An information-theoretic framework for improving imperfect predictions via Multi Model Ensemble forecasts

M. Branicki and A. J. Majda

Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, USA

Abstract

This work focuses on elucidating issues related to an increasingly common technique of *Multi Model Ensemble* (MME) forecasting. This approach is aimed at improving the statistical accuracy of imperfect time-dependent predictions by combining information from a collection of reduced-order dynamical models. Despite some operational evidence in support of the use of the MME strategy for mitigating the prediction error, the mathematical framework justifying this approach has been lacking. Here, this problem is considered within a probabilistic/stochastic framework which exploits tools from information theory to derive a simple criterion for an improvement of probabilistic predictions within the MME framework relative to single model predictions. The emphasis is on a systematic understanding of the benefits and limitations associated with the MME approach, on uncertainty quantification, and on the development of practical design principles for constructing an MME with an improved predictive performance. The condition for prediction improvement via the MME approach stems from convexity of the *relative entropy* which is used here as a measure of the lack of information in the imperfect models relative to resolved characteristics of the truth dynamics. It is also shown how practical guidelines for MME prediction improvement can be implemented in the context of forced response predictions from equilibrium with the help of the linear response theory utilizing the fluctuation-dissipation formulas at the unperturbed equilibrium. The general theoretical results are illustrated using exactly solvable non-Gaussian test models.

1 Introduction

Dynamical prediction of complex multi-scale systems based on imperfect models and spatio-temporally sparse observations of the truth dynamics is a notoriously difficult problem which is, nevertheless, essential in many applications such as climate-atmosphere science [14, 61], materials science [10, 30], neuroscience [62], or systems biology and biochemistry [57, 66, 12, 29]. Due to the high-dimensional, multi-scale nature of such time-dependent problems, it is challenging to obtain even statistically accurate predictions of the coarse-grained characteristics of the truth dynamics. Advances in computing power and new theoretical insights have spurred the development of a plethora of reduced-order models (e.g., [15, 14, 56, 61]) and data assimilation techniques (e.g., [3, 28, 27, 16, 52]). Various ways of minimizing uncertainties in imperfect predictions and validating the reduced-order models have been developed in this context (e.g., [49, 40, 50, 8, 48, 9]). Data assimilation aside, one of the most important challenges in improving imperfect predictions concerns the mitigation of model error in reduced-order dynamical models. Recent developments provide new techniques for mitigating coarse-graining errors, and for counteracting errors due to neglecting nonlinear interactions between the resolved and unresolved processes in reduced-order models; these include the stochastic superparameterization [24, 51] and reduced subspace closure techniques [64, 65, 63].

This work focuses on elucidating issues related to an increasingly common technique of *Multi Model Ensemble* (MME) predictions which is complementary to improving individual imperfect models. The heuristic idea behind MME prediction is simple: given a collection of models, consider the prediction obtained through a linear superposition of individual model forecasts in the hope of mitigating the overall prediction error. While there is some evidence in support of the MME approach for improving imperfect predictions, particularly in atmospheric sciences (e.g., [60, 67, 13, 25, 70, 71, 68, 69]), a systematic framework justifying this approach has been lacking. In particular, it is not obvious which imperfect models, and with what weights, should be included in the MME forecast in order to improve imperfect predictions within this framework. Consequently, virtually all operational MME prediction systems for weather and climate are based on equal-weight ensembles [25, 70, 68, 71, 69] which are likely to be far from optimal [13].

The main focus of the present work is on a systematic understanding of benefits and limitations associated with the MME approach to improving imperfect predictions; important practical issues in this context are the following:

(a) How to measure the skill (*statistical accuracy*) of dynamic MME predictions relative to single model predictions?

(b) Is there a condition guaranteeing an improvement of predictions via the MME approach relative to single model predictions?

(c) How to construct an MME for best prediction skill at short, medium and long time ranges?

Here, we consider the MME prediction within a probabilistic/stochastic framework which exploits tools from information theory in order to systematically understand the characteristics of such an approach. This probabilistic framework can be utilized in two different contexts: First, when dealing with deterministic imperfect models, one can consider a time-dependent *probability density function* constructed by initializing the models from a given distribution of initial conditions. Second, the probabilistic prediction framework arises naturally when using stochastic reduced-order models in imperfect predictions which is an increasingly common approach (e.g., [15, 59, 60, 47, 45, 36, 37, 38]). In many operational settings dynamic predictions can be obtained through a weighted superposition of forecasts obtained from a collection of imperfect models (e.g., [25, 70, 68, 71, 69]). However, the individual imperfect models are usually highly complex and not easily tuneable and it is desirable to consider the possibility of prediction improvement by adjusting only the ensemble weights. In order to shed light on the issues (a)-(c) above, we set out an information-theoretic framework capable of

(i) quantification of uncertainty and improving the imperfect predictions via the MME approach;

(ii) providing practical guidelines for improving dynamic MME predictions given a small collection of available imperfect models.

In this work we derive a simple criterion for improving probabilistic predictions via the MME approach. Moreover, we provide a simple justification of why the MME prediction can have a better prediction skill than the single best model in the ensemble. Finally, we derive systematic guidelines for constructing finite model ensembles which are likely to have a superior predictive skill over any single model in the ensemble. These results stem largely from the convexity of *relative entropy* (e.g., [11]) which is used here as a measure of a lack of information in the imperfect models relative to the resolved characteristics of the truth dynamics. We show that the guidelines for MME prediction improvement in the context of a forced perturbation from an equilibrium can be implemented with the help of the linear response theory and the 'fluctuation-dissipation' approach for forced dissipative systems [47, 35, 2, 41, 23, 50, 39]); this approach follows from the earlier work on improving imperfect predictions in the presence of model error in the single-model setup (see, for example, [32, 42, 33, 49, 40, 50, 18, 8, 48]). When considering prediction improvement for the initial value problem the practical implementation of the condition for skill improvement through MME can be carried out in the hindcast/reanalysis mode (e.g., [31]). Although we focus here on mitigating prediction error via the MME approach, it is worth stressing that the ultimate goal in imperfect reduced-order prediction should involve a synergistic approach that combines improvement reduced-order models with an MME framework for both data assimilation and prediction.

This paper is structured as follows: First, in section 2 we motivate the need for a systematic analysis of MME prediction. In section 3 we derive the information-theoretic criterion for improving the skill of MME predictions relative to single model predictions. A set of particularly useful results is discussed in §3.2 where Gaussian models are used in MME; this approach provides a helpful intuition for dealing with the general results of §3. Section 4 combines the analytical estimates of §3 with simple numerical tests which are based on statistically exactly solvable models described in §4.1. We conclude in §5 by summarizing the most important results and discuss directions for further research in this area, including extensions to a multi model ensemble approach to data assimilation. Technical details associated with the analytical estimates derived in §3 are presented in Appendices.

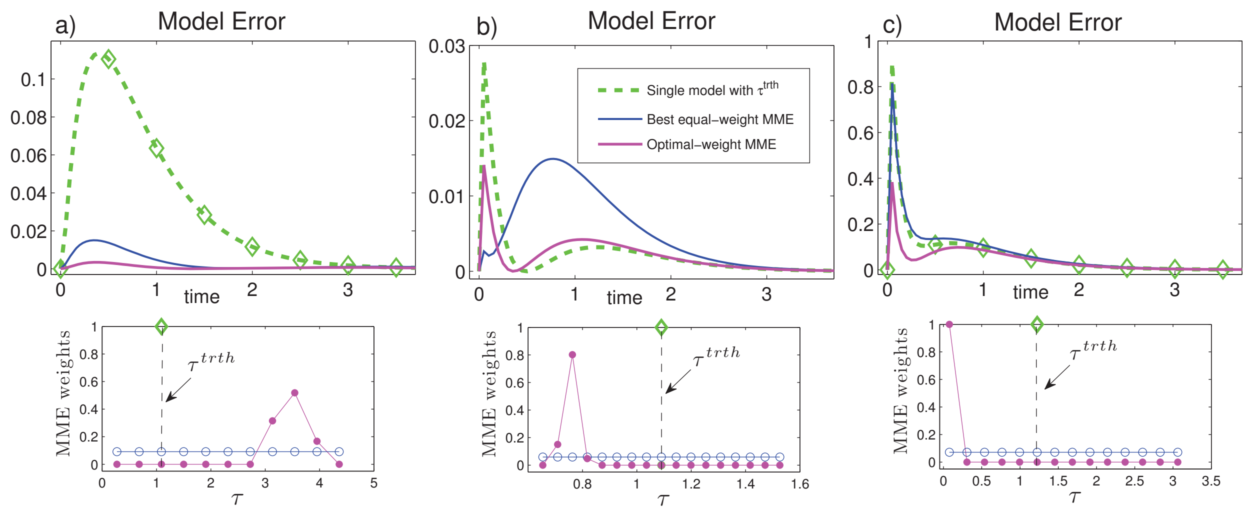

Figure 1: Dependence of the prediction error within the MME framework on the nature of the non-Gaussian truth dynamics (given by (36) in §4.1.2); here, the MME mixture density (1) contains Gaussian models (24) with correct statistics of the initial conditions and correct marginal equilibrium for the resolved dynamics. Top-row insets show the evolution of the error via the relative entropy (2) for predicting the resolved truth dynamics with: (**Left**) symmetric fat-tailed equilibrium density with initial statistical conditions in a stable regime of the truth dynamics; (**middle**) symmetric fat-tailed equilibrium density with initial conditions in an unstable regime of the truth, (**right**) skewed equilibrium density with initial conditions in a stable regime of the truth. The bottom insets show the weights α_i in MME density (1); the optimal-weight MME (9) minimizes the time-averaged relative entropy by adjusting the ensemble weights. Note that the performance of MME prediction, as well as the structure of the optimal-weight MME, depend strongly on the nature of the truth dynamics.

2 Motivating examples

Consider the dynamics of a high-dimensional, nonlinear system where only a small subset of its dynamical variables can be reasonably modelled or accessed through empirical measurements. The *resolved* dynamics of the full system is affected by nonlinear, multi-scale interactions with *unresolved* processes which cannot be observed or even correctly modelled (e.g., [44]). Nevertheless, we are interested in a statistically accurate prediction of the resolved dynamics using a collection of imperfect reduced-order dynamical models which approximate or neglect the interactions between the resolved and unresolved dynamics. To this end, assume that the state vector of dynamical variables in the true high-dimensional system decomposes as $\mathbf{v} = (\boldsymbol{u}, \boldsymbol{v})$, where $\boldsymbol{u} \in \mathbb{R}^K, K < \infty$, denotes the resolved variables and $\boldsymbol{v} \in \mathbb{R}^N$ denotes the unresolved variables; we tacitly assume that $K \ll N$ which is natural when dealing with complex multi-scale dynamics such as the turbulent dynamics of geophysical flows (e.g., [44]). The time-dependent probability density associated with the Multi Model Ensemble (MME) of imperfect reduced-order models on the subspace of resolved variables $\boldsymbol{u}$ is given by a convex superposition of the model densities in the form

$$\pi^{\mathrm{MME}}_{\boldsymbol{\alpha},t}(\boldsymbol{u}) \equiv \sum_{i=1} \alpha_i \pi^{\mathrm{M}_i}_t(\boldsymbol{u}), \qquad \int \mathrm{d}\boldsymbol{u}\, \pi^{\mathrm{MME}}_{\boldsymbol{\alpha},t}(\boldsymbol{u}) = 1, \qquad \sum_{i=1} \alpha_i = 1, \; \alpha_i \geqslant 0, \tag{1}$$

where $\pi^{\mathrm{M}_i}_t$ represent probability densities associated with the imperfect models M_i in some class $\mathcal{M}$. We are particularly interested in mitigating the prediction error via the MME approach by adjusting the model weights α_i in (1) with fixed characteristics of the individual imperfect models $\mathrm{M}_i \in \mathcal{M}$. The lack of information at time t between the MME and the truth statistics on the resolved subspace of variables is measured using the relative entropy [34]

$$\mathcal{P}(\pi_t, \pi^{\mathrm{MME}}_{\boldsymbol{\alpha},t}) = \int \mathrm{d}\boldsymbol{u}\, \pi_t \ln \frac{\pi_t}{\pi^{\mathrm{MME}}_{\boldsymbol{\alpha},t}}\,. \tag{2}$$

This non-negative functional satisfying $\mathcal{P}(\pi_t, \pi^{\mathrm{MME}}_{\boldsymbol{\alpha},t}) = 0$ only when $\pi_t = \pi^{\mathrm{MME}}_{\boldsymbol{\alpha},t}$ is not a proper metric. However, it possesses a number of desirable properties such as convexity in the pair $(\pi, \pi^{\mathrm{MME}})$ and it satisfies a 'triangle equality' for a certain class of densities discussed later (see also [47, 8, 48]); moreover, the relative entropy is invariant under general changes of variables [42, 44] which is a desirable from a measure of model error in applications. We use the relative entropy (2) as an information-based measure of the time-dependent error in imperfect probabilistic predictions; additional measures of predictive skill were introduced earlier in the context of uncertainty quantification in the single model context are discussed in §4.3.1 (see also [40, 50, 45, 8, 48]). Here, we show that the information-theoretic approach is very useful when considering improving imperfect predictions in the MME context. In particular, this setting helps address the following general questions:

- What characteristics of a MME lead to uncertainty reduction relative to a single imperfect model?
- How to construct a MME for best prediction at short, medium and long time ranges?

In the subsequent sections we derive a simple condition for improvement of dynamical predictions via the MME approach relative to the best single imperfect model. However, before embarking on a detailed analysis, some motivating examples are presented in figure 1 which shows that not every MME prediction is superior to the single model prediction, and the structure of the optimal-weight MME depends on both the truth dynamics and the imperfect models in the ensemble. The top-row insets show the evolution of the prediction error in terms of the relative entropy (2) in three different dynamical regimes of a non-Gaussian truth dynamics (described later in §4.1.2). In all cases the statistics of the initial conditions and the marginal equilibrium for the resolved dynamics in the imperfect Gaussian models M_i coincide with those of the truth dynamics; in addition, the single model predictions are carried out with an imperfect model tuned to have the correct correlation time τ^{trth} of the resolved dynamics at equilibrium. The bottom row in figure 1 shows the weight structure of the MME with individual models in the ensemble labelled by the correlation time τ of their equilibrium dynamics; the optimal-weight MME is obtained in this case by minimizing the average relative entropy $\frac{1}{T}\int_0^T \mathcal{P}(\pi_t, \pi_t^{\mathrm{MME}})\mathrm{d}t$ over the whole time interval considered. Note that the error of the MME prediction relative to the single model prediction varies significantly between the three configurations in (a)-(c); moreover, the structure of the optimal-weight MME changes drastically from an MME containing only models with $\tau^{\mathrm{M}_i} > \tau^{trth}$ in (a), to an MME with $\tau^{\mathrm{M}_i} < \tau^{trth}$ in (b), to an MME containing a single imperfect model with the smallest correlation time in the ensemble in (c). The difference between the configuration in (a) and (b) lies in the initial statistical conditions: in (a) the initial conditions are such that the resolved dynamics is in a stable regime, while in (b) the resolved dynamics is initially in a transient unstable phase. The configuration shown in figure 1c corresponds imperfect predictions of the resolved non-Gaussian dynamics when the truth equilibrium statistics is significantly skewed. (See §4.3.2 for more details.)

Clearly, the performance of the MME approach for improving imperfect predictions depends on both the structure of the MME and on the nature of the truth dynamics. The above examples highlight the need for a more analytical insight which would allow to understand when and why the MME approach leads to improved predictions. In the next section we focus solely on this topic and we obtain a sufficient condition for prediction improvement via the MME approach. The general theoretical results derived in §3 are discussed further in §4 based on two simple but revealing test models described in §4.1.

3 Information-theoretic estimates of the predictive skill of MME

Here, we develop an information-theoretic framework for assessing the potential improvement of imperfect predictions through the *Multi Model Ensemble* (MME) approach. First, in §3.1 we derive a condition for improving the predictive skill within the MME framework; it turns out that this condition requires evaluating only certain *least-biased* estimates of the truth which are obtained by maximizing the Shannon entropy subject to moment constraints, making this approach amenable to applications. Implications of this information-based criterion are discussed in §3.1.3 for both the initial value problem and the forced response prediction. More insight and intuition can be gained by restricting the MME prediction problem to the Gaussian mixture configuration which is discussed in §3.2. The results presented here exploit the convexity of the relative entropy (2) between the truth and the MME density in (1) which measures the lack of information in the MME density relative to the truth. Further details, along with some simple proofs of the facts established below, are relegated to Appendix A.

3.1 Improving predictions through MME framework

Consider imperfect probabilistic predictions of the truth dynamics on the subspace of resolved variables $\boldsymbol{u} \in I\!R^K$ based on the Multi Model Ensemble (MME) with density $\pi^{\text{MME}}_{\boldsymbol{\alpha},t}$ in (1). As in §2, we assume that the truth dynamics has a probability density function denoted by $p_t(\boldsymbol{u},\boldsymbol{v})$, $\boldsymbol{v} \in I\!R^N$, $K \ll N$, and the corresponding marginal density on the resolved subspace is $\pi_t(\boldsymbol{u}) = \int p_t(\boldsymbol{u},\boldsymbol{v})\mathrm{d}\boldsymbol{v}$. Given some class $\mathcal{M}$ of reduced-order models for the resolved dynamics $\boldsymbol{u}(t)$, the *best single model*, M_*, for making predictions at time t is given by

$$\mathcal{P}(\pi_t, \pi_t^{\text{M}^*}) = \min_{\text{M}\in\mathcal{M}} \mathcal{P}(\pi_t, \pi_t^{\text{M}}), \qquad\qquad \mathcal{P}(\pi_t, \pi_t^{\text{M}}) = \int \mathrm{d}\boldsymbol{u}\, \pi_t \ln\frac{\pi_t}{\pi_t^{\text{M}}}, \tag{3}$$

where π_t is the truth density, π_t^{M} represents the probability density associated with the models $\text{M} \in \mathcal{M}$, and the relative entropy $\mathcal{P}(\pi_t, \pi_t^{\text{M}})$ measures the lack of information in π_t^{M} relative to the truth marginal density π_t (see [49, 40, 50, 8, 48]). The best single model $\text{M}^*_{\mathcal{I}} \in \mathcal{M}$ for making predictions over the time interval $\mathcal{I}$, is given by

$$\mathcal{P}_{\mathcal{I}}(\pi, \pi^{\text{M}^*_{\mathcal{I}}}) = \min_{\text{M}\in\mathcal{M}} \mathcal{P}_{\mathcal{I}}(\pi, \pi^{\text{M}}), \tag{4}$$

where $\mathcal{P}_{\mathcal{I}}(\pi, \pi^{\text{M}}) := ⨍_{\mathcal{I}} \mathcal{P}(\pi_t, \pi_t^{\text{M}})\mathrm{d}t$ measures the average lack of information over the time interval $\mathcal{I}$. We introduce the following information measure to quantify the performance of the MME prediction relative to the single-model prediction with model $\text{M}_\diamond$

$$\mathfrak{P}^{\text{MME},\text{M}_\diamond}_{\alpha,\mathcal{M},\mathcal{I}} = ⨍_{\mathcal{I}} \Big(\mathcal{P}(\pi_t, \pi_t^{\text{MME}}) - \mathcal{P}(\pi_t, \pi_t^{\text{M}_\diamond})\Big)\mathrm{d}t = ⨍_{\mathcal{I}} \Big(\mathcal{P}(\pi_t^{\text{L}}, \pi_t^{\text{MME}}) - \mathcal{P}(\pi_t^{\text{L}}, \pi_t^{\text{M}_\diamond})\Big)\mathrm{d}t; \tag{5}$$

where $⨍_{\mathcal{I}}$ denotes the integral average over the time interval $\mathcal{I}$ and the second important equality in (5) arises from the 'triangle equality' satisfied by the relative entropy (e.g., [47]), namely

$$\mathcal{P}\Big(\pi_t, \pi_t^{\text{M}}\Big) = \mathcal{P}\Big(\pi_t, \pi_t^{\text{L}}\Big) + \mathcal{P}\Big(\pi_t^{\text{L}}, \pi_t^{\text{M}}\Big), \tag{6}$$

where π_t^{L} is the practically measurable, least-biased density associated with the resolved truth dynamics which maximizes the Shannon's entropy given L moment constraints at time t (see (12) below and [55, 47, 40, 50, 8]). We now have the following:

INFORMATION CRITERION I. The MME prediction over the time interval $\mathcal{I}$ has a smaller error than the single model prediction with $\text{M}_\diamond$ if

$$\mathfrak{P}^{\text{MME},\text{M}_\diamond}_{\alpha,\mathcal{M},\mathcal{I}} < 0\,. \tag{7}$$

This simple information-based criterion circumvents the need for the truth density π and it implies that if the lack of information in the MME prediction $\pi^{\text{MME}}_{\alpha}$ relative to the least-biased estimate π^{L} of the truth is smaller than the lack of information in the single model density $\pi^{\text{M}_\diamond}$ relative to π^{L}, the same holds for the truth density π in lieu of π^{L}. Note also that the single model $\text{M}_\diamond$ in (5) does not have to coincide with the best imperfect model $\text{M}^*_{\mathcal{I}}$ in (4) which is unknown in practice. For example, one might consider $\text{M}_\diamond$ to be the best single model $\text{M}^*_{\mathcal{I},\text{L}}$ relative to the least-biased truth estimate which is defined as

$$\mathcal{P}_{\mathcal{I}}\Big(\pi^{\text{L}}, \pi^{\text{M}^*_{\mathcal{I},\text{L}}}\Big) = \min_{\text{M}\in\mathcal{M}} \mathcal{P}_{\mathcal{I}}\Big(\pi^{\text{L}}, \pi^{\text{M}}\Big), \tag{8}$$

and it clearly depends on the L moment constraints used to estimate the truth density; note that even if $\mathcal{P}_{\mathcal{I}}(\pi^{\text{L}}, \pi^{\text{M}^*_{\mathcal{I},\text{L}}}) = 0$, there might exist an information barrier $\mathcal{P}_{\mathcal{I}}(\pi, \pi^{\text{M}^*_{\mathcal{I},\text{L}}}) = \mathcal{P}_{\mathcal{I}}(\pi, \pi^{\text{L}})$ in the imperfect predictions which can be reduced if more detailed truth estimates are considered [40, 50, 8]. We now have the following two useful facts:

FACT 1. Consider the best model $\text{M}^*_{\mathcal{I},\text{L}}$ in (8) for predicting the resolved truth dynamics $\boldsymbol{u}(t)$ over the time interval $\mathcal{I}$. The prediction of the MME with $\{\text{M}_i\} \in \mathcal{M}$ can be superior to the prediction with $\text{M}^*_{\mathcal{I},\text{L}}$ unless the density of $\text{M}^*_{\mathcal{I},\text{L}}$ coincides with the least-biased marginal density π^{L}; i.e., there might exist a set of models $\{\text{M}_i\} \in \mathcal{M}$ and the corresponding weights $\{\alpha_i\}$ such that $\mathfrak{P}^{\text{MME},\text{M}^*_{\mathcal{I},\text{L}}}_{\boldsymbol{\alpha},\mathcal{M},\mathcal{I}} < 0$ in (7).

FACT 2. Consider the ***optimal-weight MME*** for a given set of imperfect models $\{\text{M}_i\} \in \mathcal{M}$ which is defined relative to the least-biased truth estimate π^{L} as

$$\mathcal{P}_{\mathcal{I}}\Big(\pi^{\text{L}}, \pi^{\text{MME}}_{\boldsymbol{\alpha}^*_{\mathcal{I},\text{L}}}\Big) = \min_{\boldsymbol{\alpha}}\ \mathcal{P}_{\mathcal{I}}\Big(\pi^{\text{L}}, \pi^{\text{MME}}_{\boldsymbol{\alpha}}\Big), \tag{9}$$

where $\boldsymbol{\alpha}$ is the vector of weights in the MME mixture density (1) containing fixed models $\{\text{M}_i\} \in \mathcal{M}$. For a fixed number of constraints L the lack of information $\mathcal{P}_{\mathcal{I}}(\pi^{\text{L}}, \pi^{\text{MME}}_{\boldsymbol{\alpha}^*_{\mathcal{I},\text{L}}})$ corresponds to an information barrier for MME predictions with models $\{\text{M}_i\} \in \mathcal{M}$ over the time interval $\mathcal{I}$. Moreover, if the predictive skill cannot be improved via the MME approach, then $\pi^{\text{MME}}_{\alpha^*_{\mathcal{I},\text{L}}} = \pi^{\text{M}^*_{\mathcal{I},\text{L}}}$, and the information barriers in the single model and MME predictions coincide.

Simple justification of the above facts is illustrated in figure 2 and it follows immediately from the convexity of the relative entropy in the second argument (e.g., [11])

$$\mathcal{P}_{\mathcal{I}}\Big(\pi, \sum_i \alpha_i \pi^{\text{M}_i}\Big) \leqslant \sum_i \alpha_i \mathcal{P}_{\mathcal{I}}\Big(\pi, \pi^{\text{M}_i}\Big), \qquad \alpha_i \geqslant 0,\ \ \sum \alpha_i = 1, \tag{10}$$

the 'triangle equality' satisfied by $\mathcal{P}$ in (6), and the definition in (8). Fact 1 becomes obvious upon considering the fixed-time configuration sketched in figure 2a in the case when $\mathcal{P}(\pi^{\text{L}}, \pi^{\text{M}^*}) > 0$ for the best model in the ensemble (extension of these arguments to the whole time interval $\mathcal{I}$ is straightforward due to the linearity of integration and the fact that $\mathcal{P} \geqslant 0$). If $\mathcal{P}_{\mathcal{I}}(\pi^{\text{L}}, \pi^{\text{M}^*_{\mathcal{I},\text{L}}}) = 0$ then $\pi^{\text{L}}_t = \pi_t^{\text{M}^*_{\mathcal{I},\text{L}}}$ by the properties of the relative entropy. Fact 2 is established by considering the two fixed-time configurations in figure 2. In figure 2b the MME information barrier (9) at time t (red shaded) is the same as that of the single model prediction and equal to $\mathcal{P}(\pi_t, \pi_t^{\text{M}_*})$, while the MME information barrier of MME in figure 2a is reduced to $\mathcal{P}(\pi_t, \alpha^*\pi_t^{\text{M}^*} + (1-\alpha^*)\pi_t^{\text{M}_1}) < \mathcal{P}(\pi_t, \pi_t^{\text{M}^*})$. Clearly, the choice of the imperfect models in MME is important for its improved performance over the single model $\text{M}_\diamond$. (Examples of prediction improvement via MME without reducing the single model information barrier are shown in different configurations in figures 3,6 and 7 discussed in the subsequent sections.)

Clearly, the above general facts relate to important practical issues in prediction problems, such as:

- **(i)** Assessment of prediction improvement for a given MME containing a discrete collection $\{\text{M}_i\} \in \mathcal{M}$ of imperfect models, based solely on the prediction errors $\mathcal{P}(\pi_t, \pi_t^{\text{M}_i})$. Ideally, one would like to optimize the MME weights, α_i, via a minimization of $\mathcal{P}(\pi_t, \sum_i \alpha_i \pi_t^{\text{M}_i})$; however, this requires repeated evaluations of the relative entropy $\mathcal{P}$ which might be not feasible for realistic problems.
- **(ii)** Derivation of guidelines for constructing an MME from a given set of imperfect models that would guarantee prediction improvement when only partial knowledge of the truth dynamics is available.

It turns out that a significant insight into the above issues can be derived within the information-theoretic framework by exploiting the condition (7) and the convexity of the relative entropy in (10) which leads to the following simplified but practical criterion:

INFORMATION CRITERION II. Consider improving imperfect predictions via the MME approach in the case when only the fidelity $\mathcal{P}_{\mathcal{I}}(\pi^{\text{L}}, \pi^{\text{M}_i})$ of individual ensemble members $\text{M}_i \in \mathcal{M}$ can be estimated. MME prediction is preferable to single model predictions with $\text{M}_\diamond \in \mathcal{M}$ if

$$\mathcal{P}_{\mathcal{I}}(\pi^{\text{L}}, \pi^{\text{M}_\diamond}) + \Delta > \sum_{i\neq\diamond} \beta_i\, \mathcal{P}_{\mathcal{I}}(\pi^{\text{L}}, \pi^{\text{M}_i}), \qquad \beta_i = \alpha_i\,(1-\alpha_\diamond)^{-1}, \quad \sum_{i\neq\diamond} \beta_i = 1. \tag{11}$$

where $\Delta \geqslant 0$ is the uncertainty parameter and π_t^{L} is the least-biased density maximizing the Shannon's entropy given L constraints, as in (6).

Remarks:

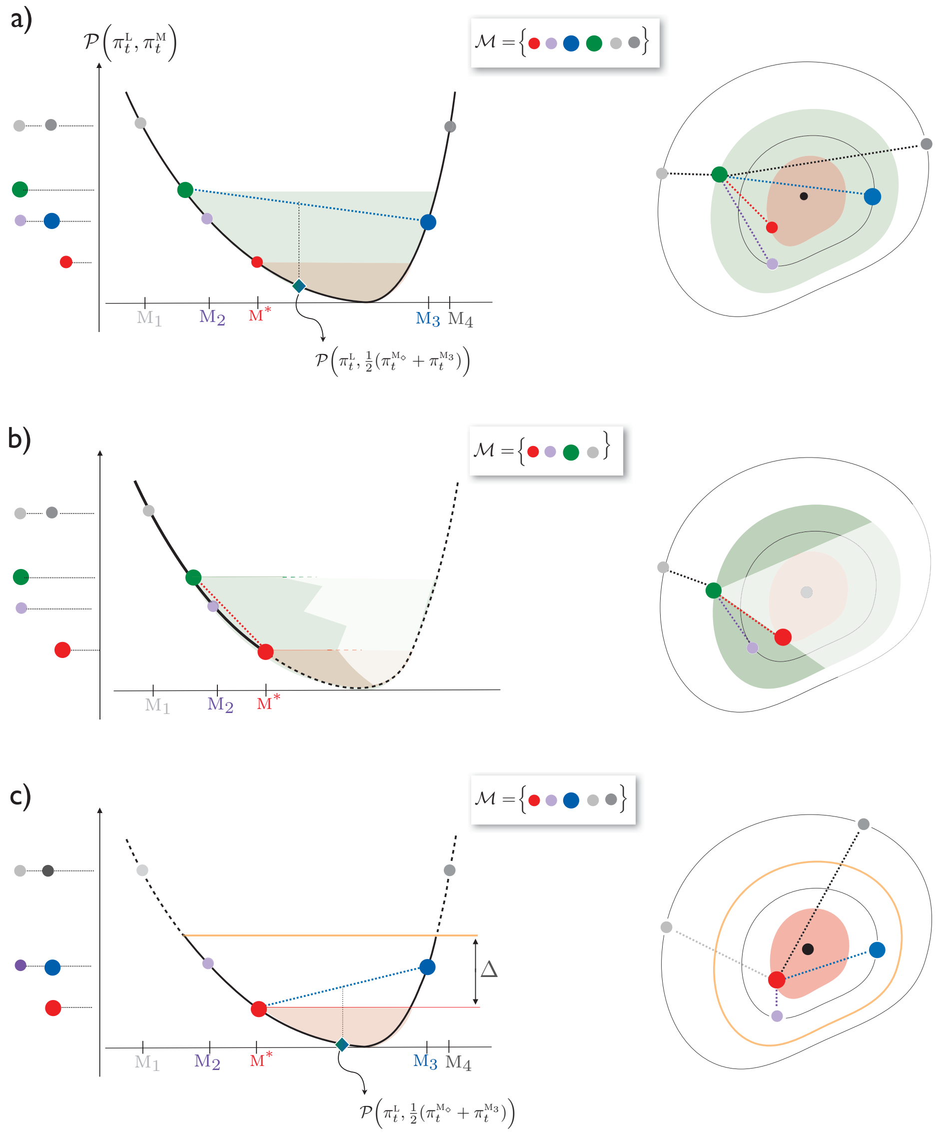

Figure 2: Consequences of the convexity of the relative entropy $\mathcal{P}$ in (10) which are exploited in considerations of prediction improvement within the Multi Model Ensembles (MME) framework with probability density function π_t^{MME} in (1). One-dimensional (left column) and two-dimensional (right column) sketches are shown for a fixed time in order to illustrate two distinct possibilities in MME prediction which depend on the class of available models $\mathcal{M}$: (a) MME can perform better than any individual model in the ensemble $\mathcal{M}$ (e.g., $\pi^{\mathrm{MME}} = \frac{1}{2}(\pi^{\mathrm{M}_\diamond} + \pi^{\mathrm{M}_1})$ see FACT 1), (b) MME cannot outperform the best single model M^* in $\mathcal{M}$. In (a) the information barrier in MME prediction is smaller than the barrier in the single model prediction (see FACT 2). Note also that the error of MME prediction cannot exceed that of the worst model in the ensemble. The configuration in (c) illustrates the information criterion in (11) and the need for introducing the uncertainty parameter $\Delta \geqslant 0$ when only the estimates on the prediction error $\mathcal{P}(\pi^{\mathrm{L}}, \pi^{\mathrm{M}_i})$ of individual models $\mathrm{M}_i \in \mathcal{M}$ are known.

- The uncertainty parameter Δ plays an important role in the above setup. For $\mathrm{M}_\diamond, \mathrm{M}_i \in \mathcal{M}$ the condition (11) implies that $0 \leqslant \mathcal{P}_{\mathcal{I}}(\pi, \pi^{\mathrm{MME}}) \leqslant \mathcal{P}_{\mathcal{I}}(\pi, \pi^{\mathrm{M}_\diamond}) + \Delta$ (see Appendix A). For $\Delta = 0$ (11) provides a sufficient condition for prediction improvement which is too restrictive, in light of FACT 1, since for $\mathrm{M}_\diamond = \mathrm{M}^*_{\mathcal{I},\mathrm{L}}$ no MME would satisfy it (see figure 2c). For $\Delta \neq 0$ the condition in (11) is no longer sufficient for reducing the prediction error; however, it allows for a possible improvement of the predictive performance via the MME approach at the risk of increasing the prediction error by a controllable value Δ relative to $\mathcal{P}_{\mathcal{I}}(\pi, \pi^{\mathrm{M}_\diamond})$ which is also true when $\mathrm{M}_\diamond = \mathrm{M}^*_{\mathcal{I},\mathrm{L}}$ (compare the configurations $\pi^{\mathrm{MME}} = \alpha\pi^{\mathrm{M}^*} + (1-\alpha)\pi^{\mathrm{M}_1}$ with $\pi^{\mathrm{MME}} = \alpha\pi^{\mathrm{M}^*} + (1-\alpha)\pi^{\mathrm{M}_2}$ when $\mathcal{P}(\pi^{\mathrm{L}}, \pi^{\mathrm{M}_1}) = \mathcal{P}(\pi^{\mathrm{L}}, \pi^{\mathrm{M}_2})$). Guidelines for generating the ensemble models and for probing the local geometry of $\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\cdot)$ are presented in §3.1.4.

- The formulation (11) is particularly useful when considering the improvement of the forced response prediction from equilibrium π_{eq}, since then π_t^{L} in (11) can be directly estimated based on the linear response theory and the fluctuation-dissipation formulas which utilize the information from the unperturbed equilibrium (see §3.1.2 and [47, 2, 41, 22, 23, 50, 39])).

3.1.1 Condition for improving imperfect predictions via the MME approach based on the least-biased density representation

It turns out that a significant insight can be gained by representing the condition (11) through the least-biased densities of the imperfect models in the MME density (1); the least-biased approximation, π^{L}, of the true density π given a set of L statistical constraints belongs to the exponential family of densities which maximizes the Shannon entropy $\mathcal{S}=-\int\pi^{\mathrm{L}}\ln\pi^{\mathrm{L}}$ subject to (see, e.g., [55, 47])

$$\int\pi^{\mathrm{L}}(\boldsymbol{u})E_i(\boldsymbol{u})\mathrm{d}\boldsymbol{u}=\int\pi(\boldsymbol{u})E_i(\boldsymbol{u})\mathrm{d}\boldsymbol{u},\quad i=1,\dots,\mathrm{L},$$

where E_i are some functionals on the space of the resolved variables $\boldsymbol{u}$; here we assume these functionals to be i-th tensor power of $\boldsymbol{u}$, i.e., $E_i(\boldsymbol{u})=\boldsymbol{u}^{\otimes i}$, so that their expectations yield the components of the first L statistical moments of π about the origin. Consequently, the least-biased densities of the truth and of the imperfect models are given by (see, e.g., [55, 1, 47, 49])

$$a)\;\;\pi_t^{\mathrm{L}_1}=C_t^{-1}\exp\Big(-\sum_{i=1}^{\mathrm{L}_1}\theta_i(t)E_i(\boldsymbol{u})\Big),\qquad b)\;\;\pi_t^{\mathrm{M,L}_2}=\big(C_t^{\mathrm{M}}\big)^{-1}\exp\Big(-\sum_{i=1}^{\mathrm{L}_2}\theta_i^{\mathrm{M}}(t)E_i(\boldsymbol{u})\Big),\tag{12}$$

where the normalization factors C_t and C_t^{M} are chosen so that $\int\pi_t^{\mathrm{L}_1}\mathrm{d}\boldsymbol{u}=\int\pi_t^{\mathrm{M,L}_2}\mathrm{d}\boldsymbol{u}=1$. While the Gaussian approximations of any density π can always be obtained, existence of π^{L} for $\mathrm{L}>2$ is not guaranteed [55]. We denote the expected values of the functionals E_i in (12) with respect to $\pi_t^{\mathrm{L}_1}$ as $\bar{E}_i$ and with respect to $\pi_t^{\mathrm{M,L}_2}$ as $\bar{E}_i^{\mathrm{M}}$; it is convenient to write these expectations in the vector form as

$$a)\;\overline{\boldsymbol{E}}=\big(\bar{E}_1,\dots,\bar{E}_{\mathrm{L}_1}\big)^T,\qquad b)\;\overline{\boldsymbol{E}}^{\mathrm{M}}=\big(\bar{E}_1^{\mathrm{M}},\dots,\bar{E}_{\mathrm{L}_2}^{\mathrm{M}}\big)^T;\tag{13}$$

note that $\boldsymbol{\theta}_t=\boldsymbol{\theta}\big(\bar{\boldsymbol{E}}_t\big)$ and $\boldsymbol{\theta}_t^{\mathrm{M}}=\boldsymbol{\theta}^{\mathrm{M}}\big(\bar{\boldsymbol{E}}_t^{\mathrm{M}}\big)$ in (12) so that the normalization factors in the least-biased densities are functions of the time-dependent statistical moments, i.e, $C_t=C\big(\bar{\boldsymbol{E}}_t\big)$ and $C_t^{\mathrm{M}}=C^{\mathrm{M}}\big(\bar{\boldsymbol{E}}_t^{\mathrm{M}}\big)$.

Based on the least-biased representations (12) of the truth and model probability densities, the criterion (11) for improvement of imperfect predictions via the MME approach can be written in a form which is particularly suited for further approximations (see Appendix A for a simple proof):

FACT 3. The criterion (11) for improving imperfect predictions via the MME approach with uncertainty $\Delta\geqslant 0$ can be expressed in terms of the statistical moments $\overline{\boldsymbol{E}},\big\{\overline{\boldsymbol{E}}^{\mathrm{M}_i}\big\}$ of the truth and models as

$$\mathscr{A}_{\boldsymbol{\beta},\mathcal{I}}\Big(\pi^{\mathrm{L}_1},\big\{\pi^{\mathrm{M}_i,\mathrm{L}_2}/\pi^{\mathrm{M}_i}\big\}\Big)+\mathscr{B}_{\boldsymbol{\beta},\mathcal{I}}\Big(\big\{\overline{\boldsymbol{E}}^{\mathrm{M}_i}\big\}\Big)+\mathscr{C}_{\boldsymbol{\beta},\mathcal{I}}\Big(\overline{\boldsymbol{E}},\big\{\overline{\boldsymbol{E}}^{\mathrm{M}_i}\big\}\Big)+\Delta>0,\tag{14}$$

where

$$\mathscr{A}_{\boldsymbol{\beta},\mathcal{I}}=\fint_{\mathcal{I}}\mathrm{d}t\int\mathrm{d}\boldsymbol{u}\,\pi_t^{\mathrm{L}_1}(\boldsymbol{u})\,\mathfrak{M}(\boldsymbol{u}),\qquad \mathfrak{M}(\boldsymbol{u})=\sum_{i\neq\diamond}\beta_i\left[\log\frac{\pi_t^{\mathrm{M}_i,\mathrm{L}_2}(\boldsymbol{u})}{\pi_t^{\mathrm{M}_i}(\boldsymbol{u})}-\log\frac{\pi^{\mathrm{M}_\diamond,\mathrm{L}_2}(\boldsymbol{u})}{\pi_t^{\mathrm{M}_\diamond}(\boldsymbol{u})}\right],\tag{15}$$

is non-zero only for those model densities which are not in the least-biased form, i.e., if $\pi_t^{\mathrm{M}_i,\mathrm{L}_2}\neq\pi_t^{\mathrm{M}_i}$, and

$$\mathscr{B}_{\boldsymbol{\beta},\mathcal{I}}=\sum_{i\neq\diamond}\beta_i\fint_{\mathcal{I}}\log\Big[C^{\mathrm{M}_\diamond,\mathrm{L}_2}\big(\overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond}\big)/C_t^{\mathrm{M}_i,\mathrm{L}_2}\big(\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}\big)\Big]\mathrm{d}t\,,\qquad \mathscr{C}_{\boldsymbol{\beta},\mathcal{I}}=\sum_{i\neq\diamond}\beta_i\fint_{\mathcal{I}}\Big[\Big(\boldsymbol{\theta}_t^{\mathrm{M}_\diamond}-\boldsymbol{\theta}_t^{\mathrm{M}_i}\Big)\cdot\bar{\boldsymbol{E}}_t\Big]\mathrm{d}t\,,$$

where the weights β_i are defined in (11) and the vectors of the Lagrange multipliers are given by

$$\boldsymbol{\theta}(\overline{\boldsymbol{E}}) = \begin{cases} (\theta_1, \dots, \theta_{\mathrm{L}_1})^{\mathrm{T}}, \\ (\theta_1, \dots, \theta_{\mathrm{L}_1}, 0, \dots, 0_{\mathrm{L}_2})^{\mathrm{T}}, \end{cases} \qquad \boldsymbol{\theta}^{\mathrm{M}}(\overline{\boldsymbol{E}}^{\mathrm{M}}) = \begin{cases} (\theta_1^{\mathrm{M}}, \dots, \theta_{\mathrm{L}_2}^{\mathrm{M}}, 0, \dots, 0_{\mathrm{L}_1})^{\mathrm{T}}, & \text{if} \quad \mathrm{L}_1 \geqslant \mathrm{L}_2, \\ (\theta_1^{\mathrm{M}}, \dots, \theta_{\mathrm{L}_2}^{\mathrm{M}})^{\mathrm{T}}, & \text{if} \quad \mathrm{L}_1 < \mathrm{L}_2. \end{cases}$$

Remarks:

- The second term, $\mathscr{B}_{\boldsymbol{\beta},\mathcal{I}}$, in (14) is independent of the truth density and it involves only the model densities, $\pi_t^{\mathrm{M}_i}$, in MME.
- The last term, $\mathscr{C}_{\boldsymbol{\beta},\mathcal{I}}$, in (14) depends linearly on the expectations, $\overline{\boldsymbol{E}}_t$, with respect to the least-biased truth density $\pi_t^{\mathrm{L}_1}$; these can be estimated in the hindcast mode in the initial value problem context or from the 'fluctuation-dissipation' formulas when considering improvement of forced response predictions, as discussed below in §3.1.3.
- The expected value in $\mathscr{A}_{\boldsymbol{\beta},\mathcal{I}}$ can be evaluated as long the least-biased approximation, $\pi_t^{\mathrm{L}_1}$, of the truth π_t is known. Moreover, $\mathscr{A}_{\boldsymbol{\beta},\mathcal{I}} = 0$ if the MME contains only least-biased models.

We will exploit the consequences of the above result extensively in the following sections; the main advantage of the above 'least-biased' representation of the condition (11) lies in the fact that it depends explicitly and linearly on the statistical moments $\overline{\boldsymbol{E}}_t$ of the truth which are, in principle, amenable to approximations and estimates through the fluctuation-dissipation formulas when considering the forced response prediction (see [47, 2, 41, 22, 23, 49, 40, 50, 39], as well as §3.1.3).

3.1.2 Predictive skill of MME

Here, we represent the general criterion (11) for improving imperfect predictions via the MME approach in the formulation suitable for various time-asymptotic estimates in the context of the initial value problem. This is obtained by using the representation (14) in terms of the least-biased densities (12) which provides a formulation that is amenable to practical approximations especially when considering the forced response predictions.

Consider the evolution of the marginal density π_t associated with the truth dynamics on the resolved subspace of variables in the form

$$\pi_t(\boldsymbol{u}) = \pi_0(\boldsymbol{u}) + \delta\tilde{\pi}_t(\boldsymbol{u}), \qquad \tilde{\pi}_0 = 0, \qquad \int \tilde{\pi}_t(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = 0, \tag{16}$$

which separates the initial statistical conditions from the subsequent evolution of the marginal probability density for the resolved dynamics; the parameter δ in (16) is arbitrary at this stage but it plays the role of an ordering parameter in the time-asymptotic considerations discussed later in §3.2. The mixture density, π_t^{MME}, in (1) associated with a Multi Model Ensemble of imperfect models M_i contained in a class $\mathcal{M}$ can be written in the same form as (16) so that

$$\pi_t^{\mathrm{MME}} = \sum_i \alpha_i \pi_t^{\mathrm{M}_i}, \qquad \pi_t^{\mathrm{M}_i}(\boldsymbol{u}) = \pi_0^{\mathrm{M}_i}(\boldsymbol{u}) + \delta\tilde{\pi}_t^{\mathrm{M}_i}(\boldsymbol{u}), \qquad \tilde{\pi}_0^{\mathrm{M}_i} = 0, \qquad \int \tilde{\pi}_t^{\mathrm{M}_i}(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = 0. \tag{17}$$

Based on decompositions (16) and (17), evolution of the statistical moments $\overline{\boldsymbol{E}}_t$, $\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}$ of the truth and the models can be written as

$$\overline{\boldsymbol{E}}_t = \overline{\boldsymbol{E}}_0 + \delta\widetilde{\overline{\boldsymbol{E}}}_t, \qquad \overline{\boldsymbol{E}}_t^{\mathrm{M}_i} = \overline{\boldsymbol{E}}_0^{\mathrm{M}_i} + \delta\widetilde{\overline{\boldsymbol{E}}}_t^{\mathrm{M}_i}, \qquad \widetilde{\overline{\boldsymbol{E}}}_0 = \widetilde{\overline{\boldsymbol{E}}}_0^{\mathrm{M}_i} = 0. \tag{18}$$

Consequently, the condition (14) for improving the imperfect predictions via the MME approach can be written in a form which is more amenable to practical estimates.

FACT 4. The condition (14) for improving imperfect predictions via the MME approach with uncertainty Δ can be expressed as

$$\mathscr{A}_{\boldsymbol{\beta},\mathcal{I}}\Big(\pi^{\mathrm{L}_1}, \{\pi^{\mathrm{M}_i,\mathrm{L}_2}/\pi^{\mathrm{M}_i}\}\Big) + \widetilde{\mathscr{B}}_{\boldsymbol{\beta},\mathcal{I}}\Big(\{\overline{\boldsymbol{E}}^{\mathrm{M}_i}\}; \overline{\boldsymbol{E}}_0\Big) + \widetilde{\mathscr{C}}_{\boldsymbol{\beta},\mathcal{I}}\Big(\{\overline{\boldsymbol{E}}^{\mathrm{M}_i}\}, \delta\tilde{\overline{\boldsymbol{E}}}\Big) + \Delta > 0, \tag{19}$$

where $\mathscr{A}_{\boldsymbol{\beta},\mathcal{I}}$ is defined in (15) and

$$\widetilde{\mathscr{B}}_{\boldsymbol{\beta},\mathcal{I}} = \sum_{i\neq\diamond}\beta_i \fint_{\mathcal{I}} \mathrm{dt}\,\Big[\mathcal{P}(\pi_0^{\mathrm{L1}},\pi_t^{\mathrm{M}_\diamond,\mathrm{L2}}) - \mathcal{P}(\pi_0^{\mathrm{L1}},\pi_t^{\mathrm{M}_i,\mathrm{L2}})\Big], \qquad \widetilde{\mathscr{C}}_{\boldsymbol{\beta},\mathcal{I}} = \sum_{i\neq\diamond}\beta_i \fint_{\mathcal{I}} \mathrm{dt}\,\Big[\Big(\boldsymbol{\theta}_t^{\mathrm{M}_\diamond} - \boldsymbol{\theta}_t^{\mathrm{M}_i}\Big)\cdot\delta\widetilde{\overline{\boldsymbol{E}}}_t\Big],$$

where $\boldsymbol{\theta}_t^{\mathrm{M}_\diamond} = \boldsymbol{\theta}_t^{\mathrm{M}_\diamond}\big(\overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond}\big)$, $\boldsymbol{\theta}_t^{\mathrm{M}_i} = \boldsymbol{\theta}_t^{\mathrm{M}_i}\big(\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}\big)$, the weights β_i are defined in (11), and the least-biased truth and model densities are given in (12).

Remarks:

- The evolution of $\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}$ and $\boldsymbol{\theta}_t^{\mathrm{M}_i}$ can be computed directly from the imperfect models.
- When considering the forced response prediction to perturbations of the attractor dynamics, the expected changes, $\delta\widetilde{\overline{\boldsymbol{E}}}_t$, in the truth statistics can be estimated based on the correlations on the unperturbed attractor using the fluctuation-dissipation formulas (e.g., [49, 40, 50]). In the context of the initial value problem $\delta\widetilde{\overline{\boldsymbol{E}}}_t$ can be estimated in the hindcast/reanalysis mode (e.g., [31]).

3.1.3 Initial value problem vs forced response

The framework introduced in §3.1 applies, in principle, to two seemingly distinct cases: (i) improving imperfect predictions from given non-equilibrium statistical initial conditions, and (ii) to improving predictions of the response of the truth equilibrium dynamics to external perturbations. Given the decomposition in (18), the similarities and differences between the initial value problem and the forced response prediction can be summarized as follows:

- For the initial value problem, the initial marginal densities for the resolved dynamics, π_0 and $\pi_t^{\mathrm{M}_i}$, correspond to any smooth probability densities with the initial statistics $\overline{\boldsymbol{E}}_0$ and $\overline{\boldsymbol{E}}_0^{\mathrm{M}}$. However, in the case of the forced response prediction the statistical initial conditions are restricted to the respective equilibrium states, i.e., $\pi_0 = \pi_{eq}$ and $\pi_0^{\mathrm{M}_i} = \pi_{eq}^{\mathrm{M}_i}$, and $\overline{\boldsymbol{E}}_0 = \overline{\boldsymbol{E}}_{eq}$ and $\overline{\boldsymbol{E}}_0^{\mathrm{M}_i} = \overline{\boldsymbol{E}}_{eq}^{\mathrm{M}_i}$.
- The fundamental difference between the initial value problem discussed in §3.1.2 and the forced response prediction lies in the properties of the perturbation terms in the decomposition (16) and the existence of the decomposition (18). In particular,
  - The marginal probability density associated with the evolution of a non-degenerate truth in the initial value problem can always be written in the form (16) and (18). However, the time-dependent terms in (16) and (18) are generally small only for sufficiently short times.
  - In the case of estimating the truth response to external perturbations, the decompositions (16) and (18) apply to non-degenerate hypoelliptic noise (see [26]). For sufficiently small external perturbations the time-dependent perturbations in (16) and (18) remain small for all time. This allows for a practical assessment of the prediction improvement in the forced response via MME through the general conditions (11), (14) or their subsidiaries discussed in §3.1.4 and §3.1.2 when combined with the linear response theory exploiting the fluctuation-dissipation formulas at the unperturbed equilibrium (see, e.g., [47, 44] for more details).

3.1.4 Formal guidelines for constructing MME with superior predictive skill relative to the single model predictions

Here, we consider a perturbative approach which provides practical guidelines for constructing a useful MME from a single model $\mathrm{M}_\diamond$. As discussed earlier (FACT 1 and figure 2), the best single model for making predictions can be inferior to an ensemble of imperfect models which appropriately 'sample' the relative entropy landscape $\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\cdot)$. Such information is inaccessible if only the estimates $\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\pi^{\mathrm{M}_i})$ for individual models $\mathrm{M}_i\in\mathcal{M}$ in the ensemble are available; in such cases the criterions (11) or (14) provide the best possible guidance. However, additional MME improvements can be achieved via testing the local geometry of $\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\cdot)$ if there exists a possibility of perturbing a parameterized family of models.

First, we note that if a globally parameterized family of imperfect models is available, then the same convexity arguments as those used in FACTS 1-3 imply that the MME with densities $\pi^{\mathrm{M}}_{\epsilon_\diamond}, \{\pi^{\mathrm{M}}_{\epsilon_i}\}$ satisfying

$$\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}}, \pi^{\mathrm{M}}_{\epsilon_\diamond}) \leqslant \mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}}, \pi^{\mathrm{M}}_{\epsilon_i}) \quad \text{and} \quad \Big(\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}}, \pi^{\mathrm{M}}_{\epsilon})\Big|_{\epsilon=\epsilon_\diamond}\Big)\Big(\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}}, \pi^{\mathrm{M}}_{\epsilon})\Big|_{\epsilon=\epsilon_i}\Big) < 0, \tag{20}$$

will have an improved prediction skill relative to the single model density $\pi^{\mathrm{M}}_{\epsilon_\diamond}$. The reasons for not choosing the model with the smallest prediction error, $\min[\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}}, \pi^{\mathrm{M}}_{\epsilon_\diamond}), \mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}}, \pi^{\mathrm{M}}_{\epsilon_i})]$, are analogous to those used in FACT 1 and 3. If there is no global parameterization in the imperfect model class, consider an MME with a mixture density generated by perturbing a single model density $\pi^{\mathrm{M}_\diamond}$ so that for $\varepsilon \ll 1$

$$\pi^{\mathrm{MME}}_t = \sum_i \alpha_i \pi^{\mathrm{M}_i,\varepsilon}_t, \qquad \pi^{\mathrm{M}_i,\varepsilon}_t(\boldsymbol{u}) = \pi^{\mathrm{M}_\diamond}_t(\boldsymbol{u}) + \varepsilon\, \tilde{\pi}^{\mathrm{M}_i}_t(\boldsymbol{u}), \qquad \int \tilde{\pi}^{\mathrm{M}_i}_t(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = 0; \tag{21}$$

existence of such perturbed densities $\pi^{\mathrm{M}_i,\varepsilon}_t$ which are non-degenerate (smooth at $\varepsilon = 0$) was shown to exist under minimal assumptions on the model dynamics in [26]; the interested reader should consult [40, 50, 45] for a related treatment of the predictive skill in the single model configuration. Based on the decomposition in (21), the evolution of the statistical moments $\overline{\boldsymbol{E}}^{\mathrm{M}_i,\varepsilon}_t$ for the ensemble members can be written as

$$a)\ \ \overline{\boldsymbol{E}}^{\mathrm{M}_i,\varepsilon}_t = \overline{\boldsymbol{E}}^{\mathrm{M}_\diamond}_t + \varepsilon\,\tilde{\overline{\boldsymbol{E}}}^{\mathrm{M}_i}_t, \qquad b)\ \ \boldsymbol{\theta}^{\mathrm{M}_i,\varepsilon}_t = \boldsymbol{\theta}^{\mathrm{M}_\diamond}_t + \varepsilon\,\tilde{\boldsymbol{\theta}}^{\mathrm{M}_i}_t\big(\overline{\boldsymbol{E}}^{\mathrm{M}_\diamond}_t\big) + \mathcal{O}\big(\varepsilon^2\big), \tag{22}$$

where

$$\tilde{\boldsymbol{\theta}}^{\mathrm{M}_i}_t = \Big(\overline{\boldsymbol{E}}^{\mathrm{M}_\diamond}_t \cdot\nabla\theta^{\mathrm{M}_i}_1|_{\varepsilon=0},\ \overline{\boldsymbol{E}}^{\mathrm{M}_\diamond}_t \cdot\nabla\theta^{\mathrm{M}_i}_2|_{\varepsilon=0}, \dots,\ \overline{\boldsymbol{E}}^{\mathrm{M}_\diamond}_t \cdot\nabla\theta^{\mathrm{M}_i}_{\mathrm{L}_1}|_{\varepsilon=0}\Big)^{\mathrm{T}}.$$

The asymptotic expansions in (22) can be combined with the condition (14) to yield the following:

FACT 5. Consider a Multi Model Ensemble generated by perturbing a single model $\mathrm{M}_\diamond$ so that the statistical moments $\overline{\boldsymbol{E}}^{\mathrm{M}_i,\varepsilon}_t$ and the coefficients $\boldsymbol{\theta}^{\mathrm{M}_i,\varepsilon}_t$ in the least-biased model densities $\pi^{\mathrm{M}_i,\mathrm{L}_2}_t$ are given by (22). The criterion (14) for improving imperfect predictions via the MME approach with uncertainty $\Delta \sim \varepsilon$ can be expressed as

$$\mathscr{A}_{\boldsymbol{\beta},\mathcal{I}}\Big(\pi^{\mathrm{L}_1}, \big\{\pi^{\mathrm{M}_i,\mathrm{L}_2}/\pi^{\mathrm{M}_i}\big\}\Big) + \varepsilon\,\widetilde{\mathscr{C}}_{\boldsymbol{\beta},\mathcal{I}}\Big(\overline{\boldsymbol{E}}, \big\{\overline{\boldsymbol{E}}^{\mathrm{M}_i}\big\}\Big) + \Delta + \mathcal{O}\big(\varepsilon^2\big) > 0, \tag{23}$$

where $\mathscr{A}_{\boldsymbol{\beta},\mathcal{I}}$ is given by (15) and

$$\widetilde{\mathscr{C}}_{\boldsymbol{\beta},\mathcal{I}} = \sum_{i\neq\diamond}\beta_i ⨍_{\mathcal{I}}\Big[\Big(\tilde{\boldsymbol{\theta}}^{\mathrm{M}_\diamond}_t - \tilde{\boldsymbol{\theta}}^{\mathrm{M}_i}_t\Big)\cdot\Big(\overline{\boldsymbol{E}}_t - \overline{\boldsymbol{E}}^{\mathrm{M}_\diamond}_t\Big)\Big]\mathrm{d}t\,,$$

where $\tilde{\boldsymbol{\theta}}^{\mathrm{M}_\diamond}_t = \tilde{\boldsymbol{\theta}}^{\mathrm{M}_\diamond}_t\big(\overline{\boldsymbol{E}}^{\mathrm{M}_\diamond}_t\big)$, $\tilde{\boldsymbol{\theta}}^{\mathrm{M}_i}_t = \tilde{\boldsymbol{\theta}}^{\mathrm{M}_i}_t\big(\overline{\boldsymbol{E}}^{\mathrm{M}_\diamond}_t\big)$, and the weights β_i are defined in (11).

Remarks:

- The perturbations $\tilde{\boldsymbol{\theta}}^{\mathrm{M}_i}_t$ can be computed directly from the imperfect models M_i in MME. The evolution of the truth moments $\overline{\boldsymbol{E}}_t$ can be estimated in the hindcast/reanalysis mode in the context of the initial value problem or via the linear response theory and the fluctuation-dissipation formulas when considering the forced response predictions from equilibrium (e.g., [49, 40, 50]).
- The condition (23) simplifies for Gaussian mixture MME discussed in §3.2 since then $\mathscr{A}_{\boldsymbol{\beta},\mathcal{I}} = 0$.

3.2 Improving imperfect predictions via MME in the Gaussian framework

The analysis presented in §3.1-3.1.3 is particularly revealing in the Gaussian framework, i.e., when $\mathrm{L}_1 = \mathrm{L}_2 = 2$ in (14) or (19), due to the existence of an analytical formula for the relative entropy between two Gaussian densities (e.g., [47]). In such a case the probability density, π^{MME}_t, in (1) of the Multi Model Ensemble is a Gaussian mixture and $\mathscr{A}_{\boldsymbol{\alpha}} = 0$ in the conditions (14), (23), and (19). In order to achieve the maximum simplification of the problem while retaining the crucial features of the framework, we assume here that the reduced-order models on the subspace of the resolved variables $\boldsymbol{u} \in I\!R^K$ for predicting the marginal statistics π_t of the resolved truth dynamics are given by the family of Gaussian Ito SDE's (e.g., [58]) given by

$$\mathrm{d}\boldsymbol{u}^{\mathrm{M}} = \big(-\gamma^{\mathrm{M}}\boldsymbol{u}^{\mathrm{M}} + F^{\mathrm{M}}\boldsymbol{f}^{\mathrm{M}}(t)\big)\mathrm{d}t + \sigma_u^{\mathrm{M}}\mathrm{d}\boldsymbol{W}_u(t), \tag{24}$$

where $\gamma^{\mathrm{M}}, F^{\mathrm{M}}, \sigma^{\mathrm{M}} \in I\!R^{K\times K}$ are diagonal matrices with $\gamma_{ii}^{\mathrm{M}}, \sigma_{ii}^{\mathrm{M}} > 0$, $\|\boldsymbol{f}\|_\infty \leqslant 1$, and $\boldsymbol{W}_u(t)$ is a vector-valued Wiener process with independent components, and the mean dynamics and its covariance are given by the well-known formulas

$$\boldsymbol{\mu}_t^{\mathrm{M}} = \mathbb{E}^{\pi_t^{\mathrm{M}}}[\boldsymbol{u}] = e^{-\gamma^{\mathrm{M}}(t-t_0)}\bar{\boldsymbol{u}}_0 + F^{\mathrm{M}}\int_{t_0}^t e^{-\Gamma^{\mathrm{M}}(t-s)}\boldsymbol{f}^{\mathrm{M}}(s)\mathrm{d}s, \tag{25}$$

$$R_t^{\mathrm{M}} = \mathbb{E}^{\pi_t^{\mathrm{M}}}[\boldsymbol{u}\otimes\boldsymbol{u}] - \boldsymbol{\mu}_t^{\mathrm{M}}\otimes\boldsymbol{\mu}_t^{\mathrm{M}} = e^{-\gamma^{\mathrm{M}}(t-t_0)}R_0\, e^{-\gamma^{\mathrm{M}}(t-t_0)} + \int_{t_0}^t e^{\gamma^{\mathrm{M}}(s-t)}Q\, e^{\gamma^{\mathrm{M}}(s-t)}\mathrm{d}s, \tag{26}$$

where $Q = \sigma^{\mathrm{M}}\otimes(\sigma^{\mathrm{M}})^T$. Consequently, the MME density, π_t^{MME}, in (1) is a linear superposition of Gaussian densities with the statistics evolving according to (25)-(26).

Consider now the time-dependent marginal density, $\pi_t(\boldsymbol{u})$, of the truth on the subspace of resolved variables $\boldsymbol{u} \in I\!R^K$ so that

$$\pi_t = \pi_0 + \delta\tilde{\pi}_t, \quad \tilde{\pi}_0 = 0, \quad \int\tilde{\pi}_t(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = 0. \tag{27}$$

As discussed in §3.1.3, the interpretation of the decomposition in (27) depends on the considered problem. In the context of the initial value problem π_0 corresponds to the uncertainty in the initial conditions, and δ is an ordering parameter utilized below in short-time asymptotic expansions. When considering the forced response to small external perturbations of the truth equilibrium dynamics $\pi_0 = \pi_{eq}$, and we assume the perturbation in (27) is non-singular so that π_t is smooth at $\delta = 0$ which holds under minimal assumptions outlined in [26]. In the Gaussian setting considered here, the decomposition (27) can be used to write the second-order statistics of the truth as

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \delta\tilde{\boldsymbol{\mu}}_t, \quad \tilde{\boldsymbol{\mu}}_0 = 0, \qquad\qquad R_t = R_0 + \delta\tilde{R}_t, \quad \tilde{R}_0 = 0, \tag{28}$$

with analogous expressions for the mean $\boldsymbol{\mu}_t^{\mathrm{M}_i}$ and covariance $R_t^{\mathrm{M}_i}$ of the imperfect Gaussian models (24) in the multi model ensemble.

The general condition (14) or (19) for improving MME predictions in the Gaussian framework can be easily rewritten in terms of the centered moments, $\boldsymbol{\mu}_t, R_t,\ \boldsymbol{\mu}_t^{\mathrm{M}_i}, R_t^{\mathrm{M}_i}$, as discussed in Appendix A. Here, we first highlight a simpler and more revealing version of this condition in the context of the initial value problem which is valid only at sufficiently short times. (The short-time constraint for the initial value problem arises from the technical requirement that the time-dependent terms in the statistical moments $\delta\tilde{\boldsymbol{\mu}}_t, \delta\tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i},\ \delta\tilde{R}_t, \delta\tilde{R}_t^{\mathrm{M}_i}$ be small; see Appendix A.)

FACT 6. Consider the *initial value problem* and imperfect statistical predictions with Gaussian models M_i in (24) with correct initial statistics, i.e., $\boldsymbol{\mu}_0^{\mathrm{M}_i}{=}\boldsymbol{\mu}_0,\ R_0^{\mathrm{M}_i}{=}R_0$, and over a sufficiently short time interval $\mathcal{I} = [0\ T],\ T \ll 1$ so that $\delta\tilde{\boldsymbol{\mu}}_t, \delta\tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i},\ \delta\tilde{R}_t, \delta\tilde{R}_t^{\mathrm{M}_i}$ remain small. The Gaussian mixture MME provides improved predictions relative to the single model predictions with $\mathrm{M}_\diamond$ over the interval $\mathcal{I}$ with uncertainty Δ when

$$\big\{D_{\boldsymbol{\beta},\mathcal{I}}\big(\{\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}^{\mathrm{M}_i}\}\big) + E_{\boldsymbol{\beta},\mathcal{I}}\big(\{\tilde{R}^{\mathrm{M}_i}\}\big) + F_{\boldsymbol{\beta},\mathcal{I}}\big(\tilde{R}, \{\tilde{R}^{\mathrm{M}_i}\}\big)\big\} + \Delta + \mathcal{O}(\delta) > 0, \tag{29}$$

where

$$D_{\boldsymbol{\beta},\mathcal{I}} = \tfrac{1}{2}\sum_{i\neq\diamond}\beta_i \fint_{\mathcal{I}}\Big[(\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_\diamond})^T(R_0)^{-1}(\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_\diamond}) - (\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i})^T(R_0)^{-1}(\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i})\Big]\mathrm{d}t,$$

$$E_{\boldsymbol{\beta},\mathcal{I}} = \tfrac{1}{4}\sum_{i\neq\diamond}\beta_i \fint_{\mathcal{I}}\mathrm{tr}\Big[(\tilde{R}_t^{\mathrm{M}_\diamond} - \tilde{R}_t^{\mathrm{M}_i})(R_0)^{-1}\Big]\mathrm{tr}\Big[(\tilde{R}_t^{\mathrm{M}_\diamond} + \tilde{R}_t^{\mathrm{M}_i})(R_0)^{-1}\Big]\mathrm{d}t,$$

$$F_{\boldsymbol{\beta},\mathcal{I}} = -\tfrac{1}{2}\sum_{i\neq\diamond}\beta_i \fint_{\mathcal{I}} \mathrm{tr}\Big[\tilde{R}_t(R_0)^{-1}\big(\tilde{R}_t^{\mathrm{M}_\diamond} - \tilde{R}_t^{\mathrm{M}_i}\big)(R_0)^{-1}\Big]\mathrm{d}t,$$

with the weights β_i defined in (11).

Remarks:

- For an MME containing models (24), underdamped MME with $\gamma^{\mathrm{M}_i} \leqslant \gamma^{\mathrm{M}_\diamond}$ helps improve the short-time imperfect predictions ($E_{\mathcal{I}} > 0$) but it is not sufficient to guarantee the overall skill improvement. The interplay between the truth and model response in $D_{\mathcal{I}}$ and the truth and model response in the variance in $F_{\mathcal{I}}$ are both important. Moreover, when the truth response $\tilde{R}_t$ in the variance is sufficiently negative the short term prediction skill is not improved through the underdamped MME.
- Even if the short-time condition (29) is satisfied, the medium-range predictive skill of MME might not beat the single model (see §4.3 for examples).

It turns out that the sufficient condition for improving infinite-time forced response predictions via a Gaussian mixture MME takes an even simpler form than (29) for the initial value problem and. This fact follows from invariance of the equilibrium covariance with respect to forcing perturbations in linear Gaussian systems (24), i.e., $\tilde{R}_t = 0$ in (28), and the fact that under minimal assumptions [26] the perturbations in the mean, $\tilde{\mu}_t$, remain small for all time. Thus, we have the following (see Appendix A for details):

FACT 7. Consider the *forced response prediction* via a Gaussian mixture MME containing imperfect Gaussian models (24) with correct equilibrium mean and covariance, i.e., $\boldsymbol{\mu}_{eq}^{\mathrm{M}_i} = \boldsymbol{\mu}_{eq}$ and $R_{eq}^{\mathrm{M}_i} = R_{eq}$. The sufficient condition for improving forced response predictions to small external forcing perturbations via MME over the time interval $\mathcal{I} = [t_1 \;\; t_1{+}T]$ is independent of the truth covariance response, $\tilde{R}_t$, and it is given by

$$D_{\boldsymbol{\beta},\mathcal{I}}\Big(\{\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}^{\mathrm{M}_i}\}\Big) + \mathcal{O}(\delta) > 0, \tag{30}$$

where $\tilde{\boldsymbol{\mu}}_t$ and $\tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i}$ are, respectively, the perturbations of the truth and model mean from their equilibrium values and D_t has the same form as in (29) but with $R_0 = R_{eq}$.

Remarks:

- The condition (30) for improving the infinite-time forced response can be written as

$$\sum_{i\neq\diamond}\beta_i \fint_{\mathcal{I}} \Big[\|\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\diamond}\|^2_{R_{eq}^{-1}} - \|\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{i}\|^2_{R_{eq}^{-1}}\Big]\mathrm{d}t + \mathcal{O}(\delta) > 0, \tag{31}$$

  where $\|\boldsymbol{\mu}\|^2_R = \boldsymbol{\mu}^{\mathrm{T}} R\,\boldsymbol{\mu}$ and the weights β_i are defined in (11). The choice of MME satisfying the above condition depends on the interplay between the truth and model response in the mean, and it becomes difficult for $\tilde{\boldsymbol{\mu}}_t R_{eq}^{-1}\tilde{\boldsymbol{\mu}}_t^i < 0$; a detailed illustration of this fact is presented in §4.2.
- The the truth response in the mean $\tilde{\boldsymbol{\mu}}_t$ can be estimated from the unperturbed equilibrium based on the linear response theory incorporating fluctuation-dissipation formulas [47, 2, 41, 23, 49, 40, 50, 39].
- An MME with superior skill for predicting the infinite-time forced response, so that (30) is satisfied for $t_1 \to \infty$, $T \to 0$, the short or medium-range predictive skill of the same MME might not beat the single model (see examples in §4.3.3).
- In a more general setting (see Appendix A) when $\pi_{eq}^{\mathrm{M}_i} \neq \pi_{eq}$ so that $\boldsymbol{\mu}_{eq}^{\mathrm{M}_i} \neq \boldsymbol{\mu}_{eq}$ and $R_{eq}^{\mathrm{M}_i} \neq R_{eq}$, the interplay between the truth and model response in both the mean and covariance are important (see [49, 40, 50] for related analysis in the single-model configuration).

The insights gained from the conditions highlighted in facts 8 and 9 and its generalizations presented in Appendix A will be used when interpreting the numerical results in §4.3.

4 Tests of the theory for MME prediction

The goal of any reduced-order prediction technique is to achieve statistically accurate estimates for the evolution of the truth dynamics on the resolved subspace of the system variables. MME prediction attempts to accomplish this by combining imperfect reduced-order models, and conditions for the utility of such an approach compared to single model predictions were derived in §3. Here, in order to illustrate these analytical results, we exploit two classes of exactly solvable stochastic models, described in §4.1, which are used to generate the 'truth' dynamics. In §4.2 we use these models to provide a cautionary analytical example illustrating the limitations of ad-hoc applications of the MME framework in the presence of information barriers [48]. In §4.3 we test the information-theoretic criteria derived in §3 for improving predictions via the MME approach with the help of numerical simulations. While an exhaustive numerical study based on complex numerical models is certainly desirable, it is complementary to our goals and a subject for a separate publication.

4.1 Setup for studying the performance of MME skill using exactly solvable test models

Here, we consider two classes of stochastic models which provide the simplest possible examples of effects due to coupling between the resolved and unresolved dynamics and non-Gausianity on the prediction skill using ensembles of reduced models. The stochastic dynamics in these models may be regarded as an idealization of nonlinear couplings with other unresolved degrees of freedom in a much higher-dimensional system (see, for example, [43]). The first class of test models, described in §4.1.1, is given by two-dimensional linear Gaussian models [54, 46, 48] which linearly couple the 'resolved' and 'unresolved' dynamics. These revealing models represent the simplest non-trivial examples of dynamics where neglecting the couplings between the resolved and unresolved processes may lead to information barriers to improving the skill of imperfect predictions (see [46, 40, 50, 48]); we discuss this issue in detail in §4.2 and illustrate it numerically in §4.3.3 in the context of MME predictions of the forced response. The nonlinear, non-Gaussian test models outlined in §4.1.2 and introduced in [17] allow for incorporating a wealth of effects due to unresolved turbulent processes on the resolved dynamics; these include the intermittent bursts of instability due to nonlinear interactions with the unresolved scales and forcing fluctuations at large scales.

4.1.1 The two-dimensional linear Gaussian system

In this linear Gaussian system with the state vector $\boldsymbol{x} = (u,\, v)^T$ the 'resolved' dynamics $u(t)$ is linearly coupled to the 'unresolved' dynamics $v(t)$ according to (see [54, 46, 48])

$$\mathrm{d}\begin{pmatrix} u \\ v \end{pmatrix} = \left[L \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} F(t) \\ 0 \end{pmatrix} \right] \mathrm{d}t + \begin{pmatrix} 0 \\ \sigma \end{pmatrix} \mathrm{d}W(t), \tag{32}$$

where $W(t)$ is the scalar Wiener process, and the matrix L and its eigenvalues $\lambda_{1,2}$ are

$$L = \begin{pmatrix} a & 1 \\ q & A \end{pmatrix}, \qquad \lambda_{1,2} = \frac{1}{2}\left(a + A \pm \sqrt{(a-A)^2 + 4q} \right), \tag{33}$$

with a the damping in the resolved dynamics $u(t)$, A the damping in the unresolved dynamics $v(t)$, and q the coupling parameter between $u(t)$ and $v(t)$. We assume that the deterministic forcing $F(t)$ acts only in the resolved subspace u, and the stochastic forcing affects directly the unresolved dynamics $v(t)$ which is linearly coupled to the resolved dynamics for $q \neq 0$. Since the system (32) is linear with additive noise, it can be easily shown that it has a Gaussian attractor provided that

$$a + A = \lambda_1 + \lambda_2 < 0, \qquad aA - q = \lambda_1\lambda_2 > 0, \tag{34}$$

so that the stable the equilibrium mean $\boldsymbol{\mu}_{eq} = (\mu^u_{eq},\ \mu^v_{eq})$ and covariance R_{eq} of (32) are given by

$$\mu^u_{eq} = -\frac{AF}{\lambda_1\lambda_2}, \quad \mu^v_{eq} = \frac{qF}{\lambda_1\lambda_2}, \qquad R_{eq} = \begin{pmatrix} 1 & -a \\ -a & \lambda_1\lambda_2 + a^2 \end{pmatrix} \frac{-\,\sigma^2}{2(\lambda_1+\lambda_2)\lambda_1\lambda_2}. \tag{35}$$

The autocovariance at equilibrium depends only on the lag τ and it is given by $\mathcal{C}_{eq}(\tau) = R_{eq}\, e^{L^T\tau}$ (see [48] for details). Extensions to the nonautonomous case are trivially obtained if the stability conditions (34) are satisfied so that there exists a Gaussian measure on the attractor (see, e.g., [44, 4]) with the attractor mean, $\boldsymbol{\mu}_{att}(t) \equiv \lim_{t_0\to-\infty}\boldsymbol{\mu}(t,t_0)$, and the same autocovariance $\mathcal{C}_{eq}(\tau)$ as in the autonomous case.

In practical applications the only reliable information that can be extracted from empirical data is the low-order statistics of the resolved dynamics at equilibrium which can be often reproduced by many imperfect models. Thus a natural question arising in this context concerns the choice of model ensemble which is tuned to the equilibrium statistics but reproduces the transient dynamics. Despite the simplicity of the system (32), there exist distinct regimes of transient dynamics parameterized by $\{a, q, A, \sigma\}$ with a stable Gaussian equilibrium satisfying (34) (see [48]). Thus, this toy model may be used to provide insight into an important practical issue which concerns prediction of the transient dynamics using ensembles of imperfect reduced-order models which misrepresent or neglect the couplings between the resolved and unresolved processes. These are exactly the issues considered in §3 and they will be illustrated further using numerical tests in §4.3.

4.1.2 The nonlinear non-Gaussian model

The non-Gaussian dynamics of the second test model is given by the following nonlinear stochastic system (see [17, 5, 8, 7, 48])

$$\begin{aligned} &(a) \quad \mathrm{d}u(t) = \big[(-\gamma(t) + i\omega)u(t) + F(t)\big]\mathrm{d}t + \sigma_u \mathrm{d}W_u(t), \\ &(b) \quad \mathrm{d}\gamma(t) = -d_\gamma(\gamma(t) - \hat{\gamma})\mathrm{d}t + \sigma_\gamma \mathrm{d}W_\gamma(t), \end{aligned} \tag{36}$$

where W_u is a complex Wiener processes with independent components and W_γ is a real Wiener process. The nonlinear system (36), introduced first in a more general form in [17] for filtering multi-scale turbulent signals with hidden instabilities, has a number of attractive properties as a test model in our analysis of the skill of MME prediction exploiting reduced-order models. Firstly, it has a surprisingly rich dynamics mimicking signals in various regimes of the turbulent spectrum, including regimes with intermittently positive finite-time Lyapunov exponents due to large-amplitude bursts of instabilities in $u(t)$ [5, 8, 7] and fat-tailed probability densities for $u(t)$. The equilibrium probability densities in the above regimes have non-zero skewness when $F \neq 0$ in (36a). Secondly, exact path-wise solutions and exact second-order statistics of this non-Gaussian system can be obtained analytically, as discussed in [17].

As in the previous test model (32), we consider $u(t)$ in (36) to be the 'resolved' variable which is nonlinearly coupled with the 'unresolved' variable $\gamma(t)$ which induces damping fluctuations in the resolved dynamics; this nonlinear coupling between the resolved and unresolved dynamics is capable of generating a highly non-Gaussian dynamics which proved valuable in a number of earlier considerations concerned with uncertainty quantification and filtering of turbulent dynamical systems [53, 5, 8, 7, 48]. Moreover, it is worth noting that the additive Wiener processes driving the dynamics of u and γ may be regarded as an idealization of effects of nonlinear couplings with other unreserved degrees of freedom in a much higher dimensional system (see, for example, [**?**]). In §4.3 we will consider numerical tests employing an ensemble of reduced-order Gaussian models for predicting the resolved dynamics $u(t)$ of the non-Gaussian model (36) in its various dynamical regimes determined by the parameters of the hidden, unresolved dynamics of $\gamma(t)$; these tests confirm the estimates obtained from the general information-theoretic framework of §3 and provide insight into additional subtleties associated with MME prediction.

4.2 Information barriers in MME prediction

Prediction improvement within the MME framework is not always guaranteed and it depends on both the choice of the imperfect model ensemble and the truth dynamics, which may be seen as a trivial consequence of the fact stated in §3.1.1. The example discussed below represents the simplest non-trivial configuration illustrating the presence of barriers to prediction improvement within the MME framework, and it augments the previous considerations discussed in [40, 46, 48] in the context of single model predictions.

Consider a configuration when the truth dynamics $(u(t), v(t))$ satisfies (32) and (34) with a stable Gaussian attractor as discussed in §4.1.1, and the imperfect models for the resolved dynamics $u^{\mathrm{M}}(t)$ are given by the linear Gaussian models (24) with correct marginal equilibrium statistics so that

$$\mathbb{E}_{eq}[u^{\mathrm{M}}] = \mathbb{E}_{eq}[u] = \mu_{eq}, \quad \text{and} \quad Var_{eq}[u^{\mathrm{M}}] = Var_{eq}[u] \equiv R_{eq}, \tag{37}$$

where the equilibrium mean of the model dynamics (24) and of the resolved truth (32) are given, respectively, in (25)-(26) and (35). The two constraints in (37) imposed on the family of imperfect models (24) with parameters $(\gamma^{\mathrm{M}}, \sigma^{\mathrm{M}}, F^{\mathrm{M}})$ leave a one-parameter family of models with correct marginal equilibrium statistics which we choose to be parameterized by γ^{M}.

Consider now predictions of the 'infinite-time' response of the truth dynamics to forcing perturbations which change the forcing by $\delta\tilde{F}$ so that the marginal statistics at the new equilibrium of the truth (32) and the model (24) are given by

$$a) \quad \mu^{\delta}_{\infty} = \mu_{eq} - \frac{A}{\lambda_1\lambda_2}\delta\tilde{F}, \qquad b) \quad \mu^{\mathrm{M},\delta}_{\infty} = \mu_{eq} + \frac{1}{\gamma^{\mathrm{M}}}\delta\tilde{F}, \tag{38}$$

while the variance of u and u^{M} remain unchanged since the considered models are linear and Gaussian. In this case the condition (30) for improving the infinite-time forced response prediction via the MME approach relative to singe model predictions with $\mathrm{M}_\diamond$ becomes

$$\sum_{i\neq\diamond}\beta_i\Big[\Big(\frac{A}{\lambda_1\lambda_2} + \frac{1}{\gamma^{\mathrm{M}_\diamond}}\Big)^2 - \Big(\frac{A}{\lambda_1\lambda_2} + \frac{1}{\gamma^{\mathrm{M}_i}}\Big)^2\Big] > 0. \tag{39}$$

The above condition implies existence of two distinct configurations which, similarly to the single model predictions, are distinguished by the sign of the damping parameter A in the unresolved dynamics of (32). These two scenarios were already sketched in figure 2 and we discuss their characteristics below:

(i) No information barrier in the single model prediction ($A < 0$ in the unresolved part of the truth (32) In this case there exists an imperfect model (24) with

$$\gamma^{\mathrm{M}^*_\infty} = 1/\tau^{\mathrm{M}^*_\infty} = -\lambda_1\lambda_2/A > 0; \tag{40}$$

and correct infinite time-response to the forcing perturbations so that $\mathcal{P}(\pi^{\delta}_{\infty}, \pi^{\mathrm{M},\delta}_{\infty}) = 0$. Below M_* denotes a model (24) with $\gamma^{\mathrm{M}*}$ satisfying the constraints (37) which is optimal for infinite-time forced response predictions.

- If $\mathrm{M}_\diamond \neq \mathrm{M}_*$ the MME approach can improve the infinite-time forced response prediction based on (39), see also figure 8. In particular the MME skill is improved for any overdamped MME with $\gamma^{\mathrm{M}_i} \geqslant \gamma^{\mathrm{M}_\diamond}$, see also figure 7. If, additionally, $\mathrm{M}_* \notin \mathcal{M}$, the information barrier in MME can be reduced relative to the single model prediction (see also figure 2a).
- If $\mathrm{M}_\diamond = \mathrm{M}_*$ the MME approach cannot improve the infinite-time forced response prediction based on (39), see also figure 8. The information barrier in MME cannot be reduced relative to the single model prediction with M_*.

(ii) Information barrier in the single model prediction ($A > 0$ in the unresolved part of the truth (32) In this case the infinite-time forced response prediction is improved (at least) for any MME containing models with $\gamma^{\mathrm{M}_i} \geqslant \gamma^{\mathrm{M}_\diamond}$. The information barrier in MME prediction cannot be reduced relative to the single model prediction; this situation is depicted schematically in figure 2b. The *intrinsic barrier* to improving forced response predictions using ensembles of models (24) is given by

$$\mathcal{P}(\pi^{\delta}_{\infty}, \pi^{\mathrm{M}_*,\delta}_{\infty}) = \frac{|\delta\tilde{F}|^2}{2R_{eq}}\left(\frac{A}{\lambda_1\lambda_2}\right)^2, \tag{41}$$

which is achieved only when $\gamma^{\text{M}} \to \infty$. Recall that the model error of the optimal-weight MME, which corresponds to the information barrier of the class of models $\mathcal{M}$, coincides in such a case with the single optimal model M_* (see figures 7-8 and figure 1c for analogous situation in the context initial value problem).

This revealing example of the MME skill for forced response prediction is examined further in §4.3.3 in the case of prediction over a finite time interval, where it is shown that additional information barriers can arise if MME consists of a finite number of models.

4.3 Numerical examples

The goal of this section is two-fold. First, we illustrate the general information-theoretic criteria derived in §3 for improving imperfect predictions via the MME approach with the help of numerical simulations based on the exactly solvable stochastic test models introduced in §4.1. We stress agin that while an exhaustive numerical study based on complex models is certainly desirable, it is complementary to our goals and a subject for a separate publication. The second aim is to illustrate, in a controlled setting, predictive skill differences between the single model and MME approach under additional constraints which arise in applications. In practice, imperfect models are often approximately tuned to the marginal equilibrium statistics of the resolved dynamics which is the only available source of information. However, such a tuning procedure does not necessarily reduce the prediction error in the transient dynamics or in the response to forced perturbations from equilibrium (e.g., [40, 50, 8, 48]). The numerical examples studied below highlight the differences between the structure of MME providing skill improvement for short and medium range predictions (see also Appendix B). Thus, apart from validating the analytical estimates of §3, particular emphasis in this section is on the following issues:

- How significant are the differences in skill between the optimal-weight and equal weight MME's?
- Are MME's with good short prediction skill likely to have good medium range prediction skill?

These themes appear recurrently throughout the remaining sections.

4.3.1 Tuning reduced-order models in the Multi Model Ensemble

In the numerical examples discussed below the MME density, π_t^{MME}, is a Gaussian mixture involving the imperfect model densities, $\pi_t^{\text{M}_i}$, associated with the class $\mathcal{M}$ of linear Gaussian models (32) with correct initial conditions and correct marginal equilibrium statistics for the resolved dynamics. We assume that the marginal equilibrium mean and covariance, $\langle u\rangle_{eq}$, $Var_{eq}[u]$, of the resolved truth dynamics can be estimated from measurements. The following result (see [48]) provides the basis for tuning the marginal equilibrium statistics of the imperfect models in MME:

Proposition 1 *Consider the linear Gaussian dynamics in (24) with coefficients* $\{\gamma^{\text{M}}, \sigma^{\text{M}}, F^{\text{M}}\}$ *and constant forcing. Provided that* $\gamma^{\text{M}} > 0$*, the equilibrium statistics of (24) is controlled by two parameters*

$$\left\{\mu_{eq}^{\text{M}} = \frac{F^{\text{M}}}{\gamma^{\text{M}}}, \quad R_{eq}^{\text{M}} = \frac{(\sigma^{\text{M}})^2}{2\gamma^{\text{M}}}\right\}, \tag{42}$$

which correspond, respectively, to the model mean and variance. There exists is a one-parameter family of models (24) with correct marginal equilibrium statistics of the resolved truth dynamics $u(t)$ *with*

$$(\sigma^{\text{M}})^2 = -2\gamma^{\text{M}}\, Var_{eq}[u], \quad F^{\text{M}} = -\gamma^{\text{M}}\, \mathbb{E}_{eq}[u], \tag{43}$$

where γ^{M} *is a free parameter and* $\mathbb{E}_{eq}[u]$ *and* $Var_{eq}[u]$ *denote the marginal equilibrium mean and variance of the truth dynamics of* $u(t)$.

Thus, the class of imperfect models with correct marginal equilibrium statistics is given by

$$\mathcal{M} := \Big\{\pi_t^{\mathrm{M}}(u) = \mathcal{N}\big(\mu(t), R(t)\big) : \quad \lim_{t\to\infty}\mathcal{P}\Big(\pi_t, \pi_t^{\mathrm{M}}\Big) = 0, \quad \pi_{t_0}^{\mathrm{M}} = \pi_{t_0}\Big\}, \tag{44}$$

where $\mathcal{P}$ is the relative entropy (2). Given the constraints on the initial conditions and the equilibrium model densities in the family $\mathcal{M}$, there is one free parameter left in the models (24) which we choose to be the correlation time $\tau^{\mathrm{M}_i} = 1/\gamma^{\mathrm{M}_i}$. Therefore, the MME densities (1) can be written in this case as

$$\pi_{t;\boldsymbol{\alpha},[\tau]}^{\mathrm{MME}}(u) = \sum_{i=1}^{I}\alpha_i\pi_{t;\tau_i}^{\mathrm{M}_i}(u), \qquad \alpha_i \geqslant 0, \quad \sum\alpha_i = 1, \tag{45}$$

so that the time-dependent MME density, $\pi_{t;\boldsymbol{\alpha},[\tau]}^{\mathrm{MME}}(u)$, is parameterized by the weights $\boldsymbol{\alpha} \equiv [\alpha_1, \dots, \alpha_I]$ and the distribution of the correlation times denoted by $[\tau]$; here, we assume that $[\tau]$ is given by a vector of correlation times evenly distributed between τ_{min} and τ_{max}

$$[\tau] = \Big\{\tau_{min}, \tau_{min} + \Delta\tau, \dots, \tau_{min} + (n-1)\Delta\tau, \tau_{max}\Big\}, \qquad \Delta\tau = (\tau_{max} - \tau_{min})/n, \tag{46}$$

and that $\tau^{trth} \in [\tau]$ denotes the correct correlation time of the marginal dynamics $u(t)$ in (36). In general, the Gaussian Itô diffusions in (24) cannot reproduce the marginal two-point equilibrium statistics of the true resolved dynamics (see [48] for details). However, there exists a linear Gaussian model (24) with the correct correlation time, $\tau^{\mathrm{M}} = \tau^{trth}$, where

$$1/\tau^{\mathrm{M}} \equiv (R_{eq}^{\mathrm{M}})^{-1}\int_0^\infty \langle u^{\mathrm{M}}(t)u^{\mathrm{M}}(t+\tau)\rangle \mathrm{d}\tau, \qquad 1/\tau^{trth} \equiv Var_{eq}^{-1}[u]\int_0^\infty\langle u(t)u(t+\tau)\rangle\mathrm{d}\tau. \tag{47}$$

In the analysis of §4.2 and §4.3 we will assume that the single model predictions are carried out using a model with correct correlation time for the resolved dynamics; this setup is justified by the fact that the correlation time estimates are usually the next easiest quantity to estimate from the measurements, apart from the mean and covariance. Finally, we adopt the following characterization of the ensemble structure:

- ***Balanced MME*** is given by imperfect models (24) with correlation times $\tau^{\mathrm{M}} = 1/\gamma^{\mathrm{M}}$ $\boldsymbol{\{\tau^{\mathrm{M}_i}\}_{i\in I} < \tau^{trth} < \{\tau^{\mathrm{M}_j}\}_{j\in J}}$, $\#I = \#J$ and correct marginal equilibrium statistics for the resolved dynamics,
- ***Underdamped MME*** is given by imperfect models (24) with correlation times $\boldsymbol{\tau^{\mathrm{M}_i} \geqslant \tau^{trth}}$ (so that $\gamma^{\mathrm{M}_i} \leqslant 1/\tau^{trth}$) and correct marginal equilibrium statistics for the resolved dynamics, (48)
- ***Overdamped MME*** is given by imperfect models (24) with correlation times $\boldsymbol{\tau^{\mathrm{M}_i} \leqslant \tau^{trth}}$ (so that $\gamma^{\mathrm{M}_i} \geqslant 1/\tau^{trth}$) and correct marginal equilibrium statistics for the resolved dynamics.

The skill (i.e., statistical accuracy) of the imperfect predictions is assessed using two information measures [21, 19, 20, 8, 48] exploiting the relative entropy (2): (i) the *model error*

$$\mathscr{E}_t^{\mathrm{M}} = \mathcal{P}(\pi_t(u), \pi_t^{\mathrm{M}}(u)), \qquad \mathscr{E}_t^{\mathrm{MME}} = \mathcal{P}(\pi_t(u), \pi_t^{\mathrm{MME}}(u)), \tag{49}$$

and (ii) the *internal prediction skill*

$$\mathscr{I}_t = \mathcal{P}(\pi_t(u), \pi_{eq}(u)), \qquad \mathscr{I}_t^{\mathrm{M}} = \mathcal{P}(\pi_t^{\mathrm{M}}(u), \pi_{eq}^{\mathrm{M}}(u)), \qquad \mathscr{I}_t^{\mathrm{MME}} = \mathcal{P}(\pi_t^{\mathrm{MME}}(u), \pi_t^{\mathrm{MME}}(u)), \tag{50}$$

for the truth, a single model M, and an MME relative to their respective equilibria.

4.3.2 Tests of MME prediction skill in the context of initial value problem

Here, we use the test models described in §4.1.1-4.1.2 in order to provide a more complete picture of MME prediction and augment the analytical results of §3 with simple numerical simulations. Particular focus is on the issues raised at the beginning of §4.3 which are not easily tractable analytically; these include differences between the equal-weight and optimal-weight ensembles, information barriers, and the change of structure in the optimal-weight ensemble depending on the prediction horizon.

Gaussian truth & Gaussian mixture MME

We begin by considering the simplest possible configuration in which both the truth dynamics and the imperfect models in the Multi Model Ensemble are Gaussian. The truth dynamics is given by the two-dimensional Gaussian model (32) outlined in §4.1.1 where the resolved dynamics is linearly coupled to the unresolved dynamics. The MME density $\pi^{\mathrm{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ is a finite Gaussian mixture associated with the one-parameter class $\mathcal{M}$ (44) of linear Gaussian models (32) which is parameterized by the weights vector $\boldsymbol{\alpha}$ and the distribution $[\tau]$ (46) of correlation times in the imperfect models (24); correct statistical initial conditions and correct marginal equilibrium statistics for the resolved dynamics are imposed.

Figure 3 illustrates the dependence of MME prediction skill on the structure of the ensemble (see (48)) for fixed initial uncertainty $Var_0[u]$ in the resolved dynamics of the Gaussian truth (32); in all cases the performance of equal-weight MME (48) and optimal-weight MME (4) is compared with predictions of a single model $\mathrm{M}_\diamond \in \mathcal{M}$ which has a correct correlation time τ^{trth} (47). The *optimal-weight* Gaussian MME is obtained by minimizing the relative entropy between the MME density $\pi^{\mathrm{MME}}_{t;\boldsymbol{\alpha},[\tau]}(u)$ and the marginal truth density $\pi_t(u)$, as in (9); recall that the error of optimal-weight MME prediction corresponds to an information barrier in the MME predictions (see FACT 2 in §3) and it is useful for assessing the skill of the equal-weight MME. The information criterion (11) for each of the considered cases is indicated in the corresponding insets. Below we summarize the most important points revealed by the simulations:

- The equal-weight MME tends to outperform the single model predictions with correct correlation time τ^{trth} provided that the MME is either underdamped or balanced (see (48) and figure 3); this is reminiscent of the short time results summarized in Fact 6 of §3.2.
- Information barriers in MME prediction in this setting are reduced relative to the single model prediction for balanced or underdamped MME (48) and moderate uncertainty $Var_0[u]$ in the initial conditions for the resolved dynamics in (32). For $Var_0[u] \ll Var_{eq}[u]$ the optimal-weight MME collapses onto the most underdamped model in the ensemble (see (48). For $Var_0[u] \sim Var_{eq}[u]$ the optimal-weight MME collapses onto the most overdamped model in the ensemble.
- Weight optimization in MME improves the prediction skill improvement (figure 3); however, this type of optimization is impractical since it requires good estimates of the truth dynamics. Optimization of the correlation times $[\tau]$ in (46) is impractical but improves the prediction skill.

Non-Gaussian truth & Gaussian mixture MME

Here, the non-Gaussian truth dynamics is given by the exactly solvable model (36) where the resolved dynamics is nonlinearly coupled with the unresolved dynamics which induces fluctuations in the effective damping of the resolved component. This non-Gaussian case with fat-tailed and skewed probability densities provides a very useful test bed for validating the analytical estimates derived in §3 and for exploring further intricacies; in fact, one example exploiting this revealing model was already shown in figure 1 of §2. As in the previous configuration, the imperfect models in MME are in the class $\mathcal{M}$ (44) of linear Gaussian models (24) so that the MME density $\pi^{\mathrm{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ in (45) is given by a Gaussian mixture. The optimal-weight MME, whose prediction error corresponds to an information barrier (FACT 2 in §3), is obtained by minimizing the relative entropy between the MME statistics and the marginal truth density, as in (9). The single model prediction is carried out using $\mathrm{M}_\diamond \in \mathcal{M}$ with correct correlation time τ^{trth} of the resolved equilibrium dynamics $u(t)$.

Figures 4-6 illustrate the dependence of the predictive skill of the Gaussian mixture MME (45) as a function of time for increasing variance $Var_0[u]$ of the initial statistics for the resolved variable $\pi_0(u)$ in the non-Gaussian truth dynamics (36). In the configuration examined in figure 4 the marginal equilibrium statistics $\pi_{eq}(u)$ of the truth (36) is symmetric but highly non-Gaussian (see regime II in [8]) and the dynamics is initiated from a stable regime, i.e, when $\pi_0(\gamma) = \mathcal{N}(\alpha\langle\gamma\rangle_{eq}, \beta Var_{eq}[\gamma]), \alpha > 0, \beta \ll 1$. In figure 5 the dynamics of (36) is initiated initiated from an unstable phase, i.e., $\pi_0(\gamma) = \mathcal{N}(-\alpha\langle\gamma\rangle_{eq}, \beta Var_{eq}[\gamma]), \alpha > 0, \beta \ll 1$ and the initial stage of the truth evolution is characterized by a rapid transient dynamics. Finally, in figure 6 the marginal equilibrium statistics of the resolved dynamics $u(t)$ in the truth is skewed ($\pi_{eq}(u)$ is fat-tailed with positive skewness) and $\pi_0(\gamma) = \pi_{eq}(\gamma)$. The prediction skill of the equal-weight MME over the time interval $\mathcal{I} = [0\ T]$ is shown for the spread $[\tau]$ in (46) with the best skill (solid blue) and for the spread $[\tau]$ with the worst skill (dotted blue) within the maximum spread of $[\tau]_{max} = 10\,\tau^{trth}$; these are defined as:

- ***Best equal-weight MME*** corresponds to the ensemble with density $\pi^{\text{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ in (45) with $\alpha_i = const.$ and the smallest prediction error (49) within the examined spread $[\tau]_{max}$ of correlation times of the models in MME.
- ***Worst equal-weight MME*** corresponds to the ensemble with density $\pi^{\text{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ in (45) with with $\alpha_i = const.$ and the largest prediction error (49) within the examined spread $[\tau]_{max}$ of correlation times of the models in MME.

Considering these two extreme cases helps judge the sensitivity of the equal weight MME to the spread of the correlation times in the ensemble; the information criterion (11) for each of the considered cases is indicated in the corresponding insets. Based on the examples illustrated in figures 4-6, we recapitulate the general features of MME prediction in this setting as follows:

• **Symmetric marginal attractor density $\boldsymbol{\pi_{eq}(u)}$ of the truth**: MME prediction skill tends to be superior to that of the single model $\mathrm{M}_\diamond$ with the correct correlation time τ^{trth} for the resolved dynamics except when $\pi_0(\gamma)$ is in unstable regime of the truth dynamics in (36). The following trends in the structure of MME (see (48)) are observed:

- Underdamped equal-weight MME (48) performs similarly well to the optimal-weight MME (4) for predicting the resolved dynamics, $u(t)$, when the dynamics $(u(t), \gamma(t))$ of (36) is initiated from a stable regime (figure 4). This behavior is reminiscent of the short-time estimate summarized in Fact 6 of §3.2. Similar, conclusions apply to balanced MME (48) when the dynamics is initiated in a stable regime, as above, or when $\pi_0(\gamma) = \mathcal{N}(\langle\gamma\rangle_{eq}, Var_{eq}[\gamma])$. The information barrier in MME prediction is reduced relative to the single model prediction for sufficiently small initial uncertainty $Var_0[t]$.
- When the dynamics $(u(t), \gamma(t))$ of (36) is initiated from an unstable regime (figure 5) the MME prediction does not provide a significantly improved skill over the predictions with a single model $\mathrm{M}_\diamond$ with correct correlation time τ^{trth} for the resolved dynamics.
- The sensitivity to the spread $[\tau]$ of correlation times $\tau^{\mathrm{M}_i} = 1/\gamma^{\mathrm{M}_i}$ in the equal-weight Gaussian MME (45) with models (24) shadows that of the optimal-weight MME (9) and increases with decreasing uncertainty $Var_0[u]$ of the resolved initial conditions. However, the optimal spread $[\tau]_{opt}$ for predictions over the time interval $\mathcal{I} = [0\ \ T]$ grows with the uncertainty $Var_0[u]$ (not shown).
- Weight optimization in MME improves the prediction skill improvement (figure **??**); however, this type of optimization is impractical since it requires good estimates of the truth dynamics. Optimization of the correlation times $[\tau]$ in (46) is impractical but improves the prediction skill.

• **For skewed marginal attractor density $\boldsymbol{\pi_{eq}(u)}$ of the truth** the following points are worth noting:

- The information barrier for Gaussian MME predictions in this regime coincides with the most overdamped single model M_i in the ensemble. The single model predictions based on $\mathrm{M}_\diamond$ with correct correlation time τ^{trth} differ little from the optimal single model.
- Prediction skill of equal-weight balanced MME (see (48)) over the time interval $\mathcal{I} = [0\ \ T]$ is poor and comparable to that of the single model $\mathrm{M}_\diamond$ with correct correlation time τ^{trth}. Some improvement at short times $T \leqslant 1$ and small uncertainty in the initial statistical conditions can be observed even for balanced MME, which is improved further (not shown) for an underdamped MME in line with the conclusions in FACT 8 §3.2.

4.3.3 MME prediction of forced response

In this section we augment the analytical considerations of §3.2, and the asymptotic infinite-time example discussed in §4.2, with simple numerical tests of the forced response estimation over a finite-time interval $\mathcal{I} = [0\ T]$ through a Gaussian mixture MME, as described in §4.3.1. The truth dynamics is Gaussian and given by the model (32) with hidden dynamics that induces stochastic fluctuations in the resolved dynamics. The imperfect models in MME are in the class $\mathcal{M}$ (44) of linear Gaussian models (24) so that the MME density $\pi^{\text{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ is given by the Gaussian mixture in (45) which is parameterized by the weights vector $\boldsymbol{\alpha}$ and the distribution $[\tau]$ (46) of correlation times in the imperfect models (24). The qualitative understanding of the results below can be obtained with the help of the schematic figure 2 discussed in §3.

In contrast to the initial value problem in §4.3.2, the initial statistical conditions in the tests of the forced response prediction coincide with the unperturbed marginal equilibrium statistics of the truth. The response in the resolved truth dynamics (32), the imperfect models (24), and MME (45) is induced by the forcing perturbations of the 'ramp' type which changes linearly between F_0 and $F_0 + \delta\tilde{F}$ over a time interval $[t_{min}\ \ t_{max}]$. Here, the truth response to forcing perturbations is computed directly from the test model but, as already pointed out in §3.1.3, for sufficiently small perturbations δF the truth response can be estimated via the linear response theory and the fluctuation-dissipation formulas utilizing the unperturbed equilibrium statistics (see [47, 35, 2, 41, 23, 50, 39] for additional information).

Figures 7 and 8 show two distinct examples of prediction skill for the forced response of the resolved dynamics $u(t)$. Figure 7 shows the skill of imperfect predictions of the forced response of the truth in (32) to small forcing perturbations in the presence of an 'infinite-time' information barrier (see §4.2) in the class $\mathcal{M}$ (44) of the imperfect Gaussian models (24). This configuration corresponds to that sketched in figure 2b when the optimal-weight MME for predicting the infinite-time response coincides with a single model M_* in $\mathcal{M}$ with the smallest prediction error. Thus, the information barrier for doing infinite-time response predictions cannot be reduced via the MME utilizing models from $\mathcal{M}$. Nevertheless, both the finite- and infinite-time predictive skill can be improved via the MME approach relative to the single model $\text{M}_\diamond \neq \text{M}_*$ for any *overdamped* MME as summarized below. The information criterion (11) for improving predictions relative to $\text{M}_\diamond$ with correct correlation time τ^{trth} (47) for the resolved equilibrium dynamics $u(t)$ is indicated in the corresponding insets.

Figure 8 shows the skill of imperfect predictions of the forced response of the truth in (32) to small forcing perturbations when there is no 'infinite-time' information barrier (see §4.2) in the class $\mathcal{M}$ (44) of imperfect Gaussian models (24). In this case we compare the predictive skill of three different types of MME defined in (48) with two single-model predictions. The first model $\text{M}_\diamond$ has the correct correlation time τ^{trth} for the resolved equilibrium dynamics $u(t)$ which can be assessed from empirical data. The second model M^*_∞ has the correct infinite-time forced response but it is unlikely to be known a priori. In this configuration the optimal-weight MME for predicting the infinite-time response collapses onto the single model M^*_∞ and there is no information barrier for the infinite-time forced response (consider figure 2a with $\mathcal{P}(\pi^{\text{L}}_\infty, \pi^{\text{M}^*}_\infty) = 0$). However, the prediction of the forced response over a finite time interval can be improved even relative to M^*_∞, as evidenced by the structure of the optimal-weight MME for doing predictions over a time interval $\mathcal{I} = [t_0\ \ t_0 + T]$.

Below we summarize the most important points illustrated in figures 7 and 8:

- The improvement of forced response with the Gaussian MME is controlled by the condition (31) in Fact 9 of §3.2 which in the present configuration becomes

$$\sum_{i\neq\diamond}\beta_i(\gamma^{\text{M}_i} - \gamma^{\text{M}_\diamond})(\mathcal{X}_i^2 - \tilde{\mu}\tilde{F}) > 0, \tag{51}$$

  where $\mu_\infty = \mu_{eq} + \delta\tilde{\mu}$ is the perturbed truth mean in response to the perturbed forcing $F_\infty = F_0 + \delta\tilde{F}$, and $\mathcal{X}_i^2 = \frac{1}{2}\tilde{F}^2(\gamma^{\text{M}_i} + \gamma^{\text{M}_\diamond})(\gamma^{\text{M}_i}\gamma^{\text{M}_\diamond})^{-2} > 0$. There are two obvious cases when the forced-response prediction is improved within the MME framework:

  (i) when $\mathcal{X}_i^2 - \tilde{\mu}\tilde{F} > 0$ an 'overdamped' MME with $\gamma^{\text{M}_i} \geqslant \gamma^{\text{M}_\diamond}$ in (24) yields improved prediction. In this case increasing the spread $[\tau]$ (46) of correlation times in an overdamped MME with correct equilibrium and correct statistical initial conditions improves the MME prediction skill of the forced response.

(ii) when $\mathcal{X}_i^2 - \tilde{\mu}\tilde{F} < 0$ the infinite-time forced response is improved by an 'underdamped' MME with $\gamma^{\mathrm{M}_i} \leqslant \gamma^{\mathrm{M}_\diamond}$. In this case increasing the spread $[\tau]$ (46) of correlation times in underdamped MME with correct equilibrium and correct statistical initial conditions improves the MME prediction skill of the forced response.

The configuration shown in figures 7 and 8 corresponds to the setting (i) so that an overdamped MME improves the forced response prediction. However, in the case with information barrier (figure 7) we have $\tilde{\mu}\tilde{F} < 0$ and in the case with no information barrier (figure 8) we have $0 < \tilde{\mu}\tilde{F} < \mathcal{X}_i^2$. The expected change of the truth mean $\tilde{\mu}$ can be estimated via the linear response theory and the fluctuation-dissipation formulas, while the perturbation of the model means can be estimated directly from the models.

- Weight optimization in MME provides a significant prediction skill improvement over the equal-weight MME. While this type of optimization is impractical, it helps reveal information barriers in the MME prediction (see FACT 2 of §3 and figure 2) and judge the skill of equal-weight MME. In the present setting the following cases are worth noting:

  No information barrier in MME prediction ($A < 0$ in the unresolved part of the truth (32) so that $0 < \tilde{\mu}\tilde{F} < \mathcal{X}_i^2$ in (51)). If $\mathbf{M}_\diamond \neq \mathbf{M}_*$ with optimal damping γ^{M_*} in (24), the MME approach can improve the infinite-time forced response prediction (see figure 8 and (51) above). In particular, since this configuration falls into the case (i) above, the MME skill is improved for any overdamped MME with $\gamma^{\mathrm{M}_i} \geqslant \gamma^{\mathrm{M}_\diamond}$ where $\mathrm{M}_\diamond$ has the correct correlation time τ^{trth} which can be tuned from measurements of the resolved truth equilibrium dynamics.

  There is no information barrier for doing infinite-time response predictions of the resolved dynamics in (32) within the class of models $\mathcal{M}$ containing M_*; consequently, the optimal-weight MME for infinite-time response collapses onto M_*. Forced-response prediction over a whole time interval $\mathcal{I} = [t_0\ t_0 + T]$ is different as evidenced by the nontrivial structure of the optimal-weight MME in figure 8 and the associated information barrier (see FACT 2 in §3); in this case the optimal-weight MME concentrates around two models: the model with correlation time closest to M_* in the given ensemble, and the most overdamped model which helps improve the short-time prediction skill.

  Information barrier in MME prediction ($A > 0$ in the truth mean (35) so that $\tilde{\mu}\tilde{F} < 0$ in (51)). Despite the presence of an information barrier to infinite-time forced response prediction (see §4.2), this configuration also falls into the case (i) above since $\tilde{\mu}\tilde{F} < 0$; consequently, the equal-weights overdamped MME outperforms the single model predictions with correct correlation time τ^{trth} (figure 7); moreover, the balanced MME (see (48)) with sufficiently narrow spread of $[\tau]$ of correlation times also performs satisfactorily. The information barrier in MME prediction of the forced response of (32) cannot be reduced relative to the single model prediction due to the structure of the ensemble $\mathcal{M}$ containing models (24); this is depicted schematically in figure 2b and was discussed in §4.2.

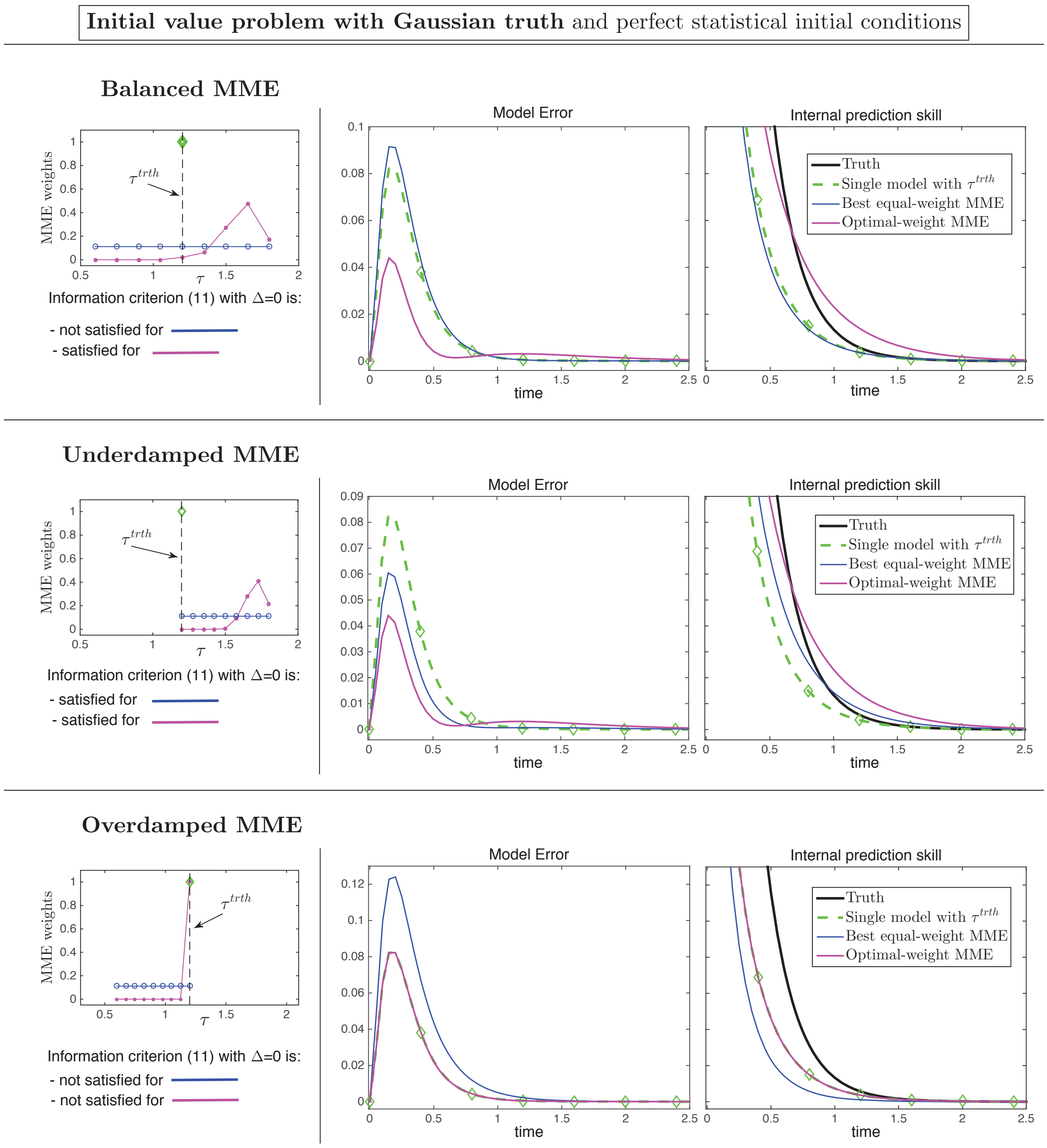

Figure 3: Initial value problem (see electronic version for colors). Prediction skill of a 9-model MME with correct statistical initial conditions for the resolved dynamics $u(t)$ of the Gaussian model (32) for three different types of ensemble structure (see (48)). The MME is a mixture of Gaussian models (24) with correct equilibrium statistics (44) and correlation times τ^{M_i} sampled around the correct correlation time τ^{trth} with the spread $[\tau] = 0.5\tau^{trth}$ (see (46)); the optimal-weight MME (magenta) is obtained by minimizing the relative entropy as in (9).
Truth parameters in (32): $A = -0.5, a = -5.5, \lambda_{1,2} = -1, -5; \sigma : 0.77, F_0 : 1, E = 0.01, \langle u\rangle_{eq} = 0.1, \langle v\rangle_{eq} = 1.35.$
Initial conditions (both truth and MME): $\langle u\rangle_0 = 1.05\langle u\rangle_{eq}, \langle v\rangle_0 = 1.1\langle v\rangle_{eq}, R_0 = 0.2R_{eq}.$

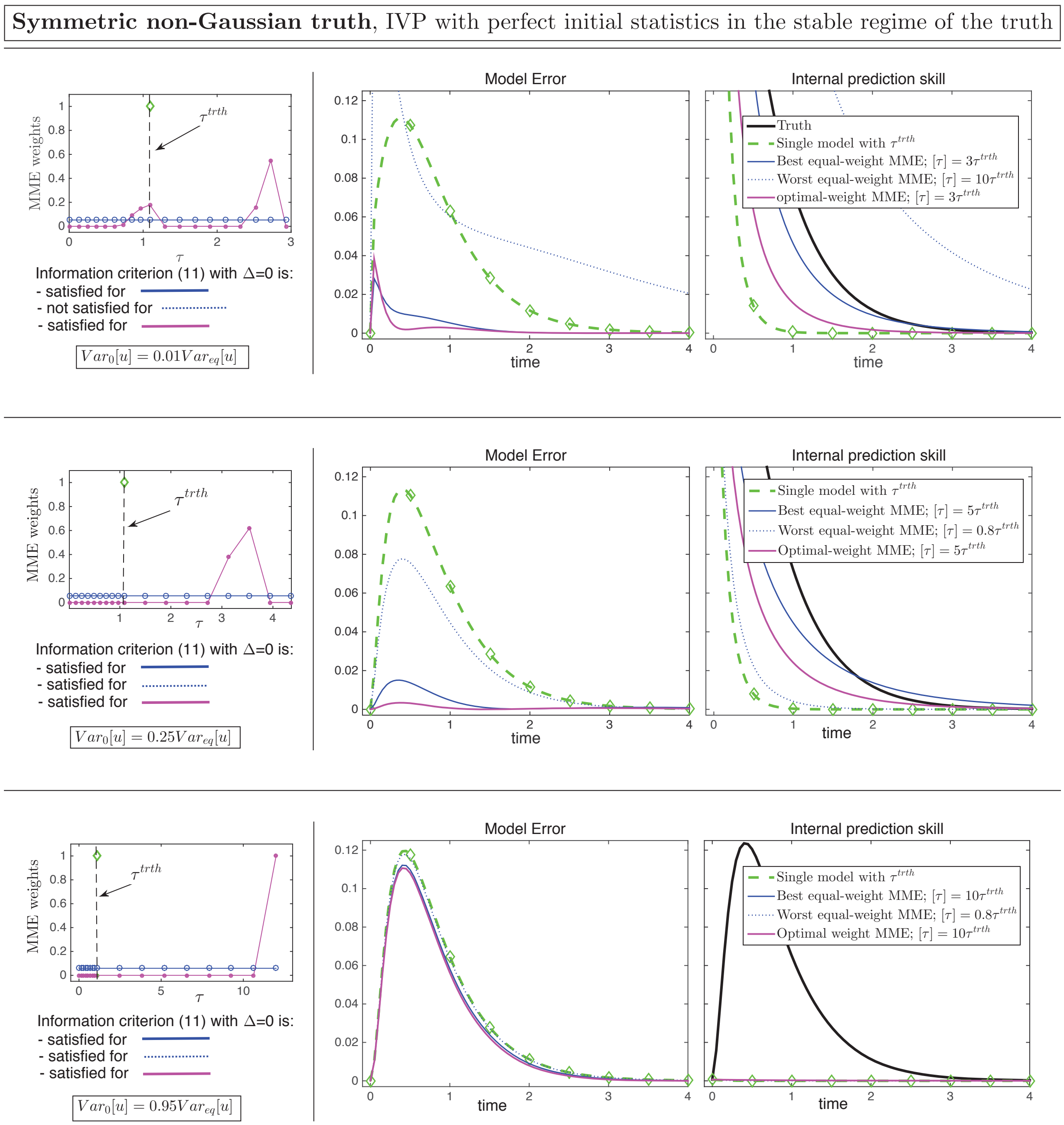

Figure 4: Initial value problem (see electronic version for colors). Prediction skill of a 17-model MME with correct statistical initial conditions for the resolved dynamics $u(t)$ of the non-Gaussian model (36) for different uncertainties $Var_0[u]$. The truth dynamics (36) is initiated in a statistically stable regime (the unresolved variable at t_0 satisfies $\gamma_0 \sim \mathcal{N}\big(1.8\langle\gamma\rangle_{eq}, 0.2Var_{eq}[\gamma]\big)$. The MME is a mixture of Gaussian models (24) with correct equilibrium statistics (44) and correlation times $\tau^{\mathrm{M}i}$ sampled around the correct correlation time τ^{trth} (see *balanced MME* in (48)) with the spread $[\tau]$ in $\pi^{\mathrm{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ defined in (46); the optimal-weight MME (magenta) is obtained by minimizing the relative entropy as in (9). *Truth parameters*: $\hat{\gamma} = 1.5, d_\gamma = 2, \sigma_\gamma = 2, \sigma_u = 0.5, F = 0$. *Initial conditions* (both truth and MME): $\langle u\rangle_0 = 0.4\langle u\rangle_{eq}$, $\langle\gamma\rangle_0 = 1.2\langle\gamma\rangle_{eq}$, $Var_0[\gamma] = 0.2Var_{eq}[\gamma]$ and $Var_0[u] = 0.01Var_{eq}[u]$ (top), $Var_0[u] = 0.25Var_{eq}[u]$ (middle), $Var_0[u] = 0.95Var_{eq}[u]$ (bottom).

Symmetric non-Gaussian truth, IVP with initial statistics in the unstable regime of the truth

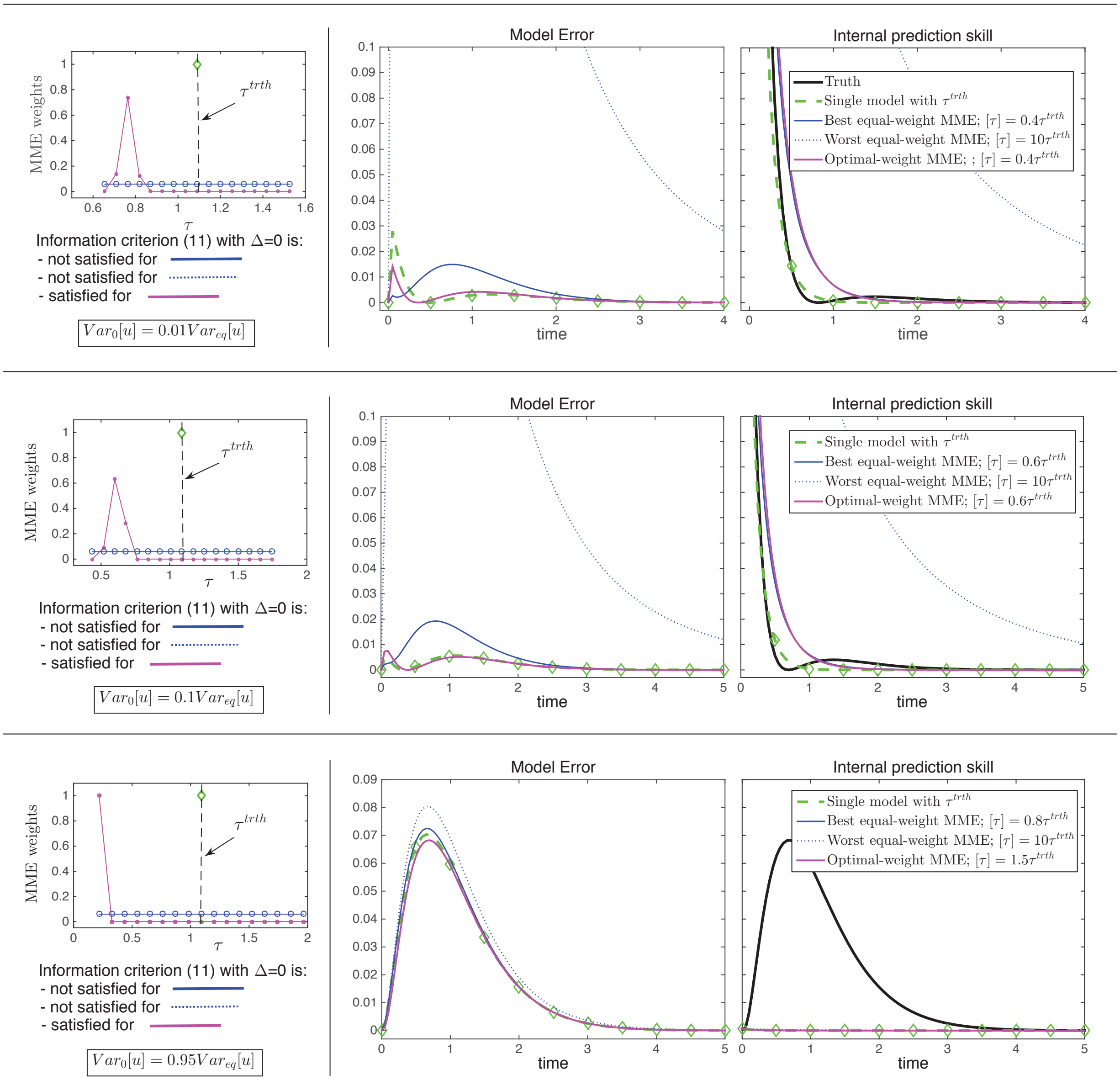

Figure 5: Initial value problem (see electronic version for colors). Prediction skill of a 17-model MME with correct statistical initial conditions for the resolved dynamics $u(t)$ of the non-Gaussian model (36) for different uncertainties $Var_0[u]$. The truth dynamics (36) is initiated in a statistically stable regime (the unresolved variable at t_0 satisfies $\gamma_0 \sim \mathcal{N}\big(-1.2\langle\gamma\rangle_{eq}, 0.2Var_{eq}[\gamma]\big)$. The MME is a mixture of Gaussian models (24) with correct equilibrium statistics (44) and correlation times τ^{M_i} sampled around the correct correlation time τ^{trth} (see *balanced MME* in (48)) with the spread $[\tau]$ in $\pi^{\mathrm{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ defined in (46); the optimal-weight MME (magenta) is obtained by minimizing the relative entropy as in (9). *Truth parameters*: $\hat{\gamma} = 1.5, d_\gamma = 2, \sigma_\gamma = 2, \sigma_u = 0.5, F = 0$. *Initial conditions* (both truth and MME): $\langle u\rangle_0 = 0.4\langle u\rangle_{eq}$, $\langle\gamma\rangle_0 = 1.2\langle\gamma\rangle_{eq}$, $Var_0[\gamma] = 0.2Var_{eq}[\gamma]$ and $Var_0[u] = 0.01Var_{eq}[u]$ (top), $Var_0[u] = 0.1Var_{eq}[u]$ (middle), $Var_0[u] = 0.95Var_{eq}[u]$ (bottom).

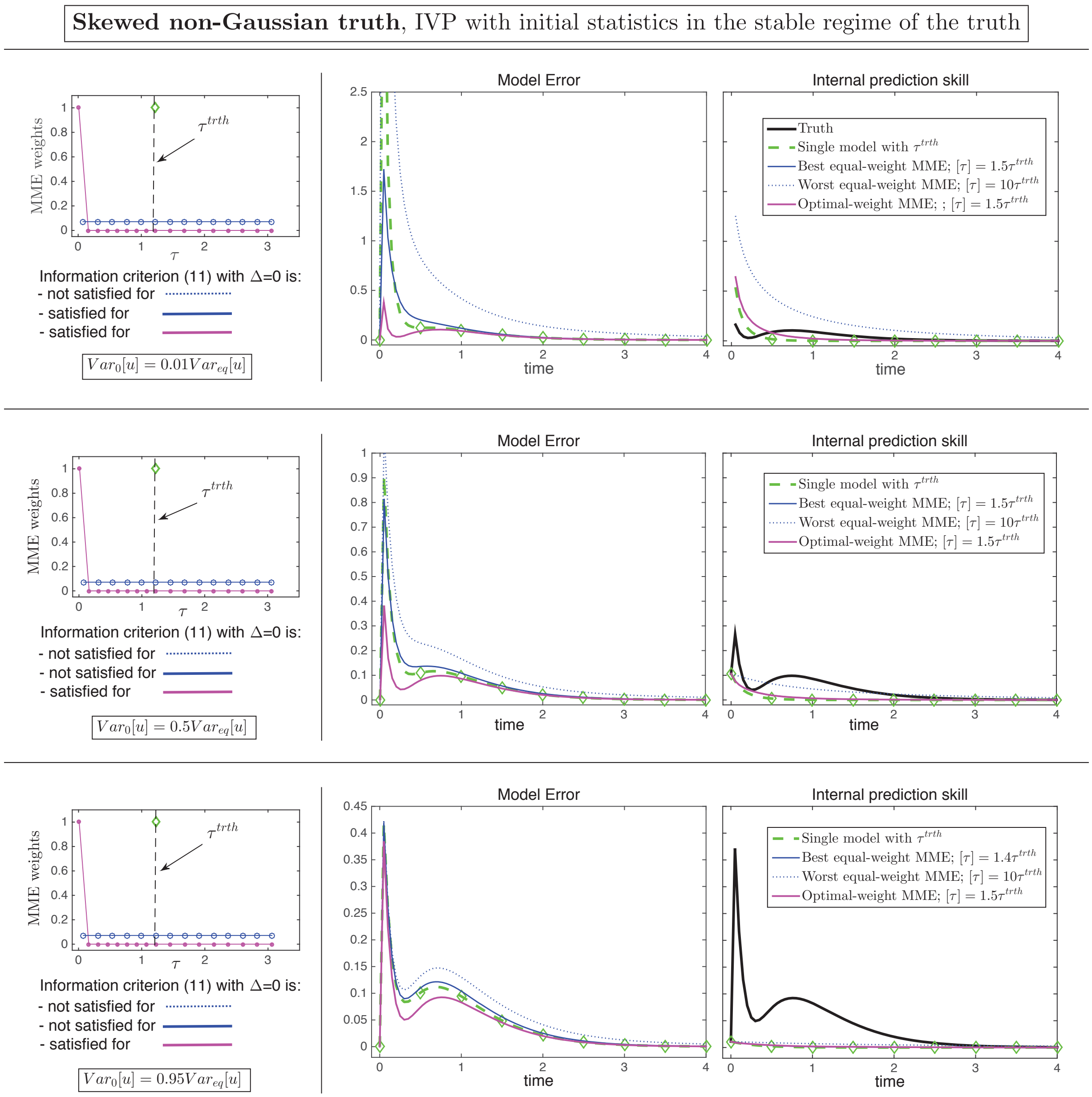

Figure 6: Initial value problem (see electronic version for colors). Prediction skill of a 17-model MME with correct statistical initial conditions for the resolved dynamics $u(t)$ of the non-Gaussian model (36) for different uncertainties $Var_0[u]$. The truth dynamics (36) is initiated in a statistically stable regime with the unresolved variable at t_0 satisfying $\gamma_0 \sim \mathcal{N}\big(\langle\gamma\rangle_{eq}, Var_{eq}[\gamma]\big)$. The MME is a mixture of Gaussian models (24) with correct equilibrium statistics (44) and correlation times τ^{M_i} sampled around the correct correlation time τ^{trth} (see *balanced MME* in (48)) with the spread $[\tau]$ in $\pi^{\mathrm{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ defined in (46); the optimal-weight MME (magenta) is obtained by minimizing the relative entropy as in (9). *Truth parameters*: $\hat{\gamma} = 1.5, d_\gamma = 10, \sigma_\gamma = 2, \sigma_u = 2, F = 1$. *Initial conditions* (both truth and MME): $\langle u\rangle_0 = 0.1\langle u\rangle_{eq}$, $\langle\gamma\rangle_0 = \langle\gamma\rangle_{eq}$, $Var_0[\gamma] = Var_{eq}[\gamma]$ and $Var_0[u] = 0.01Var_{eq}[u]$ (top), $Var_0[u] = 0.5Var_{eq}[u]$ (middle), $Var_0[u] = 0.95Var_{eq}[u]$ (bottom).

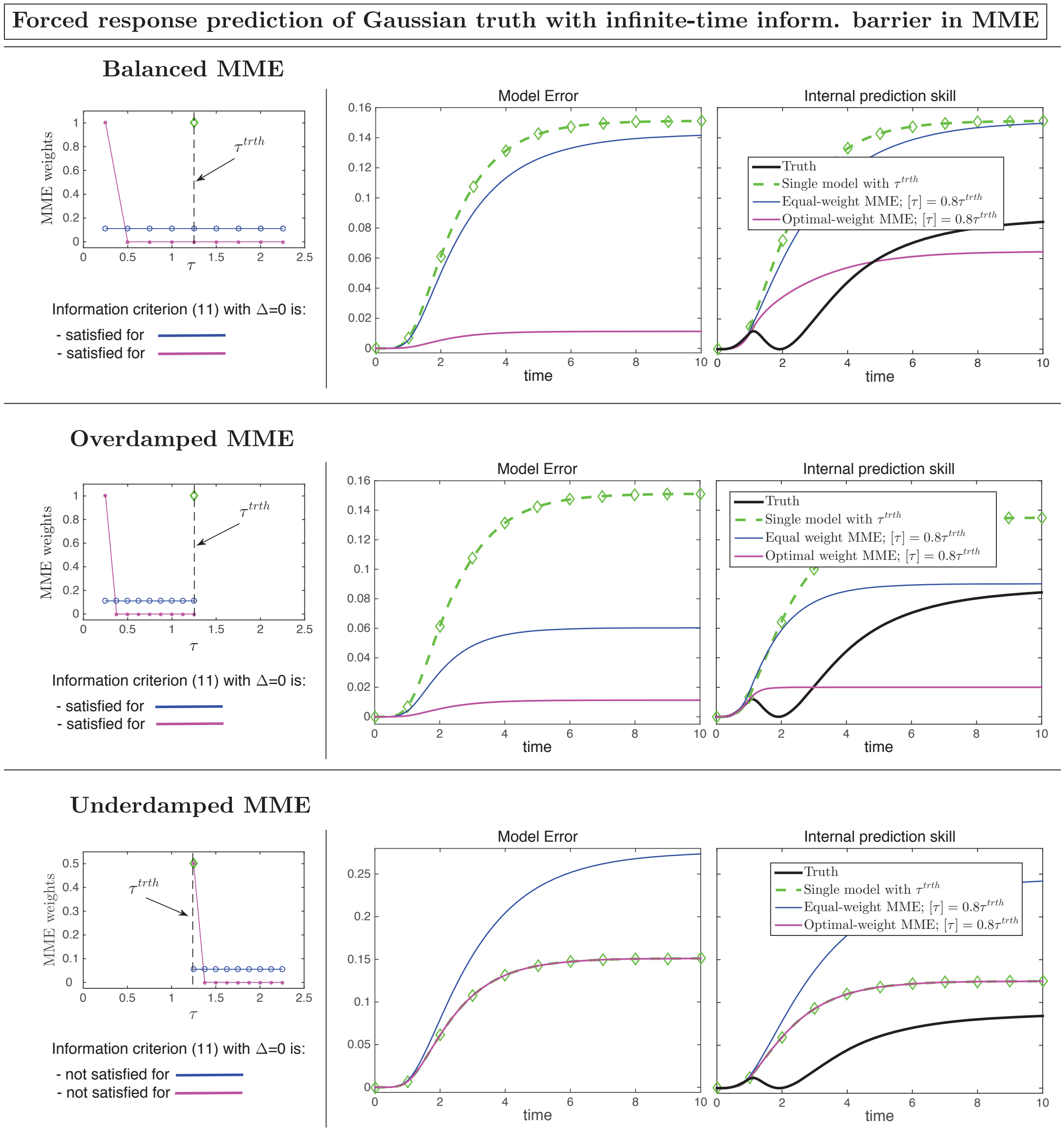

Figure 7: (See electronic version for colors) Prediction skill of a forced response of the resolved truth dynamics (32) based on a 17-model MME with correct statistical initial conditions for the resolved Gaussian dynamics $u(t)$ in (32) and three different types of MME structure (see (48)). The MME is a mixture of Gaussian models (24) with correct equilibrium statistics (44) and correlation times τ^{M_i} sampled to create a *balanced MME* (top), *overdamped MME* (middle), and *underdamped MME* (bottom) in (48)) with the spread $[\tau]$ in the MME density $\pi^{\mathrm{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ (46); the ensemble of initial conditions for the unresolved dynamics $v(t)$ in (32) is drawn from the unperturbed equilibrium $v_0 \sim \mathcal{N}\big(\langle v\rangle_{eq}, Var_{eq}[v]\big)$. The optimal-weight MME (magenta) is obtained by minimizing the relative entropy as in (9). *Truth parameters in (32)*: $A=0.5, a=-5.5, \lambda_{1,2}=-1,-4; \sigma: 0.63, F_0: -0.8,\ E=0.01, \langle u\rangle_{eq}=0.1, \langle v\rangle_{eq}=1.35$. *Forcing (both truth and MME)*: $F(t)=F_0$ for $t\leqslant 0$, $F(t)=F_0(1+0.05t)$ for $0<t\leqslant 1$, $F(t)=F_0(1+0.05)$ for $t>1$.

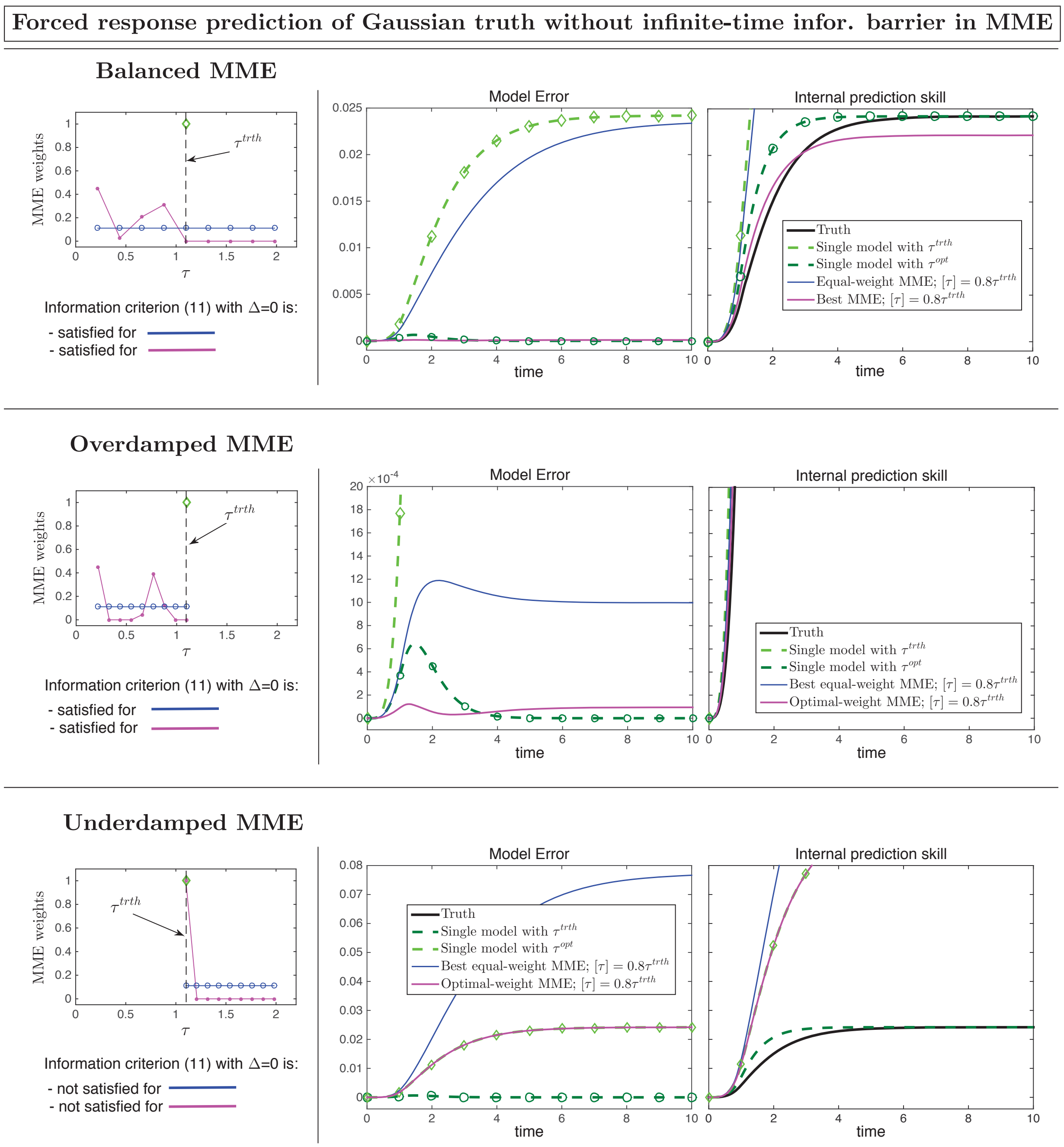

Figure 8: (See electronic version for colors.) Prediction skill of a forced response of the resolved truth dynamics (32) based on a 17-model MME with correct statistical initial conditions for the resolved Gaussian dynamics $u(t)$ in (32) and three different types of MME structure (see (48)). The MME is a mixture of Gaussian models (24) with correct equilibrium statistics (44) and correlation times τ^{M_i} sampled to create a *balanced MME* (top), *overdamped MME* (middle), and *underdamped MME* (bottom) in (48)) with the spread $[\tau] = 0.8\tau^{trth}$ in the MME density $\pi^{\mathrm{MME}}_{t;\boldsymbol{\alpha},[\tau]}$ (46); the ensemble of initial conditions for the unresolved dynamics $v(t)$ in (32) is drawn from the unperturbed equilibrium $v_0 \sim \mathcal{N}\big(\langle v\rangle_{eq}, Var_{eq}[v]\big)$. The optimal-weight MME (magenta) is obtained by minimizing the relative entropy as in (9). The correlation time τ^{opt} for the single model with perfect infinite-time response (green/circles) is given by (40) in §4.2. *Truth parameters in (32)*: $A = -5.5, a = -5.5, \lambda_{1,2} = -1, -10; \sigma : 1.48, F_0 : 0.18, E = 0.01, \langle u\rangle_{eq} = 0.1, \langle v\rangle_{eq} = 0.37$. *Forcing (both truth and MME)*: $F(t) = F_0$ for $t \leqslant 0$, $F(t) = F_0(1 + 0.05t)$ for $0 < t \leqslant 1$, $F(t) = F_0(1 + 0.05)$ for $t > 1$.

5 Conclusions

Here, we developed a framework rooted in information theory for a systematic assessment of the skill of *Multi Model Ensemble* (MME) approach which is aimed at improving the accuracy of dynamical predictions through combining probabilistic forecasts obtained from imperfect models. Despite the increasingly common use of the MME approach in applied sciences, especially in the climate and atmospheric sciences (e.g., [60, 67, 13, 70, 71, 68, 69]), a systematic framework justifying this technique was lacking. Consequently, many procedures developed in the context of MME prediction lack systematic guidelines for constructing model ensembles with improved predictive skill. Here, we focused on uncertainty quantification and a systematic understanding of the benefits and limitations of the MME approach, as well as on the development of practical design principles for constructing model ensembles with improved predictive skill. This setting should not be confused with statistical modelling in which the underlying dynamics is ignored. The main issues and results presented here focused on:

(I) The advantages/disadvantages of the MME approach relative to using a single model predictions with an ensemble of initial conditions. In particular, we derived the sufficient condition guaranteeing improvement of the skill (*time-dependent statistical accuracy*) of dynamic MME predictions relative to the single model predictions (see (11) and §3).

(II) Sensitivity of the MME skill to the unresolved truth dynamics, and guidelines for constructing MME for best prediction skill at short, medium and long time ranges (see §3, §4.3 and Appendices B, **??**). Here, a systematic insight to improving non-equilibrium prediction could be made by combining the linear response theory and fluctuation-dissipation formulas, which utilize only the unperturbed equilibrium information, within a time-dependent information theoretic framework.

Based on the information-theoretic considerations, we derived a simple condition (7) which guarantees improvement improvement of probabilistic predictions within the MME framework; this criterion uses the relative entropy (2) which measures lack of information in the dynamic predictions based on imperfect models relative to the truth dynamics on the subspace of the resolved observables. We showed for the first time why, and under what conditions, combining imperfect models results in an improved predictive performance relative to the best single imperfect model in the ensemble; the potential benefits of the MME approach result from the convexity of the relative entropy. The sufficient condition (11) for improving imperfect dynamical predictions relative to the single model predictions represents a more practical way of assessing the utility of the MME approach since it involves estimating the lack of information in predictions of the individual models in the ensemble rather than determining the lack of information in the full mixture density associated with the MME prediction.

We showed that the condition (7) for MME skill improvement can be practically implemented in the relaxed form (11) or (14) which require evaluation of the lack of information between the individual ensemble members and the least-biased estimates of the truth dynamics. This condition can be evaluated with the help of the linear response theory and the 'fluctuation-dissipation' approach (see, e.g., [47, 35, 2, 41, 23, 50, 39])) in the context of forced-response prediction when the truth equilibrium dynamics is subjected to external perturbations in forced dissipative systems; this approach follows from the earlier work in the single-model setup (see, for example, [32, 42, 33, 49, 40, 50, 18, 8, 48]). When considering prediction improvement via the MME approach for the initial value problem the implementation of the information-based condition (11) or (14) can be carried out in the hindcast/reanalysis mode; moreover techniques similar to those discussed in [19, 20, 21] could be used to effectively assess the skill of a given ensemble of imperfect models. A set of useful results was derived in §3.2 in the Gaussian framework which utilizes Gaussian models in the Multi Model Ensembles; this approach provides useful intuition and guidelines in more complex cases treated in §3. The general theoretical results were illustrated in §4.3 which combines the analytical estimates of §3 with simple numerical tests based on statistically exactly solvable Gaussian and non-Gaussian test models described in §4.1.

The ultimate goal in reduced-order prediction should involve a synergistic approach that combines MME forecasting, data assimilation [16, 52, 6, 9], and improving individual models through various stochastic superparameterization [24, 51], and reduced subspace closure techniques [64, 65, 63]. We envisage extending the present framework to account for differences in the internal prediction skill (50) of

MME and the single imperfect model in addition to the prediction error (49). The natural and important extension of this work involves combining the MME framework for improving imperfect predictions with an MME approach to data assimilation/filtering in high-dimensional turbulent systems based on imperfect models. Such a combined framework should provide a valuable tool for improving real-time predictions in complex partially observed dynamical systems.

Acknowledgements

M.B. is supported as a postdoctoral fellow on the ONR DRI grant of A.J.M: N0014-10-1-0554. The research of A.J.M. is partially supported by National Science Foundation CMG grant DMS-1025468 and the Office of Naval Research grants ONR DRI N0014-10-1-0554, N00014-11-1-0306, and the MURI award ONR-MURI N00014-12-1-0912.

A Some simple proofs of general results from §3

Here, we complement the discussion of §3 by providing simple derivations and proofs of the facts established in that section.

Information Criterion II in (11): Derivation of this criterion relies on the convexity properties of the relative entropy (e.g., [11]), which leads to the following upper bound on the lack of information in the MME mixture density π_t^{MME} (1) relative to the marginal truth density π_t

$$\mathcal{P}(\pi_t,\pi_t^{\mathrm{MME}})=\mathcal{P}(\pi_t,\textstyle\sum_i\alpha_i\pi_t^{\mathrm{M}_i})\leqslant\sum_i\alpha_i\mathcal{P}(\pi_t,\pi_t^{\mathrm{M}_i})=\sum_{i\neq\diamond}\alpha_i\mathcal{P}(\pi_t,\pi_t^{\mathrm{M}_i})+\alpha_\diamond\mathcal{P}(\pi_t,\pi_t^{\mathrm{M}_\diamond}).\tag{52}$$

where $\alpha_i\geqslant 0$ and $\sum_i\alpha_i=1$ so that

$$\mathcal{P}(\pi_t,\pi_t^{\mathrm{MME}})-\mathcal{P}(\pi_t,\pi_t^{\mathrm{M}_\diamond})=\mathcal{P}(\pi_t^{\mathrm{L}},\pi_t^{\mathrm{MME}})-\mathcal{P}(\pi_t^{\mathrm{L}},\pi_t^{\mathrm{M}_\diamond})\leqslant\textstyle\sum_{i\neq\diamond}\alpha_i\mathcal{P}(\pi_t^{\mathrm{L}},\pi_t^{\mathrm{M}_i})+(\alpha_\diamond-1)\mathcal{P}(\pi_t^{\mathrm{L}},\pi_t^{\mathrm{M}_\diamond}),\tag{53}$$

where we used the triangle equality (6) and the fact $\mathcal{P}\geqslant 0$. Clearly, the information criterion in (7) is always satisfied when the right-hand-side in (53) satisfies

$$\sum_{i\neq\diamond}\alpha_i\,⨏_{\mathcal{I}}\mathcal{P}(\pi_t^{\mathrm{L}},\pi_t^{\mathrm{M}_i})\mathrm{d}t+(\alpha_\diamond-1)⨏_{\mathcal{I}}\mathcal{P}(\pi_t^{\mathrm{L}},\pi_t^{\mathrm{M}_\diamond})\mathrm{d}t<0,\tag{54}$$

which, after rearranging terms, gives the sufficient condition in (11) with $\Delta=0$, i.e.,

$$\sum_{i\neq\diamond}\beta_i\,\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\pi^{\mathrm{M}_i})<\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\pi^{\mathrm{M}_\diamond}),\qquad\beta_i=\alpha_i\,(1-\alpha_\diamond)^{-1},\quad\sum_{i\neq\diamond}\beta_i=1.\tag{55}$$

This sufficient condition is too restrictive when $\mathrm{M}_\diamond$ coincides with the best imperfect model $\mathrm{M}^*_{\mathcal{I},\mathrm{L}}$ in (8) since there is no non-trivial MME satisfying (54). Based on the basic convexity properties illustrated in figure 2 and FACT 1, unless $\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\pi^{\mathrm{M}^*_{\mathcal{I},\mathrm{L}}})=0$, an MME with a smaller error does exist and the condition (55) needs to be relaxed in order to be applicable in such cases. The uncertainty parameter Δ in

$$\sum_{i\neq\diamond}\beta_i\,\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\pi^{\mathrm{M}_i})<\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\pi^{\mathrm{M}_\diamond})+\Delta,\qquad\beta_i=\alpha_i\,(1-\alpha_\diamond)^{-1},\quad\sum_{i\neq\diamond}\beta_i=1,\tag{56}$$

allows for including models in the ensemble with error $\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\pi^{\mathrm{M}^*_{\mathcal{I},\mathrm{L}}})\leqslant\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\pi^{\mathrm{M}})<\mathcal{P}_{\mathcal{I}}(\pi^{\mathrm{L}},\pi^{\mathrm{M}^*_{\mathcal{I},\mathrm{L}}})+\Delta$ so that the MME prediction error is $0\leqslant\mathcal{P}_{\mathcal{I}}(\pi,\pi^{\mathrm{MME}})\leqslant\mathcal{P}_{\mathcal{I}}(\pi,\pi^{\mathrm{M}_\diamond})+\Delta$, as illustrated in figure 2.

Proof of FACT 3: The proof is straightforward and follows by direct calculation consisting of two steps:

1) We start by rewriting the condition (11) in terms of the least-biased densities defined in (12) which leads to

$$\mathcal{P}(\pi_t^{\mathrm{L}1},\pi_t^{\mathrm{M}_\diamond,\mathrm{L}2})+\Delta>\sum_{i\neq\diamond}\beta_i\,\mathcal{P}(\pi_t^{\mathrm{L}1},\pi_t^{\mathrm{M}_i,\mathrm{L}2})+\sum_{i\neq\diamond}\beta_i\,\mathbb{E}^{\pi^{\mathrm{L}1}}\Big[\log\frac{\pi^{\mathrm{M}_i,\mathrm{L}2}}{\pi^{\mathrm{M}_i}}-\log\frac{\pi^{\mathrm{M}_\diamond,\mathrm{L}2}}{\pi^{\mathrm{M}_\diamond}}\Big];\tag{57}$$

note that this last term vanishes identically when $\pi^{\mathrm{M}_i,\mathrm{L}2}=\pi^{\mathrm{M}_i}$ and the MME contains only least-biased models.

2) Next, we notice that the relative entropy between two least biased densities $\pi_t^{\mathrm{L}_1}$ and $\pi_t^{\mathrm{M},\mathrm{L}_2}$ is given by

$$\mathcal{P}(\pi_t^{\mathrm{L}_1},\pi_t^{\mathrm{M},\mathrm{L}_2}) = \log C_t^{\mathrm{M}} + \boldsymbol{\theta}_t^{\mathrm{M}}\cdot\bar{\boldsymbol{E}}_t - \big(\log C_t + \boldsymbol{\theta}_t\cdot\bar{\boldsymbol{E}}_t\big) = \log\frac{C_t^{\mathrm{M}}}{C_t} + (\boldsymbol{\theta}_t^{\mathrm{M}} - \boldsymbol{\theta}_t)\cdot\bar{\boldsymbol{E}}_t, \tag{58}$$

where $\overline{\boldsymbol{E}}_t$ is the vector of expectations of the functionals E_i defined in (13) with respect to the truth marginal density π_t, and the Lagrange multipliers in (12), $\boldsymbol{\theta}_t = \boldsymbol{\theta}\big(\overline{\boldsymbol{E}}_t\big)$, $\boldsymbol{\theta}_t^{\mathrm{M}} = \boldsymbol{\theta}^{\mathrm{M}}\big(\overline{\boldsymbol{E}}_t^{\mathrm{M}}\big)$, are defined as

$$\boldsymbol{\theta}_t = \begin{cases}(\theta_1(t),\dots,\theta_{\mathrm{L}_1}(t))^{\mathrm{T}},\\ (\theta_1(t),\dots,\theta_{\mathrm{L}_1}(t),0,\dots,0_{\mathrm{L}_2})^{\mathrm{T}},\end{cases}\qquad \boldsymbol{\theta}_t^{\mathrm{M}} = \begin{cases}(\theta_1^{\mathrm{M}}(t),\dots,\theta_{\mathrm{L}_2}^{\mathrm{M}}(t),0,\dots,0_{\mathrm{L}_1})^{\mathrm{T}}, & \text{if}\quad \mathrm{L}_1\geqslant \mathrm{L}_2,\\ (\theta_1^{\mathrm{M}}(t),\dots,\theta_{\mathrm{L}_2}^{\mathrm{M}}(t))^{\mathrm{T}}, & \text{if}\quad \mathrm{L}_1 < \mathrm{L}_2.\end{cases}$$

while the normalization constants in the least biased densities are $C_t = C\big(\overline{\boldsymbol{E}}_t\big)$, $C_t^{\mathrm{M}} = C^{\mathrm{M}}\big(\overline{\boldsymbol{E}}_t^{\mathrm{M}}\big)$.

The condition in (14) is obtained by combining (57) with (58).

Proof of FACT 4: The condition in (19) for improvement of the prediction skill via MME in the context of initial value problem can be obtained as follows: Consider the representation of the true expected values $\overline{\boldsymbol{E}}_t$ of the functionals $E_i(\boldsymbol{u})$ with respect to the truth marginal density $\pi_t(\boldsymbol{u})$ in the form

$$\overline{\boldsymbol{E}}_t = \overline{\boldsymbol{E}}_0 + \delta\tilde{\overline{\boldsymbol{E}}}_t, \qquad \boldsymbol{\theta}_t = \boldsymbol{\theta}_0 + \delta\tilde{\boldsymbol{\theta}}_t\big(\overline{\boldsymbol{E}}_t\big), \qquad \tilde{\overline{\boldsymbol{E}}}_{t=0} = \tilde{\boldsymbol{\theta}}_{t=0} = 0, \tag{59}$$

these are smooth at $\delta = 0$ when the decomposition $\pi_t = \pi_0 + \delta\tilde{\pi}_t$ is smooth at $\delta = 0$ which holds under minimal hypothesis described in [26] so that

$$C_t = C_0\big(1 - \delta\tilde{\boldsymbol{\theta}}_t\cdot\overline{\boldsymbol{E}}_0\big) + \mathcal{O}\big(\delta^2\big), \tag{60}$$

The lack of information in (11) between the least-biased approximation of the truth $\pi_t^{\mathrm{L}_1}$ and the imperfect model density $\pi_t^{\mathrm{M}_i}$ can be written as

$$\mathcal{P}(\pi_t^{\mathrm{L}_1},\pi_t^{\mathrm{M}_i}) = \mathcal{P}(\pi_t^{\mathrm{L}_1},\pi_t^{\mathrm{M}_i,\mathrm{L}_2}) + \int\!\mathrm{d}\boldsymbol{u}\,\pi_t^{\mathrm{L}_1}\log\frac{\pi_t^{\mathrm{M}_i,\mathrm{L}_2}}{\pi_t^{\mathrm{M}_i}}, \tag{61}$$

similarly to the result leading to (57). The lack of information in the perturbed least-biased density, $\pi_t^{\mathrm{M}_i,\mathrm{L}_2}$, of the imperfect model relative to the least-biased perturbation of the truth, $\pi_t^{\mathrm{L}_1}$, can be expressed through (63)-(66) in the following form

$$\mathcal{P}(\pi_t^{\mathrm{L}_1},\pi_t^{\mathrm{M}_i,\mathrm{L}_2}) = \mathcal{P}(\pi_0^{\mathrm{L}_1},\pi_t^{\mathrm{M}_i,\mathrm{L}_2}) + (\boldsymbol{\theta}_t^{\mathrm{M}_i} - \boldsymbol{\theta}_0)\cdot\delta\tilde{\overline{\boldsymbol{E}}}_t + \mathcal{O}(\delta^2). \tag{62}$$

Substituting (62) into (14) leads to the desired condition (19).□

Proof of FACT 5: The condition in (23) for improvement of the prediction skill via MME obtained by perturbing single model predictions can be obtained as follows: Consider the condition (14) in the case when the ensemble members $\mathrm{M}_i\in\mathcal{M}$ are obtained from the single model $\mathrm{M}_\diamond\in\mathcal{M}$ through perturbing some parameters of the single model; we assume that the statistics of the model depends smoothly on these parameters and that the perturbations are non-singular (which required minimal assumptions [26] of hypoelliptic noise in the truth dynamics) so that the evolution of the statistical moments $\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}$ and their functions in the least-biased densities (58) of the ensemble members can be written, for $\epsilon \ll 1$, as

$$a)\;\; \overline{\boldsymbol{E}}_t^{\mathrm{M}_i,\epsilon} = \overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond} + \epsilon\,\tilde{\overline{\boldsymbol{E}}}_t^{\mathrm{M}_i}, \tag{63}$$

$$b)\;\; \boldsymbol{\theta}_t^{\mathrm{M}_i,\epsilon} = \boldsymbol{\theta}_t^{\mathrm{M}_\diamond} + \epsilon\,\tilde{\boldsymbol{\theta}}_t^{\mathrm{M}_i}\big(\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}\big) + \mathcal{O}\big(\epsilon^2\big), \tag{64}$$

$$c)\;\; C_t^{\mathrm{M}_i,\epsilon} = C_t^{\mathrm{M}_\diamond}\big(1 - \epsilon\,\tilde{\boldsymbol{\theta}}_t^{\mathrm{M}_i}\cdot\overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond}\big) + \mathcal{O}\big(\epsilon^2\big), \tag{65}$$

where

$$\tilde{\boldsymbol{\theta}}_t^{\mathrm{M}_i} = \Big(\overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond} \cdot \nabla\theta_1^{\mathrm{M}_i}|_{\varepsilon=0},\ \overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond} \cdot \nabla\theta_2^{\mathrm{M}_i}|_{\varepsilon=0}, \dots,\ \overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond} \cdot \nabla\theta_{\mathrm{L}_1}^{\mathrm{M}_i}|_{\varepsilon=0}\Big)^{\mathrm{T}}. \tag{66}$$

The lack of information in the perturbed least-biased density, $\pi_t^{\mathrm{M}_i,\mathrm{L}_2}$, of the imperfect model relative to the least-biased perturbation of the truth, $\pi_t^{\mathrm{L}_1}$, can be expressed through (63)-(66) in the following form

$$\mathcal{P}(\pi_t^{\mathrm{L}_1}, \pi_t^{\mathrm{M}_i,\mathrm{L}_2}) = \log\big(C_t^{\mathrm{M}_i}/C_t\big) + (\boldsymbol{\theta}_t^{\mathrm{M}_i} - \boldsymbol{\theta}_t)\cdot\bar{\boldsymbol{E}}_t = \mathcal{P}(\pi_t^{\mathrm{L}_1}, \pi_t^{\mathrm{M}_\diamond}) + \epsilon\,\tilde{\boldsymbol{\theta}}_t^{\mathrm{M}_i}\cdot(\overline{\boldsymbol{E}}_t - \overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond}) + \mathcal{O}\big(\epsilon^2\big), \tag{67}$$

which is obtained by combining (63)-(65). Substituting (67) into the general condition (14) leads to the desired condition (23). □

Proof of FACT 6: The proof of the condition (29) is simple but tedious and follows from the short-time asymptotic expansion of the relative entropy between the Gaussian truth and the Gaussian models. Consider the state vector $\boldsymbol{u}\in I\!R^K$ for resolved dynamics and assume that short-times the statistics of the Gaussian truth density $\pi_t^{\mathrm{G}} = \mathcal{N}(\boldsymbol{\mu}_t, R_t)$ and of the Gaussian model density $\pi_t^{\mathrm{M}_i} = \mathcal{N}(\boldsymbol{\mu}_t^{\mathrm{M}_i}, R_t^{\mathrm{M}_i})$ are

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \delta\,\tilde{\boldsymbol{\mu}}_t, \qquad R_t = R_0 + \delta\tilde{R}_t, \qquad \delta\tilde{\boldsymbol{\mu}}_0 = \delta\tilde{R}_0 = 0, \tag{68}$$

and

$$\boldsymbol{\mu}_t^{\mathrm{M}_i} = \boldsymbol{\mu}_0^{\mathrm{M}_i} + \delta\,\tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i}, \qquad R_t^{\mathrm{M}_i} = R_0^{\mathrm{M}_i} + \delta\tilde{R}_t^{\mathrm{M}_i}, \qquad \delta\tilde{\boldsymbol{\mu}}_0^{\mathrm{M}_i} = \delta\tilde{R}_0^{\mathrm{M}_i} = 0. \tag{69}$$

Then, the relative entropy between the Gaussian truth density π_t^{G} and a Gaussian model density $\pi_t^{\mathrm{M}_i}$

$$\mathcal{P}(\pi_t^{\mathrm{G}}, \pi_t^{\mathrm{M}_i}) = \tfrac{1}{2}(\Delta\boldsymbol{\mu}_t^i)^T(R_t^{\mathrm{M}_i})^{-1}\Delta\boldsymbol{\mu}_t^i + \tfrac{1}{2}\Big[\mathrm{tr}\big[R_t(R_t^{\mathrm{M}_i})^{-1}\big] - \ln\det\big[R_t(R_t^{\mathrm{M}_i})^{-1}\big] - K\Big], \tag{70}$$

with $\Delta\boldsymbol{\mu}_t^i := \boldsymbol{\mu}_t - \boldsymbol{\mu}_t^{\mathrm{M}_i}$ can be expressed as

$$\mathcal{P}(\pi_t^{\mathrm{G}}, \pi_t^{\mathrm{M}_i}) = \mathcal{P}(\pi_0^{\mathrm{G}}, \pi_0^{\mathrm{M}_i}) + \delta(X^\mu + X^R) + \delta^2(Y^\mu + Y^{\mu,R} + Y^{R,R}) + \mathcal{O}(\delta^3), \tag{71}$$

which is valid at times short enough so that the changes in moments $\delta\tilde{\boldsymbol{\mu}}, \delta\tilde{R}$, $\delta\tilde{\boldsymbol{\mu}}^{\mathrm{M}_i}, \delta\tilde{R}^{\mathrm{M}_i}$ are small; the respective coefficients in (71) are given by

$$X^\mu = \tfrac{1}{2}\Big[(\Delta\boldsymbol{\mu}_0^i)^T(R_0^{\mathrm{M}_i})^{-1}\Delta\tilde{\boldsymbol{\mu}}_t^i + (\Delta\tilde{\boldsymbol{\mu}}_t^i)^T(R_0^{\mathrm{M}_i})^{-1}\Delta\boldsymbol{\mu}_0^i\Big],$$

$$X^R = -\tfrac{1}{2}(\Delta\boldsymbol{\mu}_0^i)^T(R_0^{\mathrm{M}_i})^{-1}\tilde{R}_t^{\mathrm{M}_i}(R_0^{\mathrm{M}_i})^{-1}\Delta\boldsymbol{\mu}_0^i + \tfrac{1}{2}\mathrm{tr}\Big[(\mathcal{I} - R_0(R_0^{\mathrm{M}_i})^{-1})\tilde{R}_t^{\mathrm{M}_i}(R_0^{\mathrm{M}_i})^{-1}\Big] + \tfrac{1}{2}\mathrm{tr}\Big[\tilde{R}_t(R_0^{\mathrm{M}_i})^{-1}\Big],$$

$$Y^{\mu,\mu} = \tfrac{1}{2}(\Delta\tilde{\boldsymbol{\mu}}_t^i)^T(R_0^{\mathrm{M}_i})^{-1}\Delta\tilde{\boldsymbol{\mu}}_t^i,$$

$$Y^{\mu,R} = \tfrac{1}{2}\Big[(\Delta\boldsymbol{\mu}_0^i)^T(R_0^{\mathrm{M}_i})^{-1}\tilde{R}_t^{\mathrm{M}_i}(R_0^{\mathrm{M}_i})^{-1}\Delta\tilde{\boldsymbol{\mu}}_t^i + (\Delta\tilde{\boldsymbol{\mu}}_t^i)^T(R_0^{\mathrm{M}_i})^{-1}\tilde{R}_t^{\mathrm{M}_i}(R_0^{\mathrm{M}_i})^{-1}\Delta\boldsymbol{\mu}_0^i\Big],$$

$$\begin{aligned} Y^{R,R} = {} & \tfrac{1}{2}\Big[(\Delta\boldsymbol{\mu}_0^i)^T(R_0^{\mathrm{M}_i})^{-1}\tilde{R}_t^{\mathrm{M}_i}(R_0^{\mathrm{M}_i})^{-1}\tilde{R}_t^{\mathrm{M}_i}(R_0^{\mathrm{M}_i})^{-1}\Delta\boldsymbol{\mu}_0^i \\ & - \tfrac{1}{2}\mathrm{tr}\Big[(\mathcal{I} - R_0(R_0^{\mathrm{M}_i})^{-1})(\tilde{R}_t^{\mathrm{M}_i}(R_0^{\mathrm{M}_i})^{-1})^2\Big] - \tfrac{1}{2}\mathrm{tr}\Big[\tilde{R}_t(R_0^{\mathrm{M}_i})^{-1}\tilde{R}_t^{\mathrm{M}_i}(R_0^{\mathrm{M}_i})^{-1}\Big] + \tfrac{1}{4}\Big(\mathrm{tr}\big[\tilde{R}_t^{\mathrm{M}_i}(R_0^{\mathrm{M}_i})^{-1}\big]\Big)^2. \end{aligned}$$

For correct initial conditions, $\boldsymbol{\mu}_0^{\mathrm{M}_i} = \boldsymbol{\mu}_0$, $R_0^{\mathrm{M}_i} = R_0$, the above formulas simplify to

$$X^R = \tfrac{1}{2}\mathrm{tr}\Big[\tilde{R}_t(R_0)^{-1}\Big], \quad Y^{\mu,\mu} = \tfrac{1}{2}(\Delta\tilde{\boldsymbol{\mu}}_t^i)^T(R_0)^{-1}\Delta\tilde{\boldsymbol{\mu}}_t^i, \tag{72}$$

$$Y^{R,R} = -\tfrac{1}{2}\mathrm{tr}\Big[\tilde{R}_t(R_0)^{-1}\tilde{R}_t^{\mathrm{M}_i}(R_0)^{-1}\Big] + \tfrac{1}{4}\Big(\mathrm{tr}\big[\tilde{R}_t^{\mathrm{M}_i}(R_0)^{-1}\big]\Big)^2, \tag{73}$$

with the remaining coefficients identically zero. Substituting the relative entropy between $\mathcal{P}(\pi_t^{\mathrm{G}}, \pi_t^{\mathrm{M}_i})$ in the form (71) with the coefficients (72)-(73) into the general necessary condition (11) for improving the prediction via MME yields the condition (29). □

Proof of FACT 7: We assume that the perturbations of the equilibrium truth and model densities are smooth in response to the forcing perturbations so that the perturbed densities $\pi_t^\delta = \pi_{eq} + \delta\tilde{\pi}_t$ are differentiable at $\delta = 0$; this holds under relatively mild assumptions hypoelliptic noise as shown in [26]. Thus, based on the linear response theory combined with the fluctuation-dissipation formulas (e.g., [47]), the density perturbations remain small for sufficiently small external perturbations which also implies that the moment perturbations remain small for all time. Derivation of the condition (76) relies on the smallness of the moment perturbations which allows for an asymptotic expansion of the relative entropy as in (71) but with $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_{eq} = \boldsymbol{\mu}_{eq}^{\mathrm{M}_i}$, $R_0 = R_{eq} = R_{eq}^{\mathrm{M}_i}$ which leads to expansion coefficients in (71)

$$X^R = \tfrac{1}{2}\mathrm{tr}\Big[\tilde{R}_t(R_{eq})^{-1}\Big], \quad Y^{\mu,\mu} = \tfrac{1}{2}(\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i})^T(R_{eq})^{-1}(\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i}), \tag{74}$$

$$Y^{R,R} = -\tfrac{1}{2}\mathrm{tr}\Big[\tilde{R}_t(R_{eq})^{-1}\tilde{R}_t^{\mathrm{M}_i}(R_{eq})^{-1}\Big] + \tfrac{1}{4}\Big(\mathrm{tr}\big[\tilde{R}_t^{\mathrm{M}_i}(R_{eq})^{-1}\big]\Big)^2, \tag{75}$$

with the remaining coefficients identically zero. The general condition for improvement of forced response prediction via MME in the Gaussian framework is

$$D_{\boldsymbol{\beta},\mathcal{I}}(\{\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}^{\mathrm{M}_i}\}) + E_{\boldsymbol{\beta},\mathcal{I}}(\{\tilde{R}^{\mathrm{M}_i}\}) + F_{\boldsymbol{\beta},\mathcal{I}}(\tilde{R}, \{\tilde{R}^{\mathrm{M}_i}\}) + \Delta + \mathcal{O}(\delta) > 0, \tag{76}$$

where

$$D_{\boldsymbol{\beta},\mathcal{I}} = \tfrac{1}{2}\sum_{i\neq\diamond}\frac{\alpha_i}{1-\alpha_\diamond}\fint_{\mathcal{I}}\mathrm{d}t\Big[(\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_\diamond})^T(R_{eq})^{-1}(\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_\diamond}) - (\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i})^T(R_{eq})^{-1}(\tilde{\boldsymbol{\mu}}_t - \tilde{\boldsymbol{\mu}}_t^{\mathrm{M}_i})\Big]$$

$$E_{\boldsymbol{\beta},\mathcal{I}} = \tfrac{1}{4}\sum_{i\neq\diamond}\frac{\alpha_i}{1-\alpha_\diamond}\fint_{\mathcal{I}}\mathrm{d}t\,\mathrm{tr}\Big[(\tilde{R}_t^{\mathrm{M}_\diamond} - \tilde{R}_t^{\mathrm{M}_i})(R_{eq})^{-1}\Big]\mathrm{tr}\Big[(\tilde{R}_t^{\mathrm{M}_\diamond} + \tilde{R}_t^{\mathrm{M}_i})(R_{eq})^{-1}\Big].$$

$$F_{\boldsymbol{\beta},\mathcal{I}} = -\tfrac{1}{2}\sum_{i\neq\diamond}\frac{\alpha_i}{1-\alpha_\diamond}\fint_{\mathcal{I}}\mathrm{d}t\,\mathrm{tr}\Big[\tilde{R}_t(R_{eq})^{-1}(\tilde{R}_t^{\mathrm{M}_\diamond} - \tilde{R}_t^{\mathrm{M}_i})(R_{eq})^{-1}\Big].$$

which is very similar to the condition in FACT 6 except that there is no short time constraint due to the fact that the moment perturbations remain small in time under the above assumptions. Finally, the simplified result (30) in Fact 7 of §3.2 is obtained by taking into account that the response is due to the forcing perturbations in linear Gaussian systems (24) so that $\tilde{R}_t^{\mathrm{M}_i} = 0$ so that $X^R = Y^{R,R} = 0$ in (74), (75) and only $D_{\boldsymbol{\beta},\mathcal{I}}$, which is independent of the truth response in the covariance, remains in (76). □

B Further details of associated with the sufficient conditions for imperfect prediction improvement via MME

In §3.1.1 we discussed the condition (11) for improving imperfect predictions via MME in the least-biased density representation (14). Here, we discuss the same condition in terms of general perturbations of probability densities which provides additional insight into the essential features of MME with improved prediction skill. In particular, we show that it is difficult to improve the short-term predictive skill via MME containing models with incorrect statistical initial conditions.

The formulation presented below relies on relatively weak assumptions that the truth and model densities can be written as

$$\pi_t^{\mathrm{L}} = \pi_0^{\mathrm{L}} + \delta\tilde{\pi}_t^{\mathrm{L}}, \qquad \pi_t^{\mathrm{M}} = \pi_0^{\mathrm{M}} + \delta\tilde{\pi}_t^{\mathrm{M}}, \qquad \tilde{\pi}_0^{\mathrm{L}} = \tilde{\pi}_0^{\mathrm{M}} = 0, \qquad \int\tilde{\pi}_t^{\mathrm{L}}\mathrm{d}\boldsymbol{u} = \int\tilde{\pi}_t^{\mathrm{L}}\mathrm{d}\boldsymbol{u} = 0, \tag{77}$$

The above decomposition is always possible for the non-singular initial value problem; in the case of the forced response prediction from equilibrium (i.e., when $\pi_0^{\mathrm{L}} = \pi_{eq}^{\mathrm{L}}$, $\pi_0^{\mathrm{M}} = \pi_{eq}^{\mathrm{M}}$) such a decomposition exists for $\delta \ll 1$ under the minimal assumptions of hypoelliptic noise [26]. The possibility of estimating the evolution of statistical moments of the truth density π_t in the case of predicting the forced response within the framework of linear response theory combined with the fluctuation-dissipation approach makes this framework particularly important in this case (see [47, 2, 41, 22, 23, 49, 40, 50, 39])

FACT. Assume the decomposition (77) of the truth and model densities exists as discussed above. Then, the condition (11) for prediction improvement through MME has the following form

$$\mathscr{A}_{\boldsymbol{\beta}}\Big(\pi_0^{\mathrm{L}},\{\pi_0^{\mathrm{M}_i}\}\Big)+\delta\mathscr{B}_{\boldsymbol{\beta},\mathcal{I}}\Big(\pi^{\mathrm{L}},\{\pi^{\mathrm{M}_i}\}\Big)+\delta^2\mathscr{C}_{\boldsymbol{\beta},\mathcal{I}}\Big(\pi^{\mathrm{L}},\{\pi^{\mathrm{M}_i}\}\Big)+\tilde{\Delta}>0, \tag{78}$$

where

$$\mathscr{A}_{\boldsymbol{\beta}}\Big(\pi_0^{\mathrm{L}},\{\pi_0^{\mathrm{M}_i}\}\Big)=\sum_{i\neq\diamond}\beta_i\,\Big(\mathcal{P}(\pi_0^{\mathrm{L}},\pi_0^{\mathrm{M}_\diamond})-\mathcal{P}(\pi_0^{\mathrm{L}},\pi_0^{\mathrm{M}_i})\Big),$$

$$\mathscr{B}_{\boldsymbol{\beta},\mathcal{I}}\Big(\pi^{\mathrm{L}},\{\pi^{\mathrm{M}_i}\}\Big)=\sum_{i\neq\diamond}\beta_i\,⨍_{\mathcal{I}}\mathrm{d}t\int\!\mathrm{d}\boldsymbol{u}\,\left(\pi_0^{\mathrm{L}}\left[\frac{\tilde{\pi}_t^{\mathrm{M}_i}}{\pi_0^{\mathrm{M}_i}}-\frac{\tilde{\pi}_t^{\mathrm{M}_\diamond}}{\pi_0^{\mathrm{M}_\diamond}}\right]+\tilde{\pi}_t^{\mathrm{L}}\log\frac{\pi_0^{\mathrm{M}_i}}{\pi_0^{\mathrm{M}_\diamond}}\right),$$

$$\mathscr{C}_{\boldsymbol{\beta},\mathcal{I}}\Big(\pi_t^{\mathrm{L}},\{\pi_t^{\mathrm{M}_i}\}\Big)=\tfrac{1}{2}\sum_{i\neq\diamond}\beta_i\,⨍_{\mathcal{I}}\mathrm{d}t\int\!\mathrm{d}\boldsymbol{u}\,\left(\pi_0^{\mathrm{L}}\left[\left(\frac{\tilde{\pi}_t^{\mathrm{M}_\diamond}}{\pi_0^{\mathrm{M}_\diamond}}\right)^2-\left(\frac{\tilde{\pi}_t^{\mathrm{M}_i}}{\pi_0^{\mathrm{M}_i}}\right)^2\right]-2\tilde{\pi}_t^{\mathrm{L}}\left[\frac{\tilde{\pi}_t^{\mathrm{M}_\diamond}}{\pi_0^{\mathrm{M}_\diamond}}-\frac{\tilde{\pi}_t^{\mathrm{M}_i}}{\pi_0^{\mathrm{M}_i}}\right]\right),$$

with the weights β_i defined in (11). The following particular cases of the condition (78) for improving the predictions via the MME approach are worth noting in this general representation:

- **Initial (statistical) conditions in all models of MME are consistent with the least-biased estimate of the truth**; i.e., $\pi_0^{\mathrm{M}_i}=\pi_0^{\mathrm{L}}$. In such a case we have $\mathscr{A}_{\boldsymbol{\beta}}=0$, $\mathscr{B}_{\boldsymbol{\beta},\mathcal{I}}=0$ and the condition (78) for improvement of prediction via MME simplifies to

$$⨍_{\mathcal{I}}\mathrm{d}t\int\!\mathrm{d}\boldsymbol{u}\,\frac{(\tilde{\pi}_t^{\mathrm{L}}-\tilde{\pi}_t^{\mathrm{M}_\diamond})^2}{\pi_0^{\mathrm{L}}}+\tilde{\Delta}>\sum_{i\neq\diamond}\beta_i\,⨍_{\mathcal{I}}\mathrm{d}t\int\!\mathrm{d}\boldsymbol{u}\,\frac{(\tilde{\pi}_t^{\mathrm{L}}-\tilde{\pi}_t^{\mathrm{M}_i})^2}{\pi_0^{\mathrm{L}}}. \tag{79}$$

  In the case of forced response predictions, perturbation of the truth density $\tilde{\pi}_t^{\mathrm{L}}$ can be estimated from the statistics on the unperturbed equilibrium through the linear response theory and fluctuation-dissipation formulas exploiting only the unperturbed equilibrium information [47, 2, 41, 22, 23, 49, 40, 50, 39]).

- **Initial model densities in MME perturbed relative to the least-biased estimate of the truth**; i.e., $\pi_0^{\mathrm{M}_i}=\pi_0^{\mathrm{L}}+\epsilon\,\tilde{\pi}_0^{\mathrm{M}_i}$, $\pi_0^{\mathrm{M}_\diamond}=\pi_0^{\mathrm{L}}$. In such a case all terms in (78) are non-trivial but they can be written as

$$\mathscr{A}_{\boldsymbol{\beta}}\Big(\pi_0^{\mathrm{L}},\{\pi_0^{\mathrm{M}_i}\}\Big)=-\epsilon^2\sum_{i\neq\diamond}\beta_i\int\!\mathrm{d}\boldsymbol{u}\frac{(\tilde{\pi}_0^{\mathrm{M}_i})^2}{2\pi_0^{\mathrm{L}}}+\mathcal{O}(\epsilon^3), \tag{80}$$

$$\mathscr{B}_{\boldsymbol{\beta},\mathcal{I}}\Big(\pi_t^{\mathrm{L}},\{\pi_t^{\mathrm{M}_i}\}\Big)=\epsilon\sum_{i\neq\diamond}\beta_i\,⨍_{\mathcal{I}}\mathrm{d}t\int\!\mathrm{d}\boldsymbol{u}\Big(\tilde{\pi}_t^{\mathrm{M}_i}B_1+\tilde{\pi}_t^{\mathrm{M}_\diamond}B_2+\tilde{\pi}_t^{\mathrm{L}}B_3\Big), \tag{81}$$

$$\mathscr{C}_{\boldsymbol{\beta},\mathcal{I}}\Big(\pi_t^{\mathrm{L}},\{\pi_t^{\mathrm{M}_i}\}\Big)=\sum_{i\neq\diamond}\beta_i⨍_{\mathcal{I}}\mathrm{d}t\int\mathrm{d}\boldsymbol{u}\left[\frac{(\tilde{\pi}_t^{\mathrm{L}}-\tilde{\pi}_t^{\mathrm{M}_\diamond})^2-(\tilde{\pi}_t^{\mathrm{L}}-\tilde{\pi}_t^{\mathrm{M}_i})^2}{\pi_0^{\mathrm{L}}}+\epsilon\Big(\tilde{\pi}_t^{\mathrm{M}_i}C_1+\tilde{\pi}_t^{\mathrm{M}_\diamond}C_2+\tilde{\pi}_t^{\mathrm{L}}C_3\Big)\right], \tag{82}$$

  where $\{B_m\},\{C_m\}$, $m=1,2,3$ are functions of $\tilde{\pi}_0^{\mathrm{M}_i},\tilde{\pi}_0^{\mathrm{M}_\diamond},\tilde{\pi}_0^{\mathrm{L}}$ and ϵ. Note that unless $\epsilon=0$ (so that $\pi_0^{\mathrm{M}_i}=\pi_0^{\mathrm{L}}$), it is difficult to improve the prediction skill at short times within the MME framework since at $t=0$, we have $\mathscr{B}_{\boldsymbol{\beta},\mathcal{I}}=\mathscr{C}_{\boldsymbol{\beta},\mathcal{I}}=0$ and $\mathscr{A}_{\boldsymbol{\beta}}<0$ in (78).

References

[1] R.V. Abramov and A.J. Majda. Quantifying uncertainty for non-Gaussian ensembles in complex systems. *SIAM J Sci Comput*, 26:411–447, 2004.

[2] R.V. Abramov and A.J. Majda. Blended Response Algorithms for Linear Fluctuation-Dissipation for Complex Nonlinear Dynamical Systems. *Nonlinearity*, 20(12):2793–2821, 2007.

[3] J.L. Anderson. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, 59A:210–224, 2007.

[4] L. Arnold. *Random Dynamical Systems.* Springer-Verlag, New York, 1998.

[5] M. Branicki, B. Gershgorin, and A.J. Majda. Filtering skill for turbulent signals for a suite of nonlinear and linear Kalman filters. *J. Comp. Phys*, 231:1462–1498, 2012.

[6] M. Branicki and A.J. Majda. Dynamic Stochastic Superresolution of sparsely observed dynamical systems. *J. Comp. Phys.*, 241:333–363, 2012.

[7] M. Branicki and A.J. Majda. Fundamental Limitations of Polynomial Chaos for Uncertainty Quantification in Systems with Intermittent Instabilities. *Comm. Math. Sci.*, 11(1), 2012.

[8] M. Branicki and A.J. Majda. Quantifying uncertainty for statistical predicions with model errors in non-Guassian models with intermittency. *Nonlinearity*, 25:2543–2578, 2012.

[9] M. Branicki and A.J. Majda. Qyantifying bayesian filter performance for turbulent dynamical systems through informatoin theory. *Comm. Math. Sci.*, 2014. in press.

[10] A. Chatterjee and D. Vlachos. An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. *J. Comput-Aided Mater*, 14:253–308, 2007.

[11] T.A. Cover and J.A. Thomas. *Elements of Information Theory.* Wiley-Interscience, Hoboken, 2006.

[12] P. Das, M. Moll, H. Stamati, L.E. Kavraki, and C. Clementi. Low-dimensional, free energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proc. Natl. Acad. Sci. USA*, 103:9885–9890, 2006.

[13] F.J. Doblas-Reyes, R. Hagedorn, and T.N. Palmer. The rationale behind the success of multi-model ensembles in seasonal forecasting. Part II: Calibration and combination. *Tellus, Ser. A*, 57:234–252, 2005.

[14] K.A. Emanuel, J.C. Wyngaard, J.C. McWilliams, D.A. Randall, and Y.L. Yung. *Improving the scientific foundation for atmosphere-land ocean simulations.* Natl. Acad. Press, Washington DC, 2005.

[15] E.S. Epstein. Stochastic dynamic predictions. *Tellus*, 21:739–759, 1969.

[16] B. Gershgorin, J. Harlim, and A.J. Majda. Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation. *J. Comp. Phys.*, 229:32–57, 2010.

[17] B. Gershgorin, J. Harlim, and A.J. Majda. Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comp. Phys.*, 229:1–31, 2010.

[18] B. Gershgorin and A.J. Majda. Quantifying uncertainty for climate change and long range forecasting scenarios with model errors. Part I: Gaussian models. *J. Climate*, 25:4523–4548, 2012.

[19] D. Giannakis and A.J. Majda. Quantifying the predictive skill in long-range forecasting. Part I: Coarse-grained predictions in a simple ocean model. *J. Climate*, 25:1793–1813, 2012.

[20] D. Giannakis and A.J. Majda. Quantifying the predictive skill in long-range forecasting. Part II: Model error in coarse-grained Markov models with application to ocean-circulation regimes. *J. Climate*, 25:1814–1826, 2012.

[21] D. Giannakis, A.J. Majda, and I. Horenko. Information theory, model error, and predictive skill of stochastic models for complex nonlinear systems. *Physica D*, 241(20):1735–1752, 2012.

[22] A. Gritsun and G. Branstator. Climate response using a three-dimensional operator based on the fluctuation-dissipation theorem. *J. Atmos. Sci.*, 64:2558–2575, 2007.

[23] A. Gritsun, G. Branstator, and A.J. Majda. Climate response of linear and quadratic functionals using the fluctuation-dissipation theorem. *J. Atmos. Sci.*, 65:2824–2841, 2008.

[24] I. Grooms and A.J. Majda. Efficient Stochastic Superparameterization for Geophysical Turbulence. *Proc. Nat. Acad. Sci.*, 110(12):4464–4469, 2013.

[25] R. Hagedorn, F.J. Doblas-Reyes, and T.N. Palmer. The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus*, 57A:219–233, 2005.

[26] M. Hairer and A.J. Majda. A simple framework to justify linear response theory. *Nonlinearity*, 12:909–922, 2010.

[27] J. Harlim and A.J. Majda. Filtering turbulent sparsely observed geophysical flows. *Mon. Wea. Rev.*, 138(4):1050–1083, 2010.

[28] P. Houtekamer and H. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, 129:123–137, 2001.

[29] G. Hummer and I.G. Kevrekidis. Coarse molecular dynamics of a peptide fragment: Free energy, kinetics and long-time dynamics computations. *J. Chem. Phys.*, 118:10762–10773, 2003.

[30] M.A. Katsoulakis, A.J. Majda, and D. Vlachos. Coarse-grained stochastic processes for microscopic lattice systems. *Proc. Natl. Acad. Sci. USA*, 100:782–787, 2003.

[31] H-M Kim, P. J. Webster, and J. A. Curry. Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. *Geophys. Res. Lett.*, 39:L10701, 2012.

[32] R. Kleeman. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, 59(13):2057–2072, 2002.

[33] R. Kleeman, A.J. Majda, and I.I. Timofeyev. Quantifying predictability in a model with statistical features of the atmosphere. *Proc. Natl. Acad. Sci. USA*, 99:15291–15296, 2002.

[34] S. Kullback and R. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951.

[35] C.E. Leith. Climate response and fluctuation dissipation. *J. Atmospheric Sci.*, 32:2022–2025, 1975.

[36] E.N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141, 1963.

[37] E.N. Lorenz. A study of predictability of a 28-variable atmospheric model. *Tellus*, 17:321–333, 1968.

[38] E.N. Lorenz. The predictability of a flow which possesses many scales of motion. *Tellus*, 21:289–307, 1969.

[39] A. J. Majda, R.V. Abramov, and B. Gershgorin. High skill in low frequency climate response through fluctuation dissipation theorems despite structural instability. *Proc. Natl. Acad. Sci. USA*, 107(2):581–586, 2010.

[40] A. J. Majda and B. Gershgorin. Improving model fidelity and sensitivity for complex systems through empirical information theory. *Proc. Natl. Acad. Sci.*, 108(31):10044–10049, 2011. 10.1073/pnas.1105174108.

[41] A. J. Majda, B. Gershgorin, and Y. Yuan. Low-Frequency Climate Response and Fluctuation–Dissipation Theorems: Theory and Practice. *J. Atmos. Sci.*, 67:1186, 2010.

[42] A. J. Majda, R. Kleeman, and D. Cai. A mathematical framework for predictability through relative entropy. *Methods Appl. Anal.*, 9(3):425–444, 2002.

[43] A. J. Majda, I.I. Timofeyev, and E. Vanden Eijnden. Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.*, 60:1705, 2003.

[44] A. J. Majda and X. Wang. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*. Cambridge University Press, Cambridge, 2006.

[45] A. J. Majda and X. Wang. Linear response theory for statistical ensembles in complex systems with time-periodic forcing. *Comm. Math. Sci.*, 8(1):145–172, 2010.

[46] A.J. Majda. Challenges in Climate Science and Contemporary Applied Mathematics. *Comm. Pure Appl. Math.*, 65(7):920–948, 2012.

[47] A.J. Majda, R.V. Abramov, and M.J. Grote. *Information Theory and Stochastics for Multiscale Nonlinear Systems*, volume 25 of *CRM Monograph Series*. Americal Mathematical Society, Providence, 2005.

[48] A.J. Majda and M. Branicki. Lessons in Uncertainty Quantification for Turbulent Dynamical Systems. *Discrete Contin. Dynam. Systems*, 32(9)(9):3133–3231, 2012.

[49] A.J. Majda and B. Gershgorin. Quantifying Uncertainty in Climage Change Science through Empirical Information Theory. *Proc. Natl. Acad. Sci.*, 107(34):14958–14963, 2010.

[50] A.J. Majda and B. Gershgorin. Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error. *Proc. Natl. Acad. Sci.*, 108(31):12599–12604, 2011. 10.1073/pnas.1108132108.

[51] A.J. Majda and I. Grooms. New Perspectives on Superparameterization for Geophysical Turbulence. submitted to J. Comp. Phys. 2013.

[52] A.J. Majda and J. Harlim. *Filtering Complex Turbulent Systems*. Cambridge University Press, 2012.

[53] A.J. Majda, J. Harlim, and B. Gershgorin. Mathematical strategies for filtering turbulent dynamical systems. *Discrete Contin. Dynam. Systems*, 27:441–486, 2010.

[54] A.J. Majda and Y. Yuan. Fundamental Limitations of Ad Hoc Linear and Quadratic Multi-Level Regression Models for Physical Systems. *Discrete Cont. Dyn. Systems*, 4:1333–1363, 2012.

[55] L.R. Mead and N. Papanicolaou. Maximum entropy in the problem of moments. *J. Math. Phys.*, 25(8):2404–2417, 1984.

[56] J.D. Neelin, M. Munnich, H. Su, J.E. Meyerson, and C.E. Holloway. Tropical drying trends in global warming models and observations. *Proc Natl Acad Sci USA*, 103:6110–6115, 2006.

[57] F. Noé, Ch. Schutte, E Vanden-Eijnden, L Reich, and T R Weikl. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA*, 106(45):19011–19016, 2009.

[58] B.K. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 2010.

[59] T.N. Palmer. A nonlinear dynamical perspective on model error: a proposal for nonlocal stochastic dynamic parameterizations in weather and climate prediction models. *Quart J Roy Meteor Soc*, 127:279–303, 2001.

[60] T.N. Palmer, G.J. Shutts, R. Hagedorn, F.J. Doblas-Reyes, T. Jung, and M. Leutbecher. Representing Model Uncertainty in Weather and Climate Prediction. *Ann. Rev. Earth and Planetary Sci.*, 33:163–193, 2005.

[61] D.A. Randall. *Climate models and their evaluation. Climate change 2007: The physical science basis, contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*, pages 589–662. Cambridge University Press, 2007.

[62] A. Rangan, L. Tao, G. Kovacic, and D. Cai. Multiscale modeling of the primary visual cortex. *IEEE Eng Med Biol Mag*, 28(3):19–24, 2009.

[63] T. Sapsis and A.J. Majda. Blending modified Gaussian closure and non-Gaussian reduced subspace methods for turbulent dynamical systems. to appear in J. Nonlin. Sci. 2013.

[64] T. Sapsis and A.J. Majda. A statistically accurate modified quasilinar Gaussian closure for uncertainty quantification in turbulent dynamical systems. *Physica D*, 252:34–45, 2013.

[65] T. Sapsis and A.J. Majda. Blended reduced subspace algorithms for uncertainty quantification of quadratic systems with a stable mean state. *Physica D*, 2013. in press.

[66] S. Sriraman, I.G. Kevrekidis, and G. Hummer. Coarse master equation from Bayesian analysis of replica molecular dynamics. *J. Phys. Chem. B*, 109:6479–6484, 2005.

[67] D.B. Stephenson, C.A.S Coelho, F.J. Doblas-Reyes, and M. Balmaseda. Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions. *Tellus A*, 57:253–264, 2005.

[68] P. van der Linden and J.F.B Mitchell (eds). Ensembles: climate change and its impacts: summary of research and results from the ensembles project. *Met Office Hadley Centre, Fitzroy Road, Exeter EX1 3PB, UK*, 2009.

[69] G.J. van Oldenborgh, F.J. Doblas-Reyes, B. Wouters, and W. Hazeleger. Decadal prediction skill in a multi-model ensemble. *Clim. Dyn.*, 38:1263–1280, 2012.

[70] A.P. Weigel, M.A. Liniger, and C. Appenzeller. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Q.J.R. Meteorol. Soc.*, 134:241–260, 2008.

[71] A. Weisheimer, F.J. Doblas-Reyes, T.N. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel. ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys. Res. Lett.*, 36:L21711, 2009.