy into three parts: the first part, *the signal*, is contributed by distance between the means; the second part, *the dispersion*, is contributed by the ensemble variance and higher moments; and the third part, *the cross-term*, comes both from the means and the moments.

$$(2.41a) \quad P(p_{2L}^*, \Pi) = S + D + CT,$$

$$(2.41b) \quad S = - \sum_{k=1}^{2N} \alpha_k (\bar{\lambda}_{p_{2L}^*} - \bar{\lambda}_{\Pi})^k, \quad \text{signal},$$

$$(2.41c) \quad D = (\gamma_0 - \alpha_0) + \sum_{k=2}^{2L} (\gamma_k - \alpha_k) f_k, \quad \text{dispersion},$$

$$(2.41d) \quad CT = - \sum_{k=2}^{2L} \sum_{n=k+1}^{2N} \binom{n}{k} \alpha_n (\bar{\lambda}_{p_{2L}^*} - \bar{\lambda}_{\Pi})^{n-k} f_k, \quad \text{cross-term}.$$

This approach has been first introduced by Kleeman [12] for Gaussian constraints (only signal and dispersion because the cross-term is zero for these constraints) and later developed for more general situations by Majda, Kleeman, and Cai [16]. The signal, dispersion, and cross-term provide basic building blocks from which practical predictability measures can be constructed. The role of the cross-term (2.41d) is viewed as the fine-tuning of a practical predictability measure, since it has contributions from both the means and the moments, but involves only the non-Gaussian contributions from  $\Pi$  through  $n$ ,  $3 \leq n \leq 2N$  in (2.35).

The signal, dispersion, and cross-term possess the following important properties:

- In the case of a Gaussian  $\Pi(\lambda)$  the cross-term is identically zero.
- Both the signal and the cross-term do not depend on the ensemble Lagrange multipliers. That makes them both very attractive measures of the variation of predictive utility which can be evaluated rapidly in terms of the known moments for both  $\Pi$  and  $p$ .
- The dispersion, like the total utility, is nonnegative. The reason for that is that, by definition,

$$(2.42) \quad D = \left( \sum_k \gamma_k f_k + \gamma_0 \right) - \left( \sum_k \alpha_k f_k + \alpha_0 \right) = \max_{\vec{x} \in \mathbb{R}^{\max(2L, 2N)}} F(\vec{x}) - F(\vec{\alpha}) \geq 0,$$

where

$$F(\vec{x}) = \sum_k x_k f_k + x_0.$$

- The signal, cross-term, or the sum of both can be both positive or negative, which also implies that the dispersion can be greater than the total predictive utility. For example, for  $\Pi(\lambda)$  with zero mean, unit variance, zero skewness, and flatness Flat = 2, and for the ensemble with the mean Mean = -0.2, unitary variance, skewness Skew = 0.5 and flatness Flat = 2, the total predictive utility is  $P = 8.645 \cdot 10^{-2}$ , whereas the dispersion is  $D = 1.284 \cdot 10^{-1}$ , i.e., the dispersion exceeds the total utility by 48.5%. Moreover, both the signal and the cross-term for this set of constraints are negative,  $S = -1.082 \cdot 10^{-2}$ ,  $CT = -3.114 \cdot 10^{-2}$ , which means that combining the cross-term either with signal or with dispersion will not make all components of the total information content positive definite. However, this is a pathological example, and for the vast majority of cases encountered here the sum of the signal and the cross-term is positive. In models with a highly non-Gaussian climate, this phenomenon happens more readily [5].

Here we propose the *generalized signal*

$$(2.43) \quad \tilde{S} = S + CT.$$

Since the sum of the generalized signal (2.43) and dispersion (2.41c) is the predictive utility, the advantages and disadvantages of the generalized signal are somewhat opposite to dispersion: it is cheap to compute (because prediction Lagrange multipliers are not needed), it largely relies upon the difference between prediction and prior means (because the signal is incorporated), weakly relies upon the prediction and climate shapes (because the cross-term depends on prediction moments), and can be negative (which may be inconvenient in certain situations). The fact that the sum of generalized signal and dispersion equals relative entropy is also the reason we pick these two measures for testing in this paper. We apply them in section 4 for the practical issue from (E).

*Important technical remarks.*

- Formula (2.42) can be confusing for a reader not acquainted with convex optimization; in particular, the following question may arise: How, *minimizing* the utility, could one end up with a part of it being the *maximum*? It follows from convex optimization theory that the *constrained maximization* of a functional  $S(p)$  (in our case it is  $-P(p, \Pi)$ ) is equivalent to the *unconstrained minimization* of a *Lagrangian function*

$$\mathcal{L}(\theta_0, \dots, \theta_M) = S(p) + \sum_{m=1}^M \theta_m (\mathcal{F}_m(p) - f_m),$$

where  $\mathcal{F}_m(p)$  are the expressions for moments in (2.37), and  $f_m$  are the constraints to satisfy, such that for optimal  $p^*$  we have  $\mathcal{F}_m(p^*) = f_m$ . The  $\theta_m$  are the Lagrange multipliers. Explicit calculation yields that the dispersion from (2.41c) is

$$(2.44) \quad D = -\mathcal{L}(\gamma_0, \dots, \gamma_{2L}) - \alpha_0 - \sum_{k=2}^{2L} \alpha_k f_k.$$

Since the Lagrangian is found as the unconstrained minimum, it follows from (2.44) that the dispersion is the unconstrained maximum in the space of Lagrange multipliers  $\vec{\gamma}$ . For more extensive information see [26].

- It is clear from the structure of (2.44) that the sets of Lagrange multipliers  $\bar{\gamma}$  and  $\bar{\alpha}$  (for  $p^*$  and  $\Pi$ , respectively) can be found independently by optimizing Shannon entropies  $\mathcal{S}(p)$  and  $\mathcal{S}(\Pi)$  in (2.19), which allows an independent rescaling of the constraints for  $p^*$  and  $\Pi$  to have unit variance. This eliminates machine-dependent arithmetic range problems and facilitates convergence of modified Newton iterations in cases where  $\text{Var}(p) \ll \text{Var}(\Pi)$  or vice versa as well as tailors the optimization algorithm for parallel execution. We again emphasize that in the current paper the role of utility is not viewed only as a relative information content in the ensemble beyond that of the climate (where  $\Pi$  has to be evaluated once), but also as a suitable metric for estimating differences between two ensembles with a different number of constraints (in this case  $\Pi$  has to be evaluated as often as  $p$ ). The latter application is greatly facilitated by independent computation of  $\bar{\gamma}$  and  $\bar{\alpha}$ .

### 3. Two simplified models with statistical features of the atmosphere:

**Climate statistics.** Here we introduce two simplified models developed recently, which have important features in common with the atmosphere in a vastly simpler setting. When damping and forcing are neglected, both models are suitable discrete truncations of the Burgers–Hopf equation

$$(3.1) \quad u_t + \frac{1}{2}(u^2)_x = 0$$

in the periodic setting. For both models, the statistical behavior of the discrete system is the object of interest rather than their use as numerical approximations to (3.1). In fact both discrete systems are rather poor numerical methods for (3.1) in the continuum limit.

**3.1. The Galerkin truncation of the Burgers–Hopf equation.** One of the models is the Fourier–Galerkin truncation of (3.1), introduced recently by Majda and Timofeyev [17] as a simple model for statistical features of the atmosphere. Since the remarkable statistical features of this model have been discussed in detail recently in the applied mathematics literature [17, 18, 8, 1], next we only summarize briefly the structural and statistical features of this model. Below  $u_\Lambda$  denotes the Fourier–Galerkin truncation for  $(2\Lambda+1)$  modes  $\hat{u}_k$  with  $0 \leq |k| \leq \Lambda$ . The value  $\Lambda = 50$  is used below in describing the statistical results as well as whenever this model is utilized in the present paper so that there are 100 active degrees of freedom. The truncated Burgers–Hopf model (TBH) has the following features:

(A) Structural features of TBH.

- (1) The momentum,  $M = \int u_\Lambda$ , and the energy,  $E = \frac{1}{2} \int u_\Lambda^2$ , are conserved.
- (2) The third power,  $H = \int u_\Lambda^3$ , is conserved, and, in fact, TBH is a Hamiltonian system with Hamiltonian  $H$  [1].

(B) Statistical features of TBH.

- (1) TBH has a Gaussian invariant measure based on the Gibbs ensemble for the energy.
- (2) In a very precise sense, for most values of the mean energy  $E$  and Hamiltonian  $H$ , the statistical predictions of the energy Gibbs ensemble from (1) are confirmed, including the following.
  - (a) equipartition of energy among Fourier modes.
  - (b) A Gaussian PDF for each active Fourier mode, so the Hamiltonian  $H$  is statistically irrelevant [17, 18, 1]. The Fourier modes are the EOFs of TBH.

- (3) For statistically rare values of the Hamiltonian  $H$ , in a precise sense predicted by equilibrium statistical mechanics,
  - (a) there is a spectral tilt correction to equipartition of energy [1],
  - (b) the PDFs of individual Fourier modes remain Gaussian.
- (4) Mixing and correlation scaling law: If  $\int \hat{u}_k(t)\hat{u}_k(t+\tau)dt = C_k(\tau)$  denotes the correlation function of either the real or imaginary part of the  $k$ th Fourier mode, these correlations decay exponentially with time; with  $C_k$  being the associated correlation time, they satisfy the scaling law,

$$C_k \cong \frac{C_0}{k}, \quad 1 \leq |k| \leq 50,$$

i.e., the larger spatial scales decorrelate more slowly than the smaller spatial scales; this is a key statistical feature of TBH in common with the atmosphere [17, 18, 1].

The features in (B) document the “climatology” of TBH. It is worth remarking here that there are other discrete finite difference approximations to the Burgers–Hopf equation with similar statistical features as listed above, as well as other energy-conserving finite difference schemes sharing some of the above statistical features but also having interesting discrepancies [18]. There also are interesting systematic finite difference truncations of the Burgers–Hopf equation with a non-Gaussian statistical “climatology” [3]. In the above list and elsewhere in this paper, statistical features are determined from numerical solutions by time averaging, i.e.,

$$(3.2) \quad \langle f \rangle = \frac{1}{T} \int_{T_0}^{T_0+T} f(s)ds,$$

where  $T_0$  is a suitable startup time and  $T$  is a long time averaging window. These parameters for TBH are well documented in the above references. They will be listed below for the second toy climate model which we describe next.

**3.2. The Lorenz ’96 model.** The damped forced Lorenz ’96 (L96) model is the spatially discrete family of equations given by

$$(3.3) \quad \frac{du_j}{dt} = (u_{j+1} - u_{j-2})u_{j-1} - u_j + F, \quad j = 0, 1, \dots, J - 1,$$

with periodic boundary conditions,  $u_0 = u_J$ . The term  $-u_j$  in (3.3) represents damping (with a unit time scale of five days) while  $F$  represents constant “solar forcing” (see [15, 14]). The model in (3.3) is designed to mimic midlatitude weather and climate behavior, so periodic boundary conditions are appropriate. The unit spatial scale between discrete nodes is regarded as a nondimensional midlatitude Rossby radius  $\approx 1000$  km, and the discrete system size is set to be  $J = 40$  nodes.

In midlatitude weather systems, the main “weather waves,” the Rossby waves, have westward (toward negative  $x$ ) phase velocity, but from our own anecdotal experience, weather systems collectively move eastward (toward positive  $x$ ) with unstable behavior. The models in (3.3) have analogous behavior. To see this, note that  $u_j \equiv F$  defines a steady state solution of (3.3); in standard fashion [14] linearized stability analysis has been done at this steady state. The modes of the system are discrete Fourier modes with wavenumbers  $k$  ranging  $-20 < k \leq 20$ . Associated with each spatial wavenumber is a phase velocity and growth/decay rate as well as a group velocity. The phase and group velocities as well as the growth/decay rates for all the

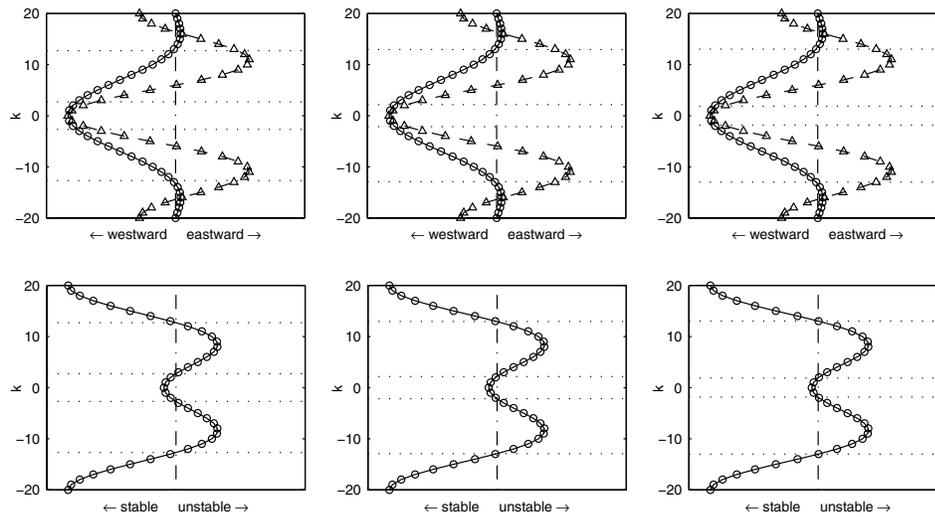


FIG. 2. The phase and group velocity profiles and bands of stable and unstable waves for the damped forced Lorenz model, forcing regimes (left to right)  $F = 4, 6, 8$ . Upper pictures: Solid line with circles—phase velocity, dashed line with triangles—group velocity. Lower pictures: Solid line with circles shows linear stability. All pictures: Horizontal dotted lines mark boundaries between stable and unstable bands, vertical dash-dotted lines denote boundaries between westward and eastward velocity directions or between regions of linear stability and instability.

wavenumbers are depicted in Figure 2 for the representative values  $F = 4, 6, 8$  utilized in this paper. Note from Figure 2 that there are bands of unstable waves in all these cases centered about the wavenumber  $|k| = 8$  with westward phase velocities and overall eastward group velocities, analogous to real weather systems. The growth rate of these linearized instabilities increases as  $F$  increases, but has been normalized in Figure 2. The instability wavelength band around spatial wavenumber  $|k| = 8$  is even a reasonable value for large-scale baroclinic instability in the atmosphere. Readers interested in a mathematical introduction to geophysical flows can consult [20].

**3.3. The inviscid Lorenz '96 model.** If dissipation and forcing are set to zero in the L96 model from (3.3), the inviscid Lorenz '96 (IL96) model results in

$$(3.4) \quad \frac{du_j}{dt} = (u_{j+1} - u_{j-2})u_{j-1}.$$

This is a finite difference approximation to the Burgers–Hopf equation in (3.1) with the following features.

(A) Structural features of IL96.

- (1) Solutions of IL96 conserve the energy  $E = \frac{1}{2} \sum_{j=0}^{J-1} u_j^2$ , and (3.4) satisfies the Liouville property.
- (2) The discrete linear momentum  $M = \sum_{j=0}^{J-1} u_j$  is not conserved by (3.4) as well as the discrete third power.

(B) Statistical features of IL96.

- (1) IL96 has a Gaussian invariant measure based on the Gibbs ensemble for energy as a consequence of (A1) above.
- (2) The equilibrium statistical predictions of (B1) are satisfied including the following.

TABLE 2

The spatial climate characteristics for statistical regimes  $F = 0, 4, 6, 8$ . The last column gives the energy of the fluctuations around the mean climate.

	Mean Climate	Mean Energy	Energy of the Mean Climate
$F = 0$	0	9.363	9.363
$F = 4$	1.193	2.386	1.674
$F = 6$	2.011	6.034	4.011
$F = 8$	2.341	9.363	6.624

- (a) equipartition of energy among the Fourier modes, which are the EOFs for this system,
  - (b) a Gaussian PDF for each of the Fourier modes.
- (3) Solutions of (3.4) have exponential decay of the correlation functions  $C_k(\tau)$ , but the correlation times of individual Fourier modes  $T_k$  are all comparable and do not exhibit the scaling behavior in (B4) as for TBH.

The IL96 model is an interesting bridge between TBH and L96 since IL96 has an equilibrium statistical formalism like TBH yet has very different behavior for decay of correlation times. On the other hand, IL96 has correlation time behavior of individual Fourier modes comparable to those of L96 with  $F = 8$ . There is a conventional wisdom that correlation times control predictability and IL96 is a useful model in this regard. The statistical behavior listed in (B) in section 3.3 is documented in the tables and figures presented below.

**3.4. The climatology for L96 and IL96.** The statistical steady state, i.e., the climatology, for L96 and IL96 with  $J = 40$  is calculated below by numerical integration of (3.3) and (3.4) through the sixth order explicit Adams–Bashforth method with time step  $\Delta t = 10^{-3}$ . This statistical behavior of solutions is calculated through time averaging an individual solution with random initial data with an averaging window  $T = 10^5$  with PDFs calculated by standard bin-counting procedures [1]. For TBH and IL96, the energy is conserved within  $10^{-8}\%$  by this numerical procedure. Below, we summarize statistical features of the climatology for L96 with the three parameter values  $F = 4, 6, 8$  while the climatology for IL96 is denoted by  $F = 0$  on the figures below.

In Table 2 we show the mean climate by calculating the time mean of each discrete node,  $u_j$ , in (3.2), which is independent of  $j$  and denoted by  $\bar{u}_F$ ; we also present the energy of this mean state compared with the mean energy of fluctuations, i.e., utilizing  $f = \frac{1}{2} \sum_{j=0}^{J-1} (u_j - \bar{u}_F)^2$  in (3.2). In Figure 3 we show the spectral variance, i.e.,  $\langle |\hat{u}_k - \bar{\hat{u}}_k|^2 \rangle$  for the individual Fourier modes for  $F = 0, 4, 6, 8$ . The equipartition of energy for the Fourier modes for IL96 is apparent as well as distinct peaks in variance around wavenumber  $k = 8$ , the most unstable wavenumber, for  $F = 4, 6, 8$ . This peak diminishes as  $F$  increases and, for  $F = 4$ , there is a pronounced second peak at  $k = 16$  indicating strong subharmonic instability. Since both L96 and IL96 are translation invariant, the Fourier modes are the EOFs, and the covariance matrix  $\langle (u_i - \bar{u}_F)(u_j - \bar{u}_F) \rangle = C_{|i-j|}$  is a discrete convolution operator in the statistical steady state. To get insight into the statistical dynamics, the correlation times of individual Fourier modes are depicted in Figure 4. There are roughly comparable correlation times for IL96 as claimed above and correlation time peaks for the band of unstable modes around  $k = 8$  for  $F = 4, 6, 8$ ; these correlations indicate strong mixing behavior in the system for  $F = 0, 6, 8$  with less pronounced mixing for  $F = 4$  and extremely slow decay of correlations for  $k = 8, 16$ . In Figure 5, we plot the covariance matrix in

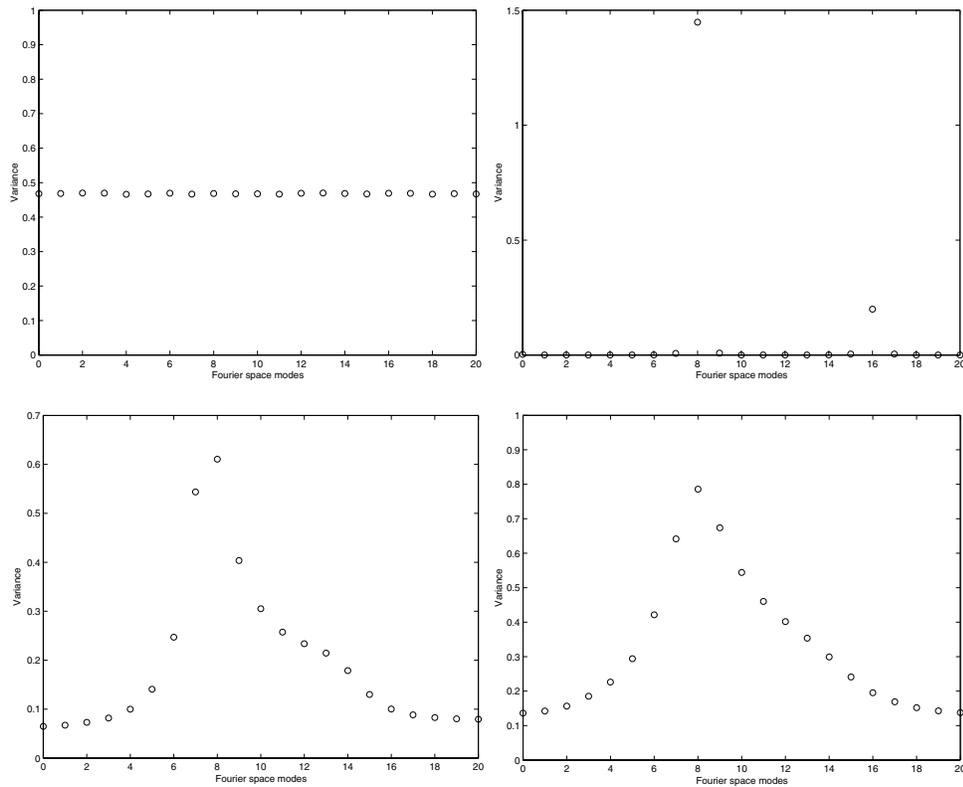


FIG. 3. The variance spectrum in Fourier space. Upper left—*I* L96 model; upper right—L96 model with constant forcing  $F = 4$ ; lower left—L96 model with constant forcing  $F = 6$ ; lower right—L96 model with constant forcing  $F = 8$ .

physical space for the four cases. For  $F = 0$  we get essentially uncorrelated behavior as predicted in (B1) in section 3.3 while the correlation matrix is successively less localized as  $F$  decreases through the values,  $F = 8, 6, 4$ .

The structure of the correlations in Figures 4 and 5 suggest strongly mixed chaotic waves for  $F = 0, 6, 8$  and weakly chaotic behavior for  $F = 4$  dominated by subharmonic instability. The snapshots from the time series of the solutions presented in Figure 6 confirm this. The PDFs for the individual Fourier modes  $\hat{u}_k$  with  $k = 0, 3, 8$  are nearly Gaussian for the cases  $F = 0, 6, 8$  with the data confirming this presented in Table 3. Note that the mode  $k = 0$  corresponds to the mean climate. On the other hand, for  $F = 4$  the PDFs (not depicted here) are highly non-Gaussian and bimodal, which is an indication of the weakly chaotic behavior observed here. Also, the PDF of an individual spatial node is skewed and non-Gaussian for  $F = 8$  with skewness  $\text{Skew} = 9.309 \cdot 10^{-2}$  and flatness  $\text{Flat} = 2.483$ . In Figure 7 we graph this PDF as well as the statistical estimator using the first four moments calculated through the maximum entropy principle from section 2.2.

Principal predictability components from section 2.2 require a Gaussian climate in the form (2.9), so it is important to know how close is the actual climate to the Gaussian. We define the two-mode third and fourth order nondimensional moments

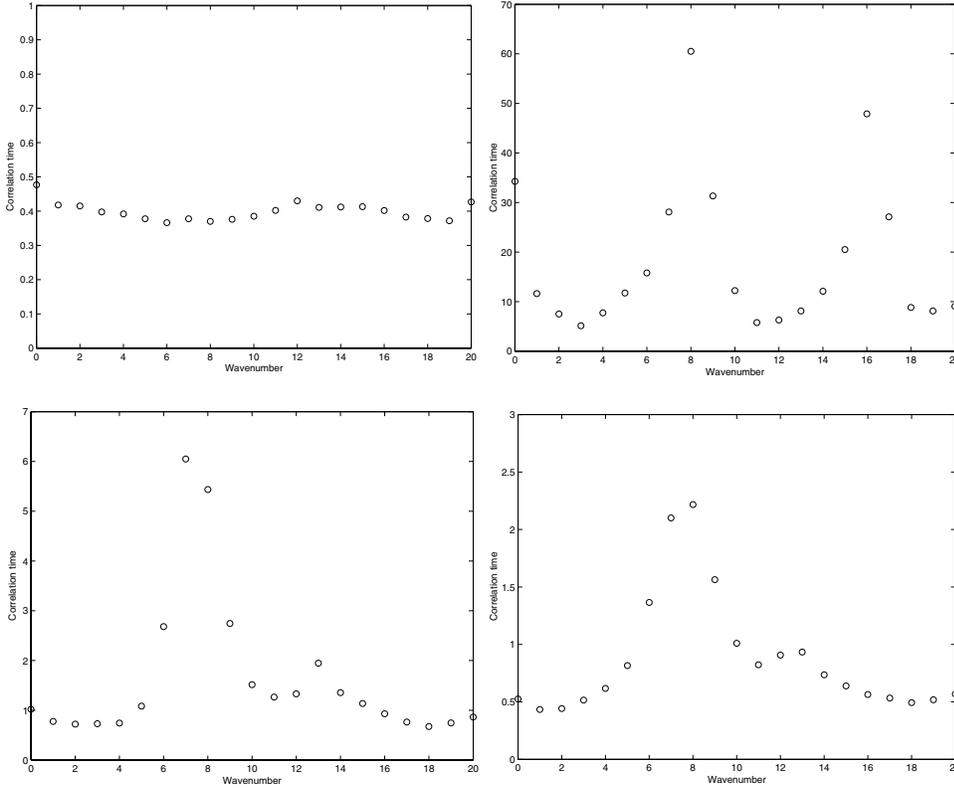


FIG. 4. The correlation time scaling in Fourier space. Upper left—IL96 model; upper right—L96 model with constant forcing  $F = 4$ ; lower left—L96 model with constant forcing  $F = 6$ ; lower right—L96 model with constant forcing  $F = 8$ .

as

$$(3.5) \quad M_{k_1 k_2}^{\alpha\beta} = \frac{\langle \text{Re}(\hat{u}_{k_1} - \bar{\hat{u}}_{k_1})^\alpha \text{Re}(\hat{u}_{k_2} - \bar{\hat{u}}_{k_2})^\beta \rangle}{(\text{Var}_{k_1})^{\alpha/2} (\text{Var}_{k_2})^{\beta/2}}.$$

For  $\alpha, \beta > 0$ ,  $3 \leq \alpha + \beta \leq 4$ ,  $\alpha \neq \beta$ , and Gaussian climate from (2.9),  $M_{k_1 k_2}^{\alpha\alpha} = 1$  and  $M_{k_1 k_2}^{\alpha\beta} = 0$ . In Tables 4 and 5 we show the moments (3.5),  $3 \leq \alpha + \beta \leq 4$  for the Fourier EOF modes  $k = 0, 3, 8$  and regimes  $F = 0, 6, 8$ . There are no significant cross-correlations of the third order between the modes except for  $M_{30}^{21}$  and  $M_{80}^{21}$  for the regimes  $F = 6, 8$ . These two cross-correlations show the wave mean flow interaction with the linearly stable  $k = 3$  and linearly unstable  $k = 8$ . It is remarkable that  $M_{30}^{21}$  is positive for both  $F = 6$  and  $F = 8$  (0.1275 and 0.1072, respectively), whereas  $M_{80}^{21}$  is negative for both  $F = 6$  and  $F = 8$  (−0.2069 and −0.1511, respectively), because the sign is defined by the mean climate mode  $k = 0$  (contributes linearly into both  $M_{30}^{21}$  and  $M_{80}^{21}$ ). The fourth order cross-correlations  $M_{k_1 k_2}^{22}$  are close to Gaussian value 1, with  $M_{38}^{22}$  exhibiting the farthest deviations from 1 (within 10%) for the regimes  $F = 6, 8$ . The off-diagonal fourth order moments  $M_{k_1 k_2}^{31}$  are nearly zero in all cases.

We construct the joint probability density function for the two-dimensional phase space  $\{k = 0, k = 8\}$  by maximizing the absolute lack of information from (2.19) according to the strategy given in section 2.2 subject to all moment constraints in

TABLE 3

The individual modal statistics for regimes  $F = 0, 6, 8$ . Note that for regimes  $F = 6, 8$  (L96) climate mode  $k = 0$  is skewed, and linearly unstable mode  $k = 8$  is sub-Gaussian (flatness  $< 3$ ).

$F = 0$	Variance	Skewness	Flatness
$k = 0$	$4.681 \cdot 10^{-1}$	$8.057 \cdot 10^{-4}$	2.853
$k = 3$	$2.356 \cdot 10^{-1}$	$1.037 \cdot 10^{-3}$	2.853
$k = 8$	$2.347 \cdot 10^{-1}$	$3.423 \cdot 10^{-3}$	2.849

$F = 6$	Mean	Variance	Skewness	Flatness
$k = 0$	2.011	$6.491 \cdot 10^{-2}$	$1.868 \cdot 10^{-1}$	2.939
$k = 3$	$-1.337 \cdot 10^{-4}$	$4.100 \cdot 10^{-2}$	$1.591 \cdot 10^{-3}$	2.998
$k = 8$	$4.476 \cdot 10^{-4}$	$3.054 \cdot 10^{-1}$	$1.543 \cdot 10^{-3}$	2.568

$F = 8$	Mean	Variance	Skewness	Flatness
$k = 0$	2.341	$1.363 \cdot 10^{-1}$	$8.128 \cdot 10^{-2}$	2.952
$k = 3$	$-1.889 \cdot 10^{-4}$	$9.246 \cdot 10^{-2}$	$-4.802 \cdot 10^{-4}$	2.989
$k = 8$	$9.339 \cdot 10^{-4}$	$3.926 \cdot 10^{-1}$	$-4.791 \cdot 10^{-3}$	2.681

TABLE 4

The third order cross-modal statistics  $M_{k_1 k_2}^{21}$  for regimes  $F = 0, 6, 8$ . Note how the climate mode  $k = 0$  interacts with  $k = 3$  and  $k = 8$  through  $M_{30}^{21}$  and  $M_{80}^{21}$  for L96 with  $F = 6, 8$ .

$F = 0$	$k_1 = 0$	$k_1 = 3$	$k_1 = 8$
$k_2 = 0$	–	$M_{03}^{21} = -1.995 \cdot 10^{-2}$	$M_{08}^{21} = 3.404 \cdot 10^{-3}$
$k_2 = 3$	$M_{30}^{21} = 1.737 \cdot 10^{-2}$	–	$M_{38}^{21} = 1.751 \cdot 10^{-4}$
$k_2 = 8$	$M_{80}^{21} = 1.815 \cdot 10^{-2}$	$M_{83}^{21} = -2.206 \cdot 10^{-2}$	–

$F = 6$	$k_1 = 0$	$k_1 = 3$	$k_1 = 8$
$k_2 = 0$	–	$M_{03}^{21} = 1.036 \cdot 10^{-2}$	$M_{08}^{21} = -1.451 \cdot 10^{-3}$
$k_2 = 3$	$M_{30}^{21} = 0.1275$	–	$M_{38}^{21} = -5.419 \cdot 10^{-3}$
$k_2 = 8$	$M_{80}^{21} = -0.2069$	$M_{83}^{21} = -9.573 \cdot 10^{-3}$	–

$F = 8$	$k_1 = 0$	$k_1 = 3$	$k_1 = 8$
$k_2 = 0$	–	$M_{03}^{21} = 4.532 \cdot 10^{-3}$	$M_{08}^{21} = 1.118 \cdot 10^{-2}$
$k_2 = 3$	$M_{30}^{21} = 0.1072$	–	$M_{38}^{21} = -2.249 \cdot 10^{-3}$
$k_2 = 8$	$M_{80}^{21} = -0.1511$	$M_{83}^{21} = -7.112 \cdot 10^{-3}$	–

TABLE 5

The fourth order cross-modal statistics  $M_{k_1 k_2}^{31}$  and  $M_{k_1 k_2}^{22}$  for regimes  $F = 0, 6, 8$  ( $M_{k_1 k_2}^{22}$  are placed in the main diagonals of the tables). Note that  $M_{38}^{22}$  are significantly below Gaussian value 1 in all regimes.

$F = 0$	$k_1 = 0$	$k_1 = 3$	$k_1 = 8$
$k_2 = 0$	$M_{03}^{22} = 0.9539$	$M_{03}^{31} = -1.295 \cdot 10^{-2}$	$M_{08}^{31} = -8.469 \cdot 10^{-4}$
$k_2 = 3$	$M_{30}^{31} = -4519 \cdot 10^{-3}$	$M_{08}^{22} = 0.971$	$M_{38}^{31} = 6.995 \cdot 10^{-3}$
$k_2 = 8$	$M_{80}^{31} = 4.102 \cdot 10^{-3}$	$M_{83}^{31} = 7.557 \cdot 10^{-3}$	$M_{38}^{22} = 0.9361$

$F = 6$	$k_1 = 0$	$k_1 = 3$	$k_1 = 8$
$k_2 = 0$	$M_{03}^{22} = 1.017$	$M_{03}^{31} = 1.746 \cdot 10^{-2}$	$M_{08}^{31} = 1.285 \cdot 10^{-3}$
$k_2 = 3$	$M_{30}^{31} = 4.675 \cdot 10^{-3}$	$M_{08}^{22} = 0.9972$	$M_{38}^{31} = 2.167 \cdot 10^{-2}$
$k_2 = 8$	$M_{80}^{31} = 2.29 \cdot 10^{-3}$	$M_{83}^{31} = 1.919 \cdot 10^{-2}$	$M_{38}^{22} = 0.9063$

$F = 8$	$k_1 = 0$	$k_1 = 3$	$k_1 = 8$
$k_2 = 0$	$M_{03}^{22} = 0.9662$	$M_{03}^{31} = 4.195 \cdot 10^{-2}$	$M_{08}^{31} = -1.219 \cdot 10^{-2}$
$k_2 = 3$	$M_{30}^{31} = 2.875 \cdot 10^{-2}$	$M_{08}^{22} = 0.9769$	$M_{38}^{31} = -2.962 \cdot 10^{-2}$
$k_2 = 8$	$M_{80}^{31} = -1.843 \cdot 10^{-2}$	$M_{83}^{31} = -2.338 \cdot 10^{-2}$	$M_{38}^{22} = 0.9141$

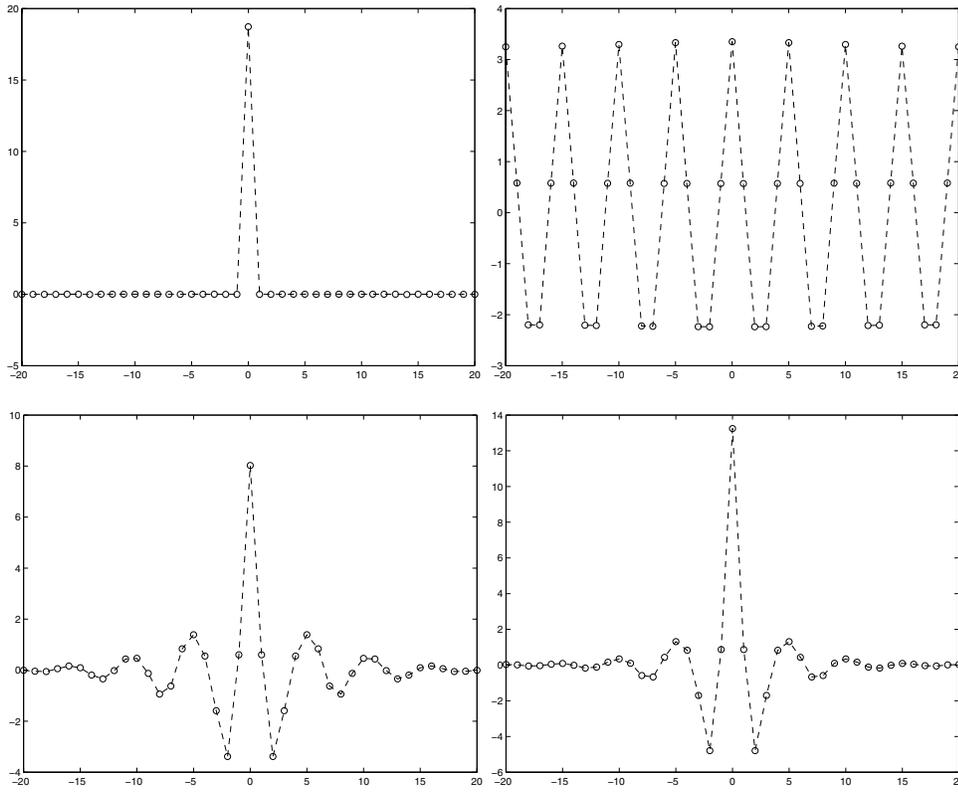


FIG. 5. The covariance matrix in form of spatial correlations. Upper left—IL96 model; upper right—L96 model with constant forcing  $F = 4$ ; lower left—L96 model with constant forcing  $F = 6$ ; lower right—L96 model with constant forcing  $F = 8$ .

Tables 3, 4, and 5 (overall ten constraints). This particular phase space is chosen for its large nontrivial skewness in the cross-third moment  $M_{80}^{21} = -0.2069$ , and the optimized PDF is shown on the left plot in Figure 8. For comparison, on the right plot in Figure 8 we show the “actual” PDF determined by a straightforward bin-counting routine from the long time series of the climate [1] with bin size 0.07. The profound similarity of the two PDFs demonstrates the success of the optimization approach in section 2.2 in determining the information content through recorded moments for a two-dimensional phase space. The PDF determined through the bin counting (right plot in Figure 8) looks jagged because of the finite size of the bins, whereas the optimized PDF on the left plot of Figure 8 is a smooth function of the phase space variables by construction. The skewness in the Fourier EOF mode  $k = 0$  affects the shape of the optimized PDF, which is significantly skewed in that direction.

**4. Quantifying the information in ensemble predictions.** Here we apply the theory developed in section 2 to ensemble prediction in the L96 and IL96 models with the climatology described in section 3 above for varying  $F = 0, 4, 6, 8$ . We also briefly mention some non-Gaussian aspects of ensemble prediction for TBH to compare and contrast these with L96 and IL96. We refer the reader to [8] for the Gaussian predictability theory in TBH. In all of the predictability studies reported here we take a long time series for the climate and at a given time,  $T_k$ , we generate a 100-member

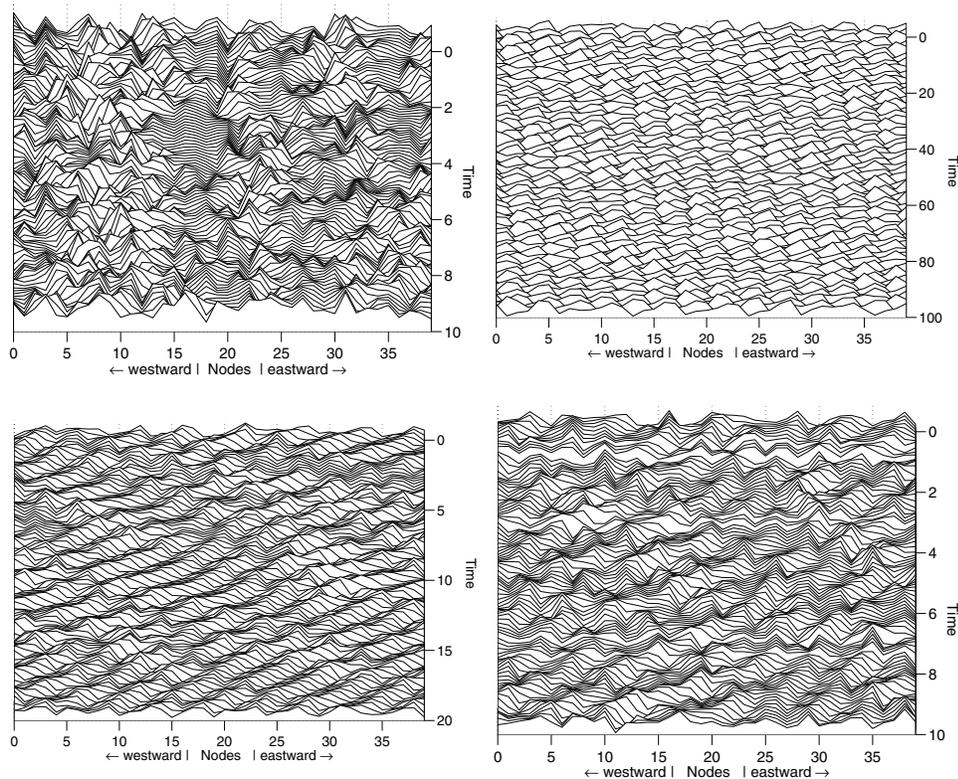


FIG. 6. The time evolution of actual mixing solution. Upper left—IL96 model; upper right—L96 model with constant forcing  $F = 4$ ; lower left—L96 model with constant forcing  $F = 6$ ; lower right—L96 model with constant forcing  $F = 8$ .

ensemble of initial data by sampling a Gaussian distribution with mean given by the value of the climate record at  $T$  and variance,  $\sigma^2$ , of each spatial node,  $u_j$ , given by  $\sigma^2 = 10^{-2}$ . This low order ensemble with  $R = 100$  in (1.4) is utilized to mimic real predictions in complex systems with small sample size. In order to understand the variability of predictive utility in a given system, a 100-member ensemble of ensembles is utilized with each of the 100 members generated by the above procedure at 100 different times,  $T_k$ , in the climate records which are sufficiently spaced to guarantee independence [12].

**4.1. The temporal behavior of predictive utility.** Here we get the first crude insight into the behavior of predictive utility  $P(p, \Pi_c) \geq P(p_4^*, \Pi_c)$  by measuring the lack of information in the climate through four moment estimators. In Figure 9, we graph the mean of predictive utility with time over the 100 member ensemble of ensembles for  $F = 0, 4, 6, 8$  and the three Fourier EOF modes,  $k = 0, 3, 8$ . Recall from section 3 that  $k = 0$  is the mean climate mode, while  $k = 8$  is a linearly unstable mode for  $F = 4, 6, 8$  with  $k = 3$ , a linearly stable mode. In each ensemble, the predictive utility is calculated through the first four moments of the prediction ensemble as described in section 2.1. It is apparent that the case  $F = 4$  is nearly integrable with weak chaotic dynamics, which results in nondecaying predictive utility (marked as a dotted line in Figure 9). On the contrary, the regimes  $F = 0, 6, 8$  are dynamically

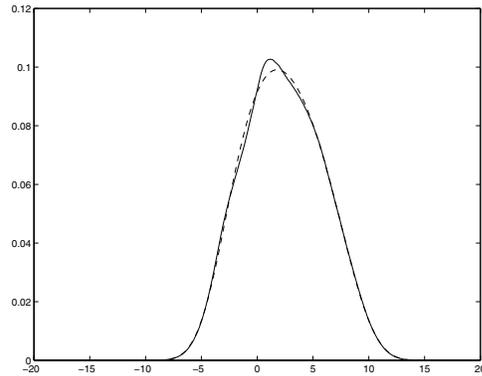


FIG. 7. Solid line: The averaged spatial PDF for statistical regime  $F = 8$ . Dashed line: The PDF with the same mean, variance, skewness, and flatness, produced by the four moment numerical estimator from (2.21).

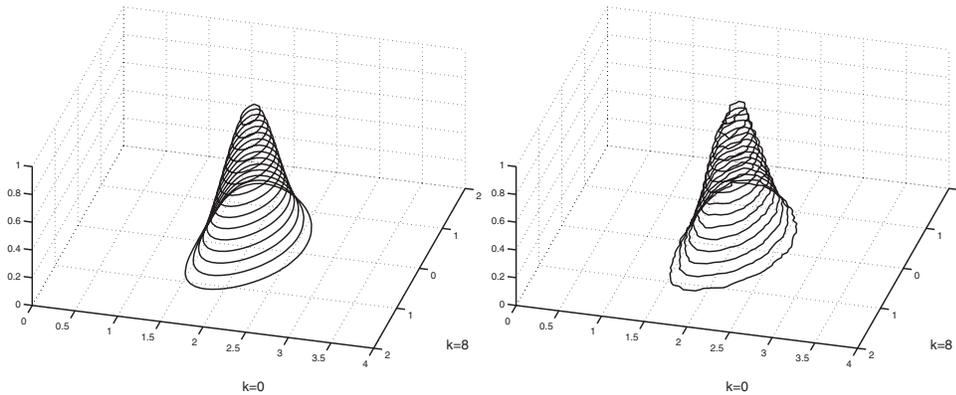


FIG. 8. The joint PDFs for the Fourier EOF modes  $k = 0$  and  $k = 8$ , regime  $F = 6$ . Left: The PDF determined by maximizing the absolute lack in (2.19) according to the strategy given in section 2.2 with all the moment constraints in Tables 3, 4, and 5 (overall ten constraints). Right: The PDF found by straightforward bin counting (bin size 0.07). The contours are successive level curves of PDFs with a height increment 0.0636.

robust chaotic systems, as observed in Figure 9, with the decay of mean predictability essentially the same for the three different Fourier modes. Clearly, for L96 with forcing  $F = 8$  there is the most rapid decay of this mean predictive utility (typical decay time is about  $t = 5$ ), whereas the regimes  $F = 0$  and  $F = 6$  have more moderate predictability (typical decay time  $t = 10$ ).

While the graphs in Figure 9 give some insight into the predictive utility as  $F$  varies, they are somewhat misleading for a fixed value of  $F$  because the distribution of predictive utility over the 100-member ensemble of ensembles typically has a rather wide spread. To demonstrate this, in Figure 10 we present the histograms of predictive utility generated by bin counting for the 100-member ensemble of ensembles as time evolves for  $F = 8$  for both the mean climate mode,  $k = 0$ , and the unstable mode,  $k = 8$ . We observe that predictability histograms can be quite broad for both  $k = 0$  and  $k = 8$  (time  $t = 2$  is a good example in Figure 10). Thus, the graphs in Figure 9

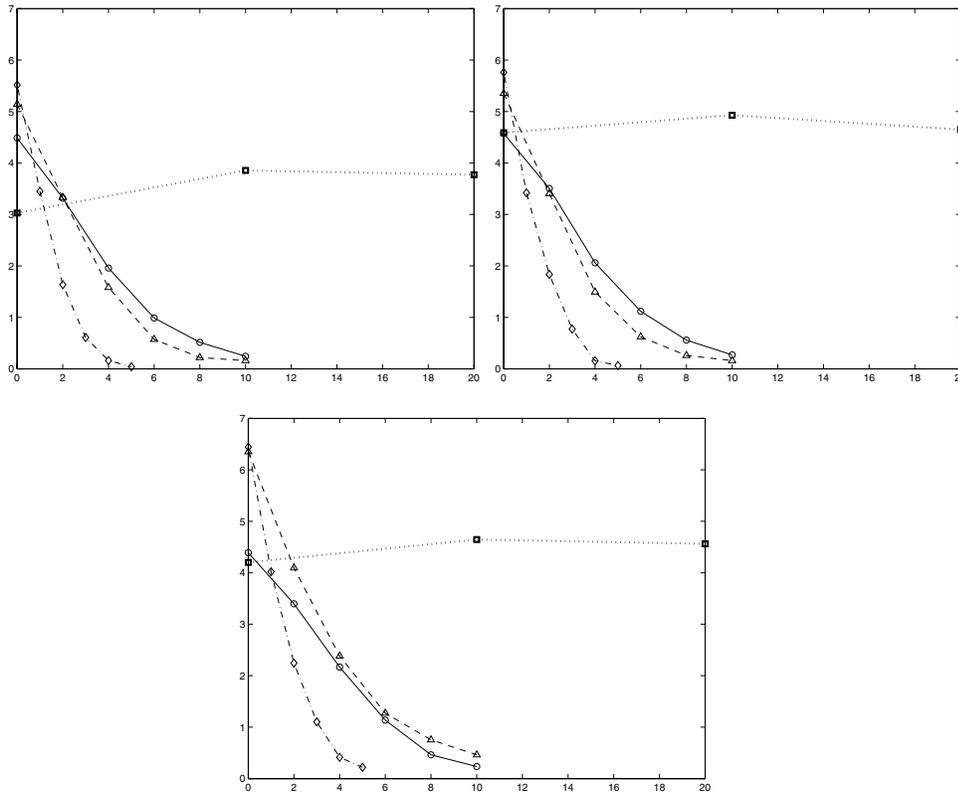


FIG. 9. The time series of the mean of predictive utility  $P(p_4^*, \Pi_c)$ , damped forced L96 model. Upper left— $k = 0$ , upper right— $k = 3$ , middle— $k = 8$ . Solid line with circles— $F = 0$ , dotted line with squares— $F = 4$ , dashed line with triangles— $F = 6$ , dot-dashed line with diamonds— $F = 8$ . There is a difference in time scales of decay:  $T \approx 5$  for  $F = 8$ ,  $T \approx 10$  for  $F = 0$ , and 6;  $F = 4$  does not decay.

need to be interpreted within the context of behavior like that in Figure 10.

Conventional wisdom says that usually the decay of time correlation functions governs the predictive utility, and it is true in most of the cases. However, in Figure 11 we plot the predictive utility for a typical ensemble prediction and the correlation function for the linearly stable Fourier modes  $k = 3$  and  $F = 4$ . It is evident that the correlation function (left panel) decays much more rapidly in time than the predictive utility (right panel) in this weakly chaotic regime. Thus, this case serves as a counterexample to the conventional wisdom in the applied literature. All the graphs in Figure 9 also provide a somewhat less powerful counterexample. From Figure 4 there is a large difference in the correlation time for mode  $k = 8$  compared with  $k = 0, 3$  for  $F = 6, 8$ . Yet from Figure 9 the mean predictive utility does not depend on the wavenumber for these cases.

**4.2. Non-Gaussian bimodal and skewed behavior for individual prediction ensembles.** Here we address and summarize the important issues in (A) and (B) in section 1 for the L96 and IL96 models. According to the results in section 2.1, for a single variable the information content in the ensemble prediction beyond a Gaussian estimator is bounded from below by the predictive utility  $P(p_4^*, p_G)$ , where

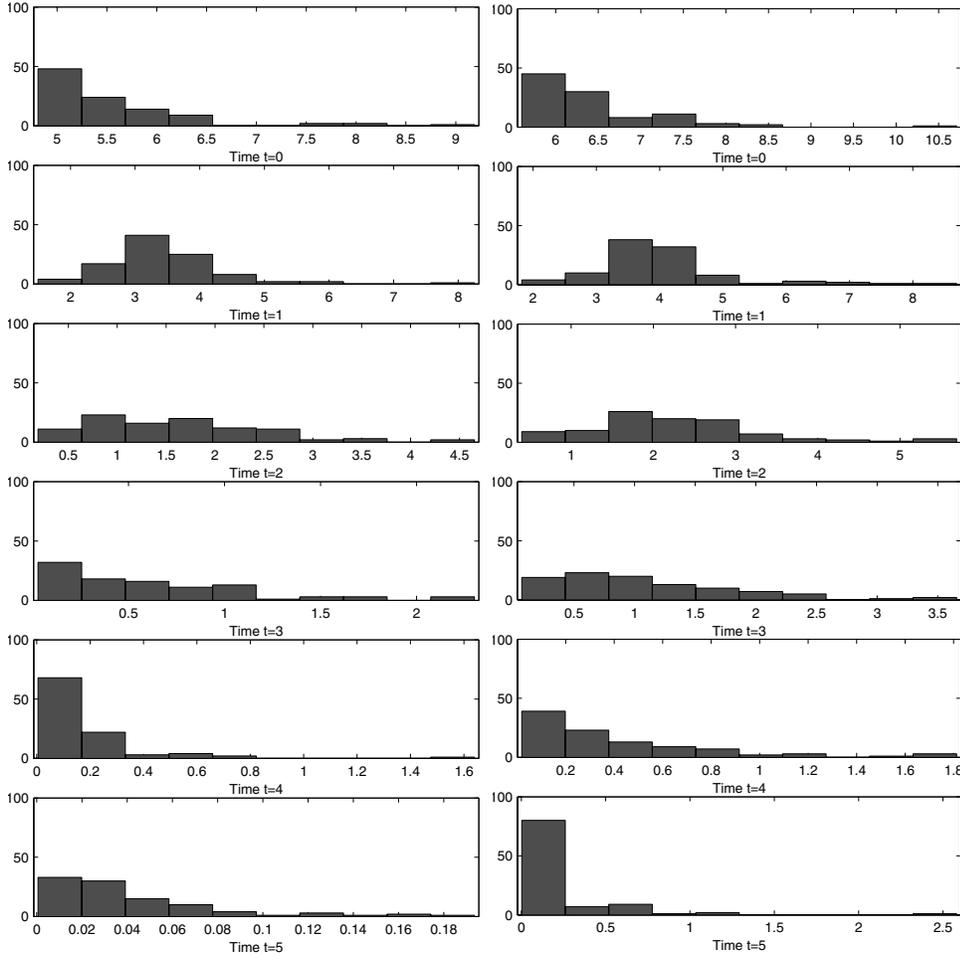


FIG. 10. The time series of histograms of the utility  $P(p_4^*, \Pi_c)$ , damped forced L96 model,  $F = 8$ , and modes  $k = 0$  and  $k = 8$ .

$p_4^*$  is the estimator utilizing four moments of the prediction. Large values of  $P(p_4^*, p_G)$  will reflect a large tendency toward a skewed and/or bimodal prediction ensemble.

In Figure 12 we plot the mean of the distribution of  $P(p_4^*, p_G)$  over the 100-member ensemble of ensembles for the four cases,  $F = 0, 4, 6, 8$  and the three Fourier EOF modes  $k = 0, 3, 8$ . Here the regime  $F = 4$  has the least amount of predictability, which is strong evidence that its ensembles are close to Gaussian, and follows from the fact that the initially generated Gaussian ensemble simply propagates undisturbed by essentially linear dynamics of  $F = 4$ . Regimes  $F = 0$  and  $F = 6$  are the most predictable, and the regime  $F = 8$  jumps back and forth between the other three. It is evident that the ensembles in cases  $F = 0$  and  $F = 6$  are more skewed and bimodal than those for  $F = 4$  and  $F = 8$ . Figure 1 with the associated tabulated values of  $P(p_4^*, p_G)$  from Table 1 is useful in interpreting highly non-Gaussian behavior in Figure 12. Recalling Table 1, the mean content of skewness and flatness in the ensembles for  $F = 0$  and  $F = 6$  can be estimated as  $\pm 0.1-0.3$  for skewness and about 2 for flatness

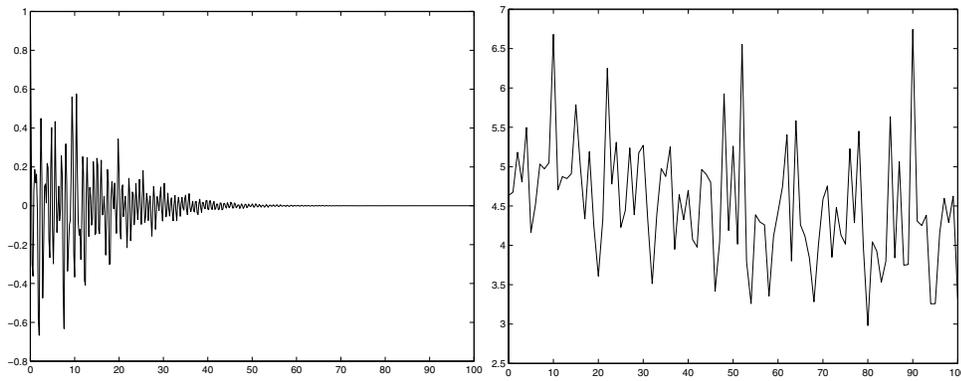


FIG. 11. *Left: The time correlation function for the damped forced L96 model, constant forcing  $F = 4$ , mode  $k = 3$ . Right: The typical predictive utility  $P(p_A^*, \Pi_c)$  time series for the same mode and forcing. Evidently, the time correlation function decays faster than the predictive utility (the latter exhibits very slow decay).*

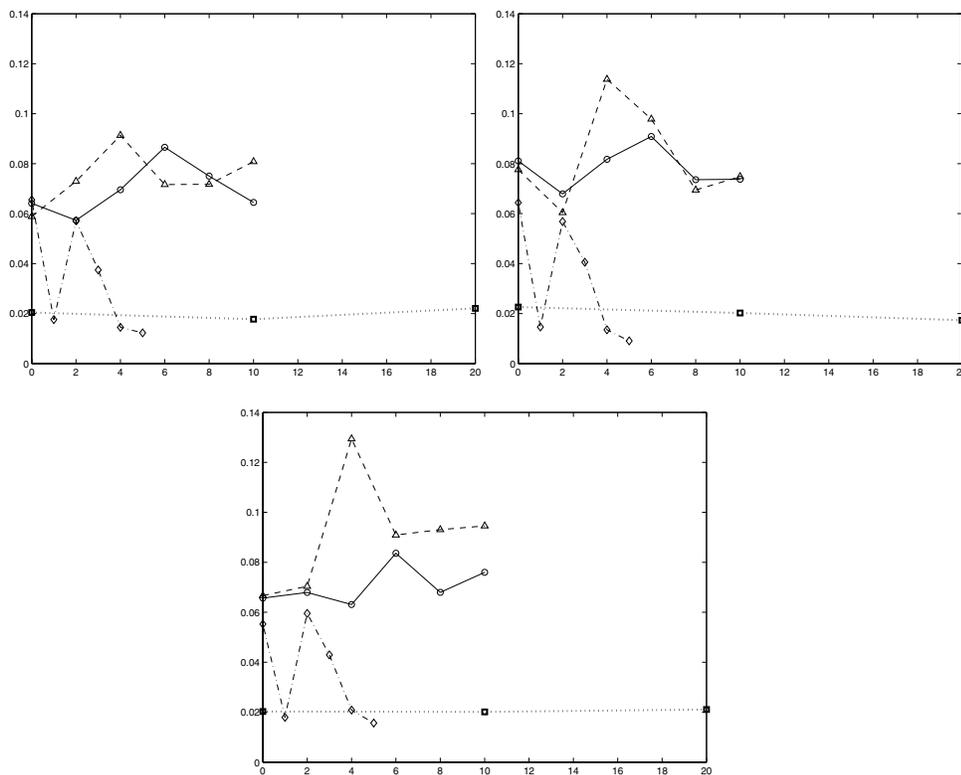


FIG. 12. *The time series of the mean of predictive utility  $P(p_A^*, p_G)$ , damped forced L96 model. Upper left— $k = 0$ , upper right— $k = 3$ , middle— $k = 8$ . Solid line with circles— $F = 0$ , dotted line with squares— $F = 4$ , dashed line with triangles— $F = 6$ , dot-dashed line with diamonds— $F = 8$ .*

(which means significant skewness and bimodality, according to Figure 1). The mean utility  $\bar{P}(p_4^*, p_G)$  exhibits systematic peaks at time  $t = 6$  for the regime  $F = 0$  (modes  $k = 0, 3, 8$ ) and at time  $t = 4$  for the regime  $F = 6$  (modes  $k = 0, 3, 8$ ). As in Figures 9 and 10 above, the mean plots show the average of a distribution which can have large spread. To illustrate this, in Figure 13 we plot the histograms of  $P(p_4^*, p_G)$  for selected times for ensemble prediction of the change in the mean climate,  $k = 0$  for  $F = 6$  and  $F = 8$ . The peak of mean utility at the time  $t = 4$  for the regime  $F = 6$  is confirmed by especially big spread on the histogram for the same time snapshot in Figure 13, where the utility in rare cases achieves 0.9 (90% of its mean value). For the regime  $F = 8$  there is a distinct decrease in statistical spread scale in Figure 13 as well as for its mean in Figure 12. The quantitative use of this measure of uncertainty utilizing  $P(p_4^*, p_G)$  is illustrated in Figures 14 and 15 with a typical ensemble prediction for the mean climate mode  $k = 0$  and  $F = 6, 8$ . The bimodality and skewness in the histogram of the 100-member ensemble prediction is typically estimated quite accurately by the four moment estimator  $p_4^*$ . Good examples are the plots for the regimes  $F = 6, t = 4$  and  $F = 8, t = 3$ . In the first example the four moment estimator follows the two peaks and a valley in between very well, whereas the Gaussian estimator bluntly produces a single peak where there is no actual trace of it. In the second example the skewness, as well as bimodality, is reproduced by the four moment estimator. The Gaussian estimator in the second example is as bad as in the first: that is, a single peak with no trace of observed skewness is exactly where it is least expected. These non-Gaussian ensembles in Figure 15 are clearly predicted by the peaks of utility in Figure 14 at corresponding times.

Next we consider quantifying the lack of information in the non-Gaussian climate compared with an ensemble prediction for a single spatial node for L96 with  $F = 8$ . The plot of the predictive utility estimator  $P(p_4^*, \Pi_c)$  for a typical ensemble prediction is given in Figure 16. Here  $\Pi_c$  has been estimated from the first four moments of the climate distribution through (2.21) as in Figure 7. The Gaussian estimator  $P(p_G, \Pi_c)$  is also plotted on this graph to compare the difference between the two estimators' information content. The difference between the two is rather small, which suggests that the ensembles are close to Gaussian in this case.

**4.3. The relative importance of signal and dispersion for atypical events where the climate lacks information.** Here, following [12] and [8], we utilize the generalized signal/dispersion decomposition of the relative entropy,  $P(p_4^*, \Pi_c)$ , to quantify the important issue in (E) in section 1, i.e., how to assess the atypical rare events where the climate lacks large information relative to the ensemble prediction at a given prediction time. For a given dynamical system such as TBH, L96, or IL96, does the signal or dispersion or both control the lack of information in the climate in these rare events? Recall that  $\Pi_c$  is the climate distribution estimated from four moments and  $p_4^*$  is the estimator for the ensemble prediction utilizing four moments. Here, for a given dynamical system, the rare ensemble predictions with large utility relative to the climate are given by those ensembles where  $P(p_4^*, \Pi_c) \geq \alpha \bar{P}(p_4^*, \Pi_c)$ , where the mean of the utility,  $\bar{P}$ , for the ensemble of ensembles has been presented in Figure 9 for illustrative examples. The number  $\alpha > 1$  is a threshold parameter; here we set  $\alpha = 1.5$  to define the atypical "rare" events. Then, given each of these rare events, we use the generalized signal/dispersion decomposition to measure the percentage of the utility given the signal and dispersion. A histogram for the percentage of the utility given by the signal or dispersion is generated by bin counting over a given time interval. If this histogram shows significant dominance of high percentage events by the signal

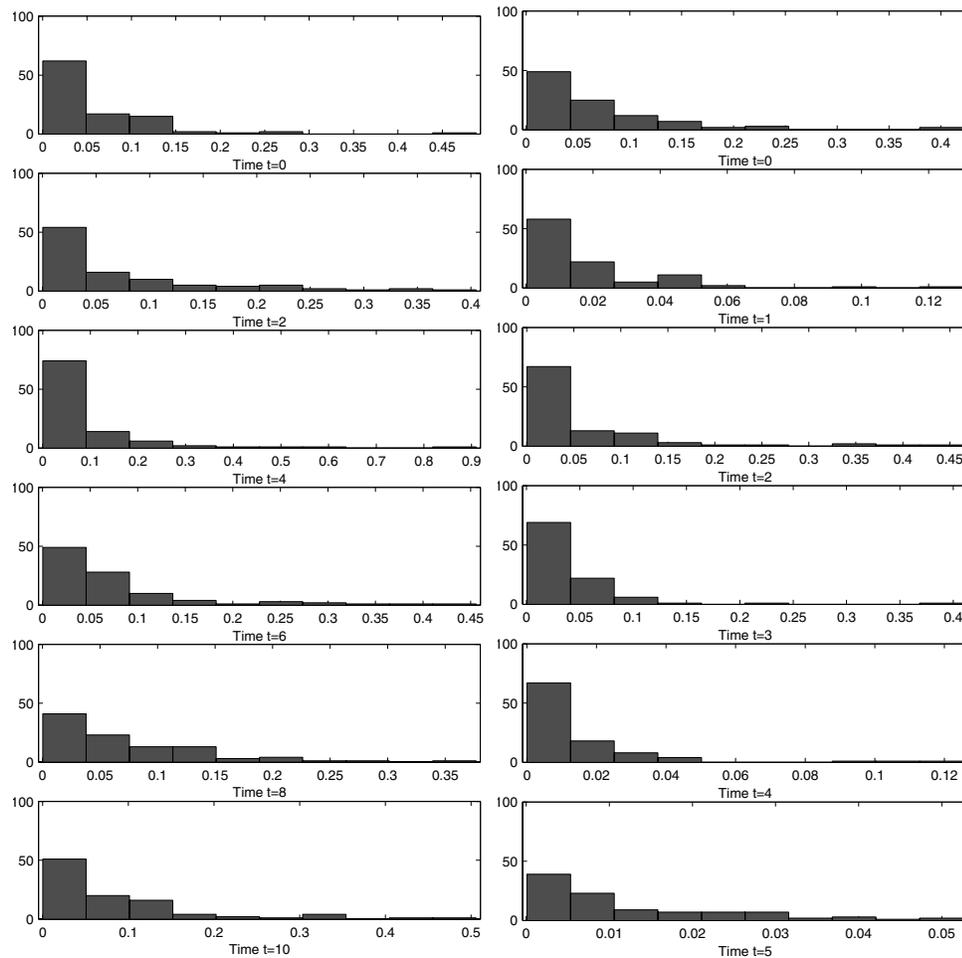


FIG. 13. The time series of histograms of  $P(p_4^*, p_G)$ , damped forced L96 model, climate mode  $k = 0$ , forcing regimes  $F = 6$  (left column) and  $F = 8$  (right column). Note the systematic difference in horizontal scale—the histograms for  $F = 6$  (left column) are wider than those for  $F = 8$  (right column).

(dispersion), then the signal (dispersion) controls the predictive utility for these rare events over that time interval. This use of the signal/dispersion decomposition is illustrated first for Fourier EOF mode  $k = 5$  for the TBH model. The above procedure was applied to the two time intervals  $0 \leq t \leq 2$  and  $3 \leq t \leq 5$  to represent short and long times for the ensemble prediction with significant utility. The histograms for the signal and dispersion for the two time intervals for this case are depicted in Figure 17. For the latter time interval it is evident that the signal controls the utility for TBH for the rare events where the climate has a significant lack of information, while there is equal importance of both signal and dispersion over the former time interval. This result is a different confirmation of the trend first reported in [8] that the signal controls variations in this predictive utility for TBH. In Figures 18 and 19, we apply the same test as described above for TBH to L96 with  $F = 8$  for the climate mode  $k = 0$ , and the unstable mode  $k = 8$ , respectively. The cases where the contribution

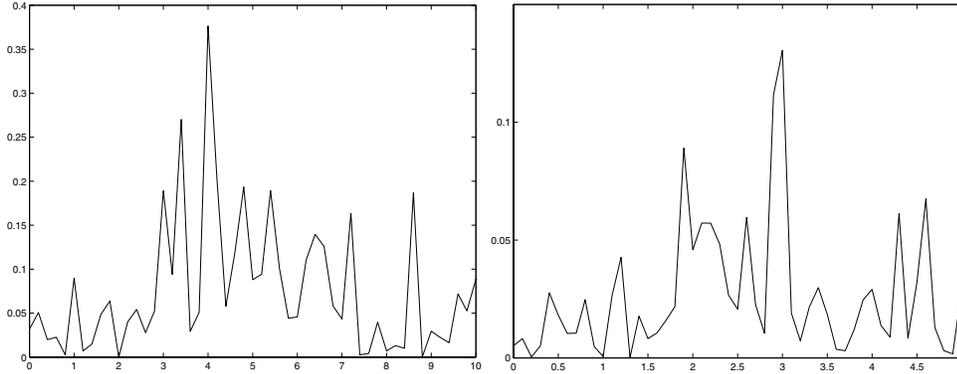


FIG. 14. The typical  $P(p_4^*, p_G)$  profile for the mode  $k = 0$ , damped forced L96 model, mode  $k = 0$ . Left—forcing  $F = 6$ , right—forcing  $F = 8$ . The peak at  $t = 4$  for  $F = 6$  (left panel) is confirmed by the non-Gaussian bimodal ensemble in Figure 15, left column. The peak at  $t = 3$  for  $F = 8$  (right panel) is confirmed by the non-Gaussian bimodal ensemble in Figure 15, right column.

of the generalized signal clearly dominates that of the dispersion are  $k = 0$  and  $k = 8$ , both for the regime  $F = 8$  and the second time interval  $3 \leq t \leq 5$ , when the mean utility is small (see Figure 9). The general trend here is that the dispersion is more important at short times, while the signal is more important at later times.

The same test, illustrated above and applied to the climate models TBH and L96 for  $F = 0, 6, 8$  in section 3 yields the results summarized below.

*Generalized signal dominates in rare events of high utility.*

- IL96:
  - mode  $k = 3$ ,  $0 \leq t \leq 4$  (high mean utility),
  - mode  $k = 8$ ,  $0 \leq t \leq 4$  (high mean utility),
  - mode  $k = 8$ ,  $6 \leq t \leq 10$  (low mean utility).
- L96,  $F = 6$ :
  - no such events, dispersion always dominates.
- L96,  $F = 8$ :
  - mode  $k = 0$ ,  $3 \leq t \leq 5$  (low mean utility),
  - mode  $k = 3$ ,  $3 \leq t \leq 5$  (low mean utility),
  - mode  $k = 8$ ,  $3 \leq t \leq 5$  (low mean utility).
- TBH, statistically irrelevant  $H = 0$ :
  - mode  $k = 1$ ,  $3 \leq t \leq 5$  (low mean utility),
  - mode  $k = 5$ ,  $3 \leq t \leq 5$  (low mean utility),
  - mode  $k = 10$ ,  $0 \leq t \leq 2$  (high mean utility),
  - mode  $k = 10$ ,  $3 \leq t \leq 5$  (low mean utility).
- TBH, statistically relevant  $H = 0.01$ :
  - mode  $k = 1$ ,  $3 \leq t \leq 5$  (low mean utility),
  - mode  $k = 5$ ,  $3 \leq t \leq 5$  (low mean utility),
  - mode  $k = 10$ ,  $0 \leq t \leq 2$  (high mean utility),
  - mode  $k = 10$ ,  $3 \leq t \leq 5$  (low mean utility).

*Remark.* The signal/dispersion decomposition is also useful for other predictability issues as discussed in (A) and (B) in section 1. For example, one can use this decomposition to measure the lack of information in small member ensemble predictions  $R \cong 100$ , compared with perfect ensemble predictions  $R \gg 100$ , in idealized

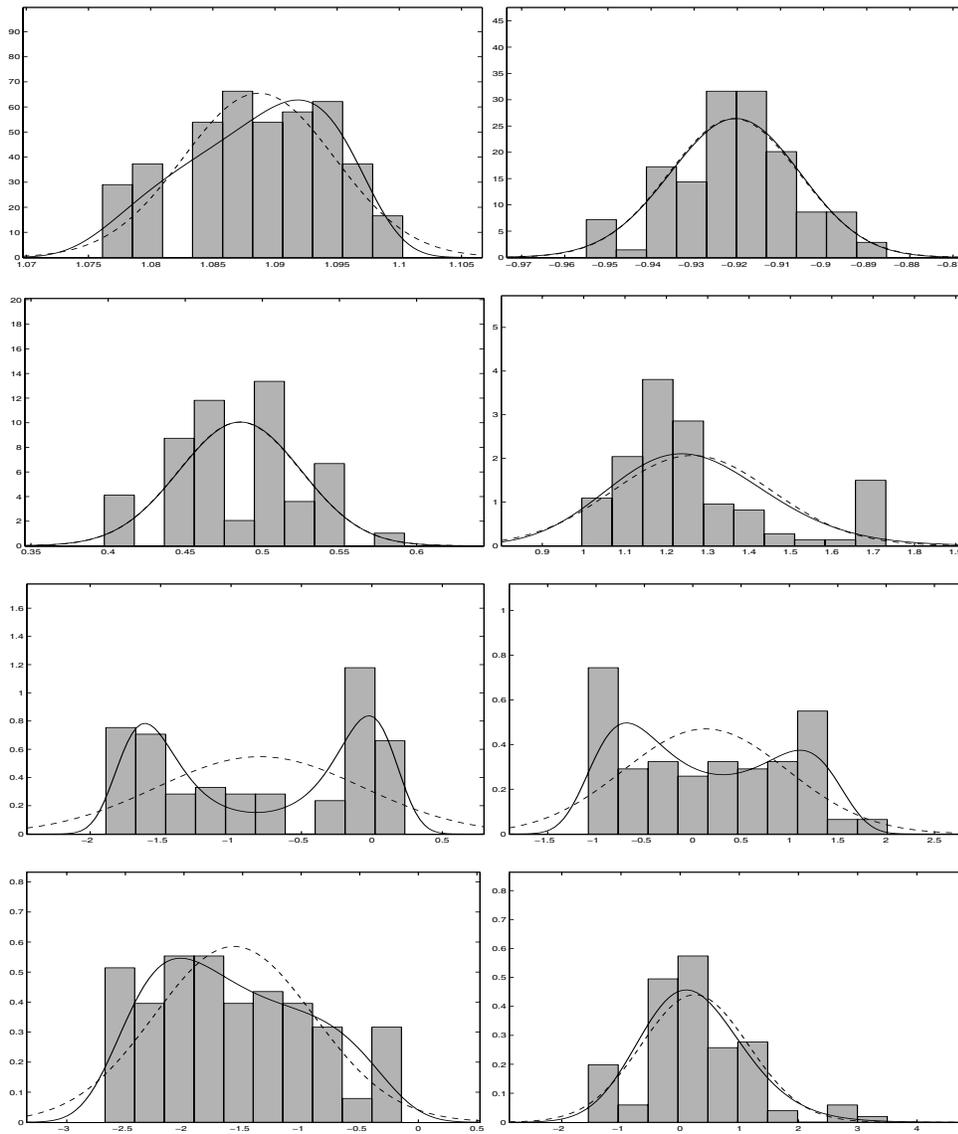


FIG. 15. The typical ensemble histograms fitted by the four-constraint estimator (solid lines) and Gaussian estimator (dashed lines), mode  $k = 0$ , damped forced L96 model with constant forcing  $F = 6$  (left column), and  $F = 8$  (right column). Left column, up  $\rightarrow$  down: Time  $t = 0, 2, 4, 6$ . Right column, up  $\rightarrow$  down: Time  $t = 1, 2, 3, 4$ . The non-Gaussian bimodal ensemble in the left column ( $F = 6$ ) for  $t = 4$  is predicted as a peak in Figure 14, left panel. The non-Gaussian bimodal ensemble in the right column ( $F = 8$ ) for  $t = 3$  is predicted as a peak in Figure 14, right panel.

systems, to decide whether the signal or the dispersion is the principal source of the lack of information in low member ensembles for a given system.

**4.4. Hierarchical estimators for utility in several variables.** Here we utilize the hierarchical approach from section 2.2 to build the improved predictability estimators from practical predictability components. Tables 3, 4, and 5 in section 3.4

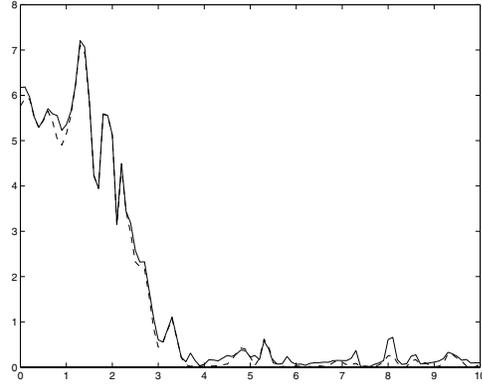


FIG. 16. The predictive utilities  $P(p_4^*, \Pi_c)$  (solid line) and  $P(p_G, \Pi_c)$  (dashed line), damped forced L96 model,  $F = 8$ , single physical space node.

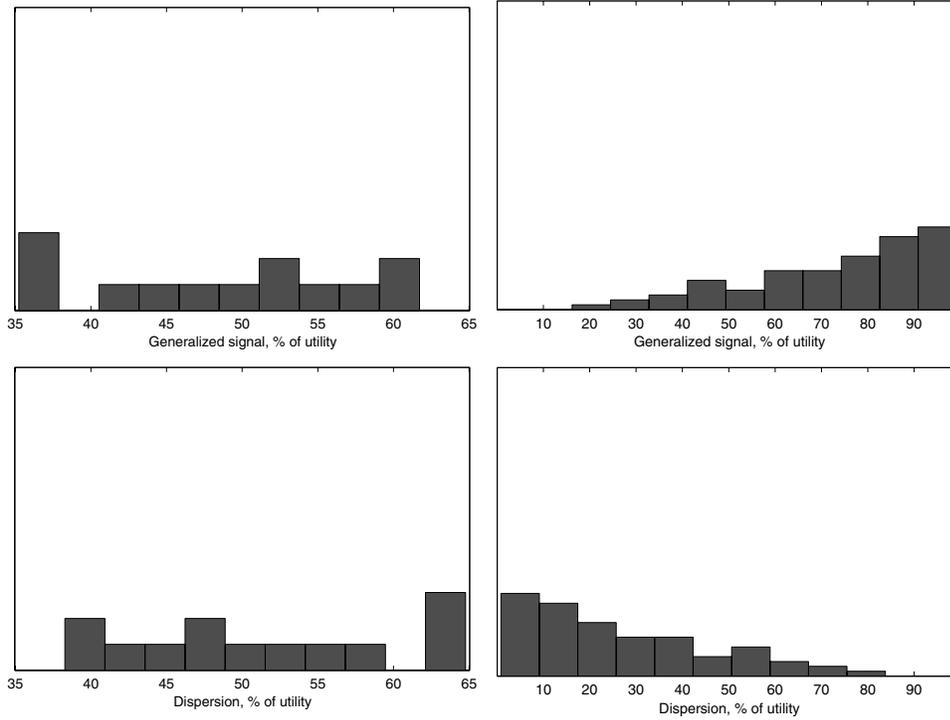


FIG. 17. The statistical significance of the generalized signal and dispersion for rare events of high total utility,  $0 \leq t \leq 2$  (left column),  $3 \leq t \leq 5$  (right column), TBH, statistically relevant Hamiltonian  $H = 0.01$ , mode  $k = 5$ .

provide sufficient evidence of climate Gaussianity up to the fourth order moments. As a first approximation, we assume this in the discussion presented next. The relation

$$(4.1) \quad P(p_4^*(\vec{Z}), \Pi_G(\vec{Z})) \geq P(p_G(\vec{Z}), \Pi_G(\vec{Z})) \geq P(p_G(\vec{X}), \Pi_G(\vec{X})),$$

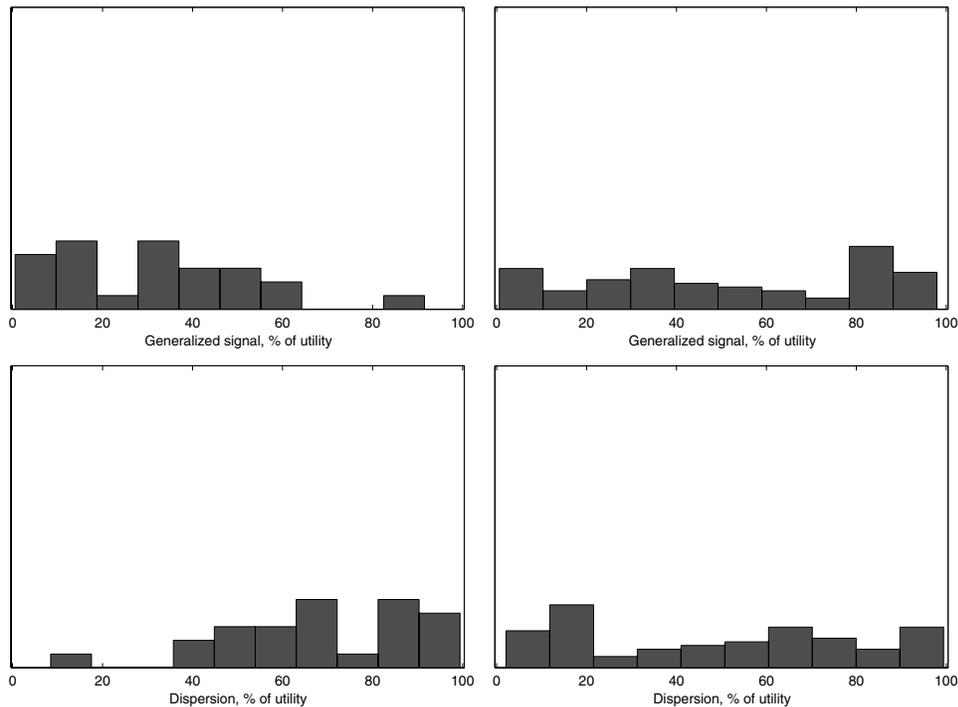


FIG. 18. *The statistical significance of the generalized signal and dispersion for rare events of high total utility,  $0 \leq t \leq 2$  (left column),  $3 \leq t \leq 5$  (right column), L96,  $F = 8$ , mode  $k = 0$ .*

where  $\vec{X}$  are the Fourier EOF modes and  $\vec{Z}$  are the principal predictability components given by (2.25), obviously follows from (2.30) and the fact that the prediction covariance matrix is diagonal in  $\vec{Z}$ -basis, which is given in (2.24). It is important to know the advantage of an improved predictability estimator over its predecessor for evaluating the minimum error provided by the latter in estimating the predictability. In Figure 20 we show the three successive predictability estimators from (4.1) for the L96 model with  $F = 8$  and for the subspace of Fourier EOF modes 0–9:  $P(p_G(\vec{X}), \Pi_G(\vec{X}))$  (shown as dash-dotted line),  $P(p_G(\vec{Z}), \Pi_G(\vec{Z}))$  (shown as dashed line), and  $P(p_4^*(\vec{Z}), \Pi_G(\vec{Z}))$  (shown as solid line). Figure 20 shows that the relative error between  $P(p_G(\vec{Z}), \Pi_G(\vec{Z}))$  and  $P(p_G(\vec{X}), \Pi_G(\vec{X}))$  grows roughly linearly with time, approximately 10% error per time unit, whereas the difference between  $P(p_4^*(\vec{Z}), \Pi_G(\vec{Z}))$  and  $P(p_G(\vec{Z}), \Pi_G(\vec{Z}))$  is not so significant (no more than 5% until  $t = 5$ ). After time  $t = 5$  the measure  $P(p_G(\vec{X}), \Pi_G(\vec{X}))$  is not useful, because it shows nearly zero utility, whereas the “actual” utility  $P(p, \Pi_G)$  should be at least 2–3 (no less than  $P(p_4^*(\vec{Z}), \Pi_G(\vec{Z}))$  anyway). According to the trends shown on Figure 20, the advantage of  $P(p_G(\vec{Z}), \Pi_G(\vec{Z}))$  over  $P(p_G(\vec{X}), \Pi_G(\vec{X}))$  is great and highlighted by the fact that it is achieved through modest additional cost of covariance matrix diagonalization (no optimization problem has to be solved for both  $P(p_G(\vec{Z}), \Pi_G(\vec{Z}))$  and  $P(p_G(\vec{X}), \Pi_G(\vec{X}))$ ). On the other hand, in this example the additional diagonal third and fourth moments in  $P(p_4^*(\vec{Z}), \Pi_G(\vec{Z}))$  only improve the estimate over  $P(p_G(\vec{Z}), \Pi_G(\vec{Z}))$  by between 15 and 20% for times longer than  $t = 5$ .

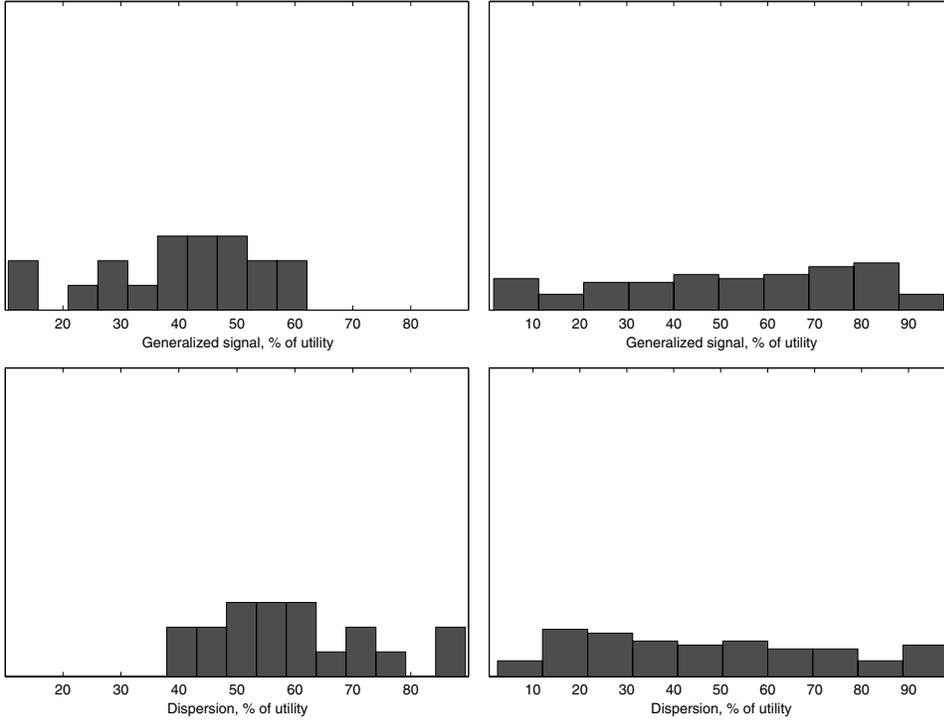


FIG. 19. The statistical significance of the generalized signal and dispersion for rare events of high total utility,  $0 \leq t \leq 2$  (left column),  $3 \leq t \leq 5$  (right column),  $L96$ ,  $F = 8$ , mode  $k = 8$ .

On Figure 21 we demonstrate the improved non-Gaussian estimator from (2.31) and (2.32), which takes into account joint two-dimensional PDFs of a given phase space. Two sampled phase spaces are studied: the three-dimensional space of principal predictability components  $\{Z_1, Z_2, Z_3\}$  defined in (2.25) for the Fourier EOF modes  $k = 0, 3, 8$ , and one two-dimensional subspace spanned by the pair  $\{Z_1, Z_2\}$ . Note that for a two-dimensional phase space the lower bound in (2.32) becomes an identity, and so the only lack of information comes from the fact that the set of measured moments in (2.33) is finite ( $L = 2$ ). The same optimization routine was used here and in Figure 8, where it was able to recreate a significantly skewed two-dimensional climate PDF for the Fourier EOF modes  $k = 0, 8$  in the forcing regime  $F = 6$ . Here we take a typical ensemble for the regime  $F = 8$  (snapshots are shown on Figure 15). The picture in the upper left corner of Figure 21 demonstrates the difference between the straightforward sum of one-dimensional estimators  $\sum_i^2 P(p_4^*(Z_i), p_G(Z_i))$  (solid lines) and improved two-dimensional estimators  $P(p_4^*(Z_i, Z_j), p_4^*(Z_i)p_4^*(Z_j)) + P(p_4^*(Z_j, Z_k), p_4^*(Z_j)p_4^*(Z_k)) + \sum_i^2 P(p_4^*(Z_i), p_G(Z_i))$  (dashed lines) for the subspace  $\{Z_1, Z_2\}$ . One can see a lot of improvement over the sum of one-dimensional estimators (dashed line over solid line), namely, the predictive utility of an improved two-dimensional estimator is on average twice that of the direct sum  $\sum_i^2 P(p_4^*(Z_i), p_G(Z_i))$  of one-dimensional estimators. The improved non-Gaussian estimators for the three-dimensional phase space are shown in the upper right, lower left, and lower right corners of Figure 21. Here we emphasize that the improved estimator in (2.32) is not

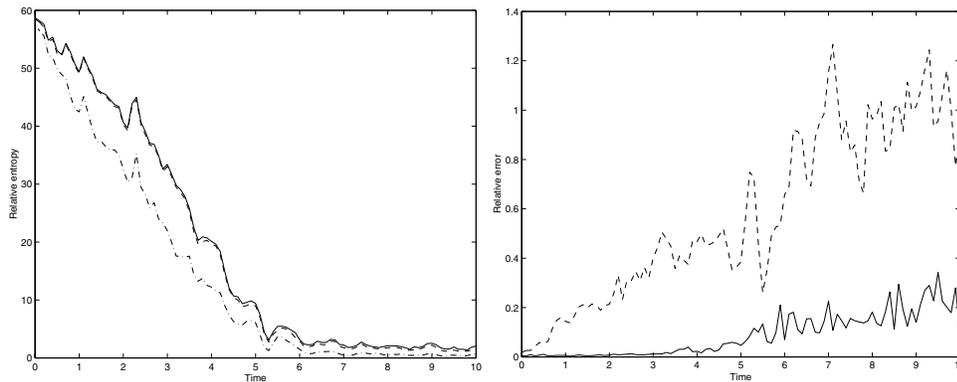


FIG. 20. Left:  $P(p_4^*(\vec{Z}), \Pi_G(\vec{Z}))$  (solid line),  $P(p_G(\vec{Z}), \Pi_G(\vec{Z}))$  (dashed line),  $P(p_G(\vec{X}), \Pi_G(\vec{X}))$  (dash-dotted line). Right: The relative errors between  $P(p_4^*(\vec{Z}), \Pi_G(\vec{Z}))$  and  $P(p_G(\vec{Z}), \Pi_G(\vec{Z}))$  (solid line), and between  $P(p_G(\vec{Z}), \Pi_G(\vec{Z}))$  and  $P(p_G(\vec{X}), \Pi_G(\vec{X}))$  (dashed line). L96 model with  $F = 8$ , Fourier EOF modes  $0 \leq k \leq 9$ .

invariant under general re-ordering of phase space variables, and therefore all three possible permutations of triads  $\{Z_1, Z_2, Z_3\}$  are shown as dashed lines on the plots in the upper right, lower left, and lower right corners of Figure 21. Solid lines denote the direct sum of one-dimensional estimators  $\sum_i^3 P(p_4^*(Z_i), p_G(Z_i))$ , which is, in accordance with the theory in (2.31) and (2.32), always equal to or below any of the other three. One can see that different estimators are more important for different times, such as the dashed line on the lower left plot which is above the rest at the time  $t = 3$ , the dashed line on the upper right plot—at the time  $t = 5$ , and the dashed line on the lower right plot—at the time  $t = 6.25$ . Since all dashed lines on the plots are different lower bounds on the same information content, clearly the best estimate is the maximum of the three dashed curves in the appropriate three panels in Figure 21. However, while in this low-dimensional phase space all possible permutations of coordinate triads can be directly checked (there are only three of them, all shown in Figure 21), such an approach is not computationally feasible in the moderate- and high-dimensional phase spaces. Instead, a suitable ordering strategy, based on individual utility of each phase space variable, has yet to be developed, allowing one to choose the best estimate with high probability without directly evaluating it.

One can interpret  $\vec{Z}$ -variables as the *principal predictability directions* in the Fourier EOF space, arranged in descending order by their predictability at a given time. Figure 22 shows the time evolution of principal predictability directions in the  $P(p_4^*(\vec{Z}), \Pi_G(\vec{Z}))$  metric for the subspace of the Fourier EOF modes  $k = 0, 3, 8$ , forcing  $F = 8$ . The upper row of pictures in Figure 22 shows the time evolution of the direction with the best predictability. Most of the time the direction with best predictability is aligned with the Fourier mode  $k = 8$ , with the exception for the time  $t = 2$ , when  $k = 3$  is dominant. Statistically, one should expect the linearly unstable mode  $k = 8$  to be predictable more often than the others because of the strong influence of the linear, less chaotic part of the dynamics. The bottom row of Figure 22 shows the time evolution of the direction with the least predictability. The trajectory of the least predictability direction lies entirely in the  $\{k_x = 0, k_y = 3\}$  plane, which is essentially a consequence of the fact that the best predictability direction is aligned with  $k = 8$  (perpendicular to the  $\{k_x = 0, k_y = 3\}$  plane). The spread of the least

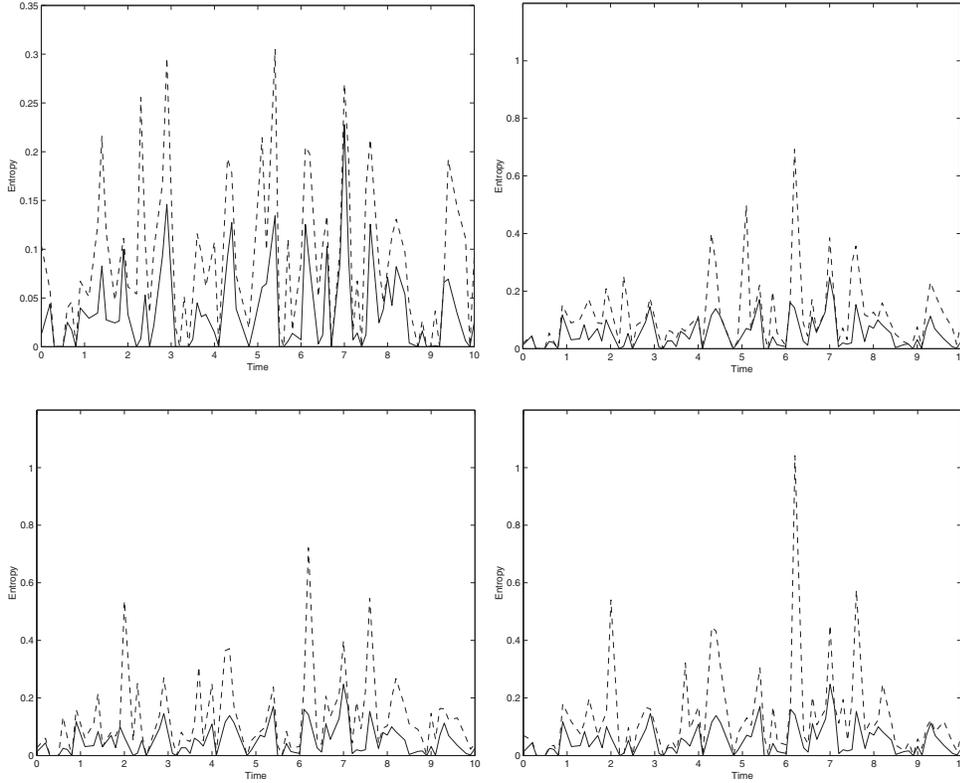


FIG. 21. Different non-Gaussian estimators for the phase space of the three principal predictability components  $\{Z_1, Z_2, Z_3\}$  for the Fourier EOF modes  $k = 0, 3, 8$ , L96 model, forcing regime  $F = 8$ . Upper left:  $P(p_4^*(Z_1)p_4^*(Z_2), p_G(Z_1)p_G(Z_2))$  (solid line),  $P(p_4^*(Z_1, Z_2), p_G(Z_1)p_G(Z_2))$  (dashed line). Upper right:  $\sum_j^3 P(p_4^*(Z_j), p_G(Z_j))$  (solid line),  $P(p_4^*(Z_1, Z_2), p_G(Z_1)p_G(Z_2)) + P(p_4^*(Z_1, Z_3), p_G(Z_1)p_G(Z_3)) + \sum_j^3 P(p_4^*(Z_j), p_G(Z_j))$  (dashed line). Lower left:  $\sum_j^3 P(p_4^*(Z_j), p_G(Z_j))$  (solid line),  $P(p_4^*(Z_1, Z_2), p_G(Z_1)p_G(Z_2)) + P(p_4^*(Z_2, Z_3), p_G(Z_2)p_G(Z_3)) + \sum_j^3 P(p_4^*(Z_j), p_G(Z_j))$  (dashed line). Lower right:  $\sum_j^3 P(p_4^*(Z_j), p_G(Z_j))$  (solid line),  $P(p_4^*(Z_1, Z_3), p_G(Z_1)p_G(Z_3)) + P(p_4^*(Z_2, Z_3), p_G(Z_2)p_G(Z_3)) + \sum_j^3 P(p_4^*(Z_j), p_G(Z_j))$  (dashed line). Note the difference in vertical scale between the upper left plot (0–0.35) and the other three (0–1.2).

predictability direction is about uniform in the  $\{k_x = 0, k_y = 3\}$  plane, i.e., directions  $k = 0$  and  $k = 3$  are statistically equally unpredictable. The intermediate predictability direction is roughly uniformly scattered in time between all three Fourier modes without any particular trends. It is remarkable that the direction of best predictability is aligned with the linearly unstable Fourier mode. The reason for this is that the unstable linear part of dynamics in the mode  $k = 8$  contributes more ordered dynamical behavior which results in good utility, whereas nonlinear interactions dominate in the modes  $k = 0$  and  $k = 3$ , bringing more chaos to the dynamics of those modes and reducing their predictability.

**5. Concluding discussion.** A mathematically rigorous, computationally feasible hierarchical strategy for estimating the lack of information in one probability distribution relative to another for many degrees of freedom in complex systems has

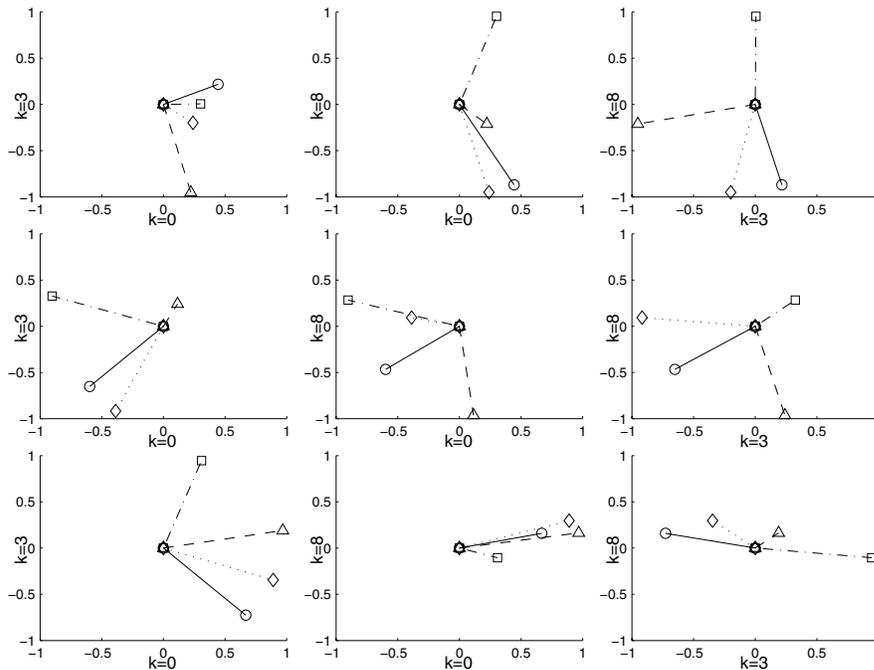


FIG. 22. The time evolution of the three principal predictability directions of the subspace  $k = 0, 3, 8$  in the forcing regime  $F = 8$ , shown as projections on the three planes (from left to right):  $\{k_x = 0, k_y = 3\}$ ,  $\{k_x = 0, k_y = 8\}$ , and  $\{k_x = 3, k_y = 8\}$ . Upper row—direction of the best predictability, bottom row—direction of least predictability, middle row—direction perpendicular to both. The predictability is measured in the  $P(p_1^*(Z), \Pi_G(Z))$  metric. Each plot contains four vectors showing time evolution:  $t = 1$  (solid line with circles),  $t = 2$  (dashed line with triangles),  $t = 3$  (dash-dotted line with squares), and  $t = 4$  (dotted line with diamonds).

been developed here. Many facets of this strategy have been tested in the context of ensemble prediction with small ensemble sizes in two “toy” climate models, the Galerkin truncated Burgers–Hopf model and the L96 model in a wide range of parameter regimes. Additional more realistic applications of those ideas are under way for quantifying the uncertainty in geostrophic turbulent flows for both the atmosphere and ocean [9] and for the low frequency variability of the atmosphere [2]. In particular, for a perfect predictability scenario, statistical estimators for the lack of information due to finite large sample size effects in ensemble prediction are quantified for the moment estimator strategies used in the present paper.

**Acknowledgments.** The authors thank Richard Kleeman and Tapio Schneider for stimulating discussions which influenced the work.

#### REFERENCES

- [1] R. ABRAMOV, G. KOVAČIČ, AND A. MAJDA, *Hamiltonian structure and statistically relevant conserved quantities for the truncated Burgers-Hopf equation*, Comm. Pure Appl. Math., 56 (2003), pp. 1–46.
- [2] R. ABRAMOV, A. MAJDA, AND R. KLEEMAN, *Information theory and predictability for low frequency variability*, J. Atmos. Sci., to appear.

- [3] R. ABRAMOV AND A. MAJDA, *Discrete approximations with additional conserved quantities: Deterministic and statistical behavior*, Methods Appl. Anal., 10 (2003), pp. 151–190.
- [4] J. ANDERSON AND W. STERN, *Evaluating the potential predictive utility of ensemble forecasts*, J. Clim., 9 (1996), pp. 260–269.
- [5] D. CAI, K. HAVEN, AND A. MAJDA, *Quantifying predictability in a simple model with complex features, I: The perfect predictability scenario*, Stochastics and Dynamics (2003), in press.
- [6] T. COVER AND J. THOMAS, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
- [7] E. KALNAY, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, New York, 2003.
- [8] R. KLEEMAN, A. MAJDA, AND I. TIMOFEYEV, *Quantifying predictability in a model with statistical features of the atmosphere*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 15291–15296.
- [9] R. KLEEMAN AND A. MAJDA, *Predictability in a model for geophysical turbulence*, J. Atmos. Sci., submitted.
- [10] R. KLEEMAN AND A. MOORE, *A theory for the limitation of ENSO predictability due to stochastic atmospheric transients*, J. Atmos. Sci., 54 (1997), pp. 753–767.
- [11] R. KLEEMAN AND A. MOORE, *A new method for determining the reliability of dynamical ENSO predictions*, Mon. Wea. Rev., 127 (1999), pp. 694–705.
- [12] R. KLEEMAN, *Measuring dynamical prediction utility using relative entropy*, J. Atmos. Sci., 59 (2002), pp. 2057–2072.
- [13] A. LASOTA AND M. MACKAY, *Chaos, Fractals, and Noise*, 2nd ed., Springer-Verlag, New York, 1994.
- [14] E. LORENZ AND K. EMANUEL, *Optimal sites for supplementary weather observations*, J. Atmos. Sci., 55 (1998), pp. 399–414.
- [15] E. LORENZ, *Predictability: A problem partly solved*, in Proceedings of the Seminar on Predictability, Shinfield Park, Reading, UK, 1996, ECMWF.
- [16] A. MAJDA, R. KLEEMAN, AND D. CAI, *A mathematical framework for quantifying predictability through relative entropy*, Methods Appl. Anal., 9 (2002), pp. 425–444.
- [17] A. MAJDA AND I. TIMOFEYEV, *Remarkable statistical behavior for truncated Burgers-Hopf dynamics*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 12413–12417.
- [18] A. MAJDA AND I. TIMOFEYEV, *Statistical mechanics for truncations of the Burgers-Hopf equation: A model for intrinsic stochastic behavior with scaling*, Milan J. Math., 70 (2002), pp. 39–96.
- [19] A. MAJDA AND X. WANG, *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*, Cambridge University Press, New York, in press.
- [20] A. MAJDA, *Introduction to PDEs and Waves for the Atmosphere and Ocean*, Courant Lect. Notes in Math. 9, AMS, Providence, RI, 2003.
- [21] T. PALMER, *Predicting uncertainty in forecasts of weather and climate*, Rep. Prog. Phys., 63 (2000), pp. 71–116.
- [22] M. ROULSTON AND L. SMITH, *Evaluating probabilistic forecasts using information theory*, Mon. Wea. Rev., 130 (2002), pp. 1653–1660.
- [23] T. SCHNEIDER AND S. GRIFFIES, *A conceptual framework for predictability studies*, J. Clim., 12 (1999), pp. 3133–3155.
- [24] Z. TOTH AND E. KALNAY, *Ensemble forecasting at NMC: The generation of perturbations*, Bull. Amer. Meteor. Soc., 74 (1993), pp. 2317–2330.
- [25] Z. TOTH, *Circulation patterns in phase space: A multinormal distribution?*, Mon. Wea. Rev., 119 (1991), pp. 1501–1511.
- [26] Z. WU, G. PHILLIPS, R. TAPIA, AND Y. ZHANG, *A fast Newton algorithm for entropy maximization in phase determination*, SIAM Rev., 43 (2001), pp. 623–642.