Interplay between information theory, uncertainty quantification, and improving reduced-order predictions

Michał Branicki

School of Mathematics & Maxwell Institute, U. Edinburgh, UK

A. J. Majda

Dept. of Mathematics, Courant Institute, NYU, USA

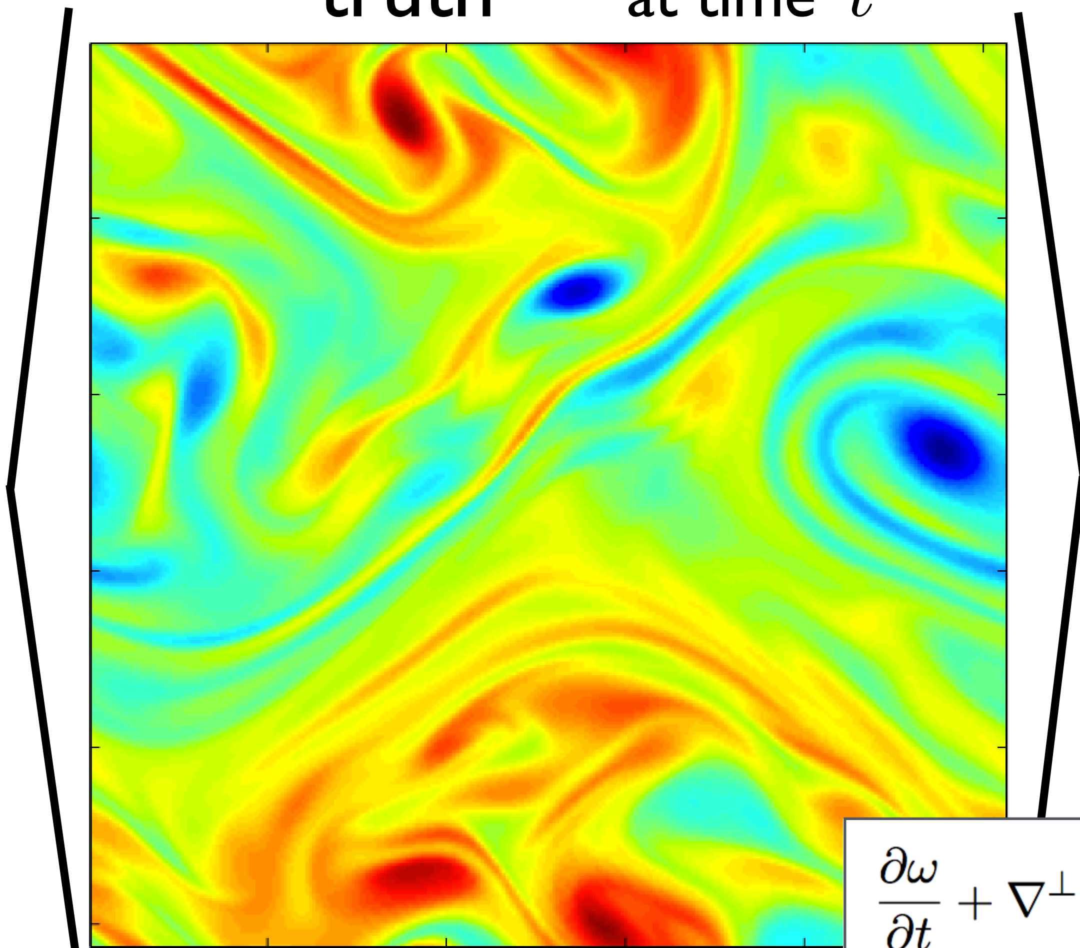
Outline:

> "All models are wrong but some are useful" *G.E.P. Box*

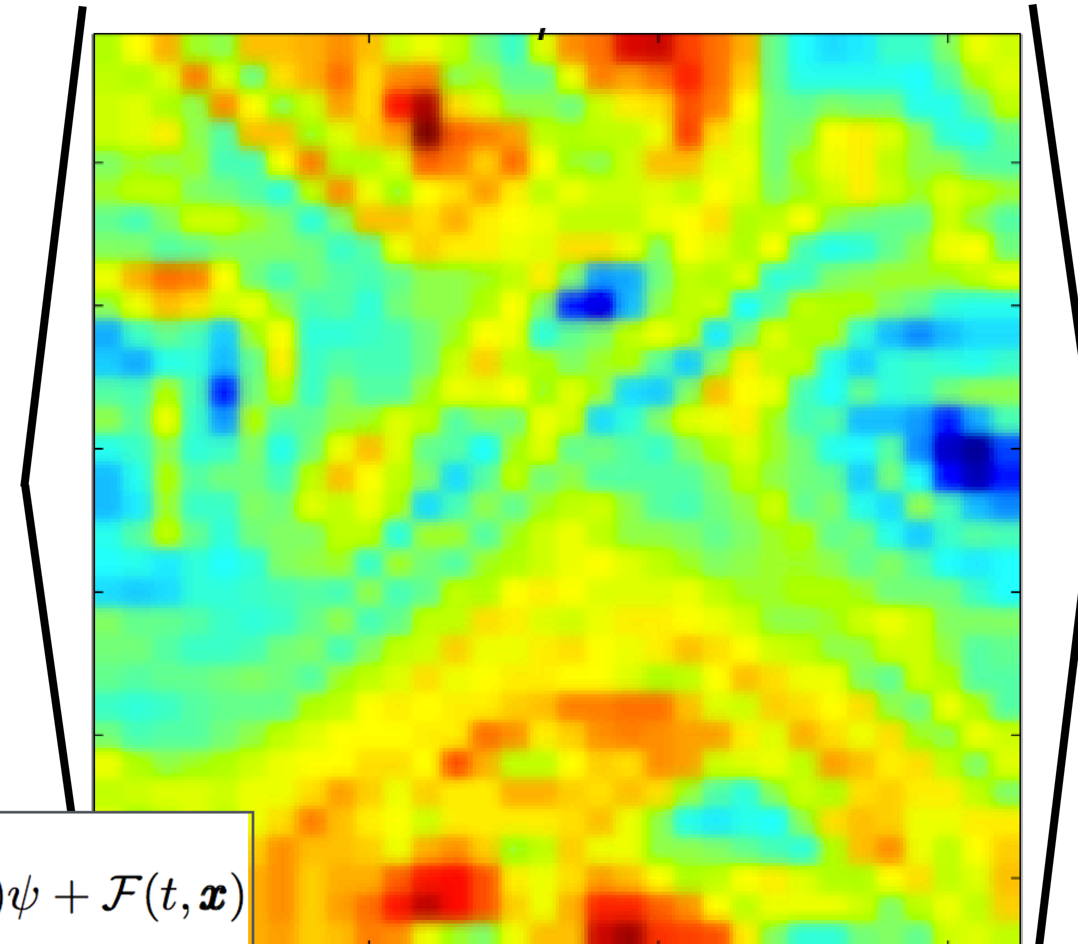
- Predictions from 'coarse' models tuned from data
- Why information theory ?
- Multi-model ensemble predictions & info theory
- Simple example
- Summary & outlook

Accurate probabilistic predictions from **coarse-grained models**

"truth" at time t



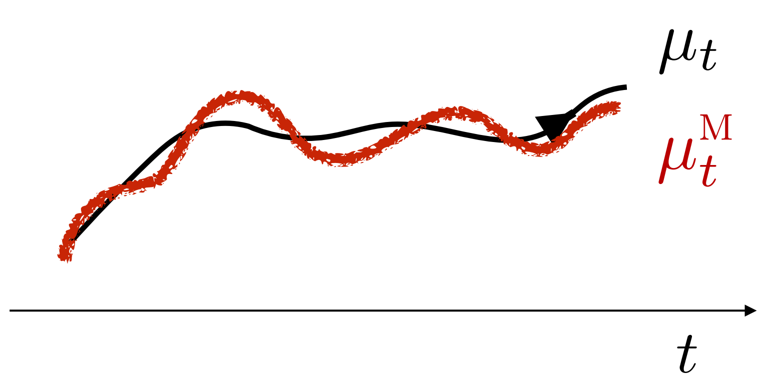
"coarse" finite-dim model



$$\frac{\partial \omega}{\partial t} + \nabla^{\perp}\psi \cdot \nabla \omega = \mathcal{D}(\Delta)\psi + \mathcal{F}(t, \boldsymbol{x})$$

$$\frac{\mathrm{d}\hat{\omega}_{\boldsymbol{k}}(t)}{\mathrm{d}t} \propto \sum_{|\boldsymbol{m}| \neq |\boldsymbol{k}-\boldsymbol{m}|} \frac{\boldsymbol{m}^{\perp} \cdot \boldsymbol{k}}{|\boldsymbol{m}|^2} \hat{\omega}_{\boldsymbol{m}}(t) \hat{\omega}_{\boldsymbol{k}-\boldsymbol{m}}(t)$$

$$\frac{\mathrm{d}\hat{\omega}^{\mathrm{M}}_{\boldsymbol{k}}(t)}{\mathrm{d}t} \propto \sum_{\substack{|\boldsymbol{m}| \neq |\boldsymbol{k}-\boldsymbol{m}| \\ |\boldsymbol{m}| \in \mathcal{K}}} \frac{\boldsymbol{m}^{\perp} \cdot \boldsymbol{k}}{|\boldsymbol{m}|^2} \hat{\omega}^{\mathrm{M}}_{\boldsymbol{m}}(t) \hat{\omega}^{\mathrm{M}}_{\boldsymbol{k}-\boldsymbol{m}}(t)$$

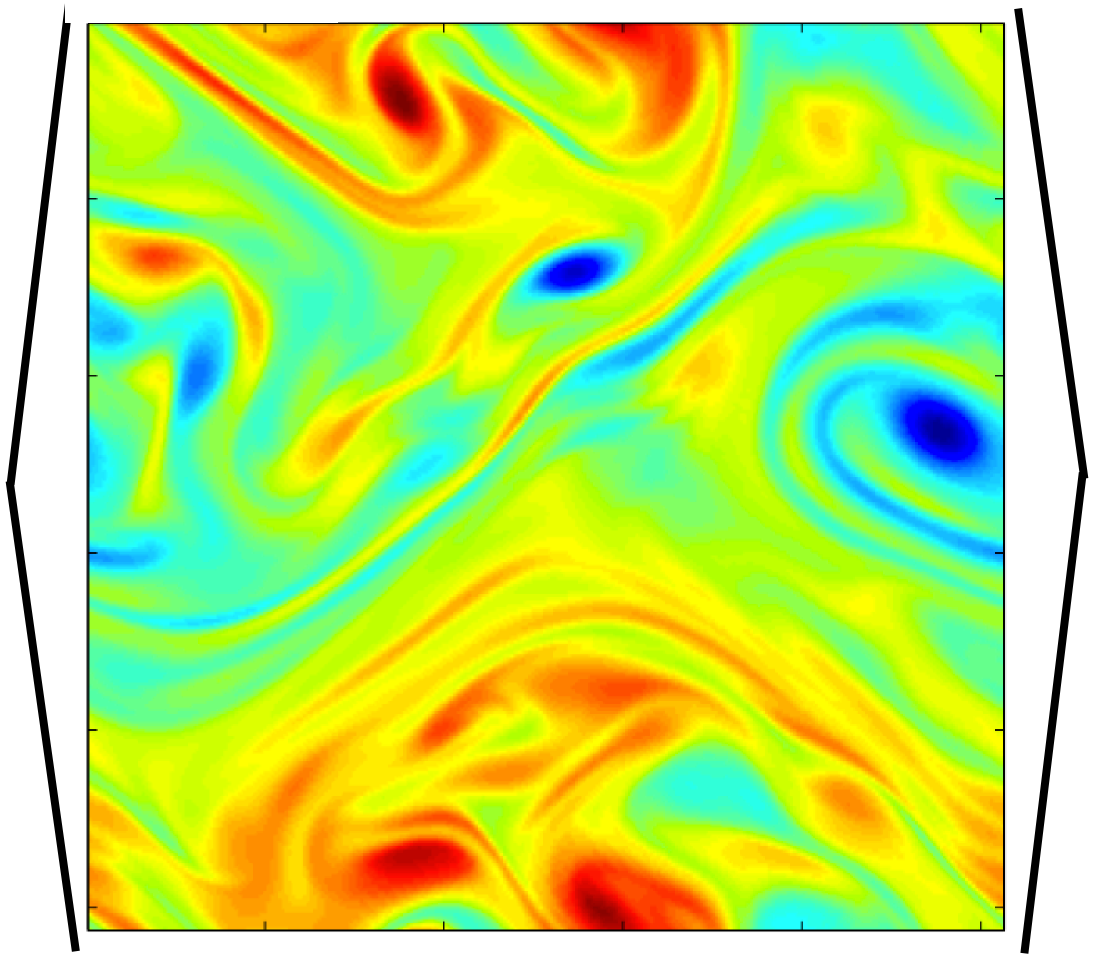


$$(\Omega_{\mathcal{K}}, \mathscr{A})$$

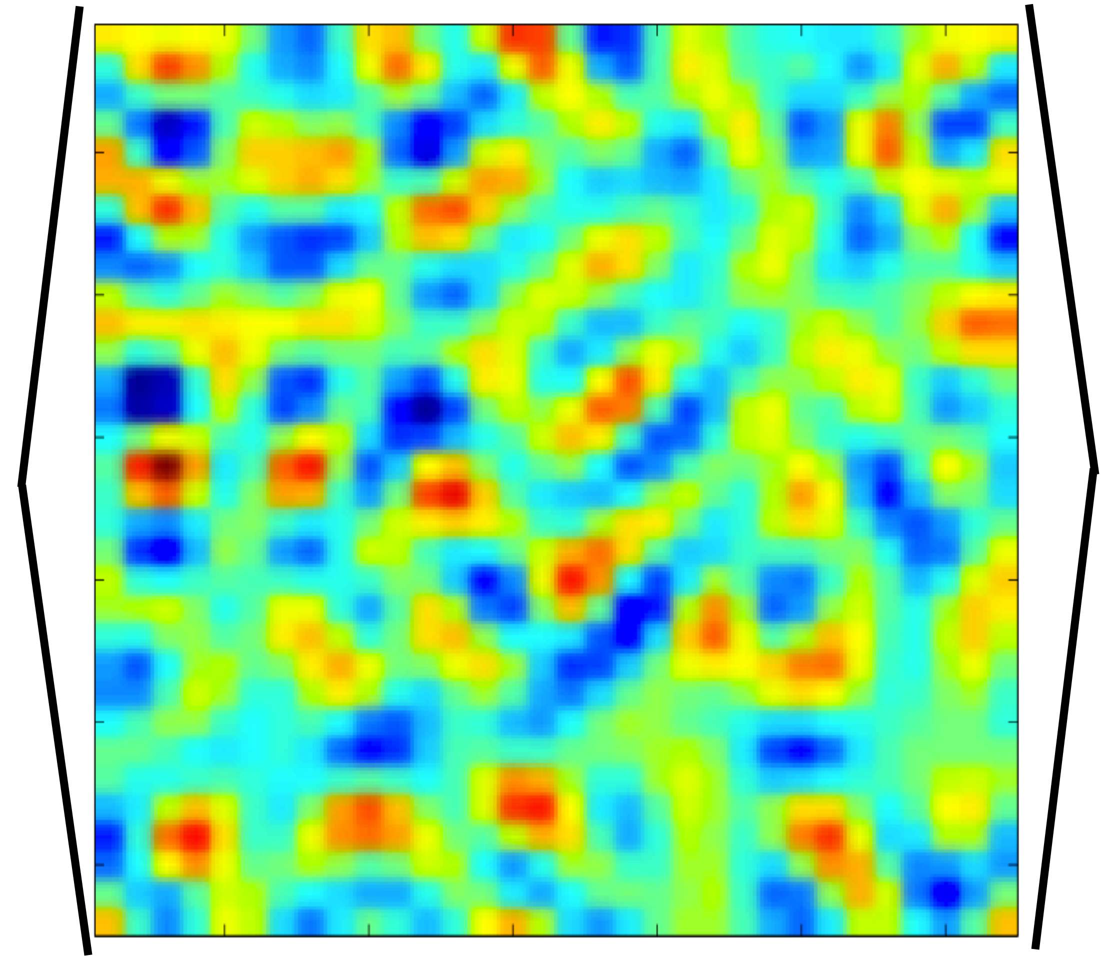
$$\Omega_{\mathcal{K}} := span\big(\{\hat{\omega}_k\}_{k \in \mathcal{K}}\big)$$

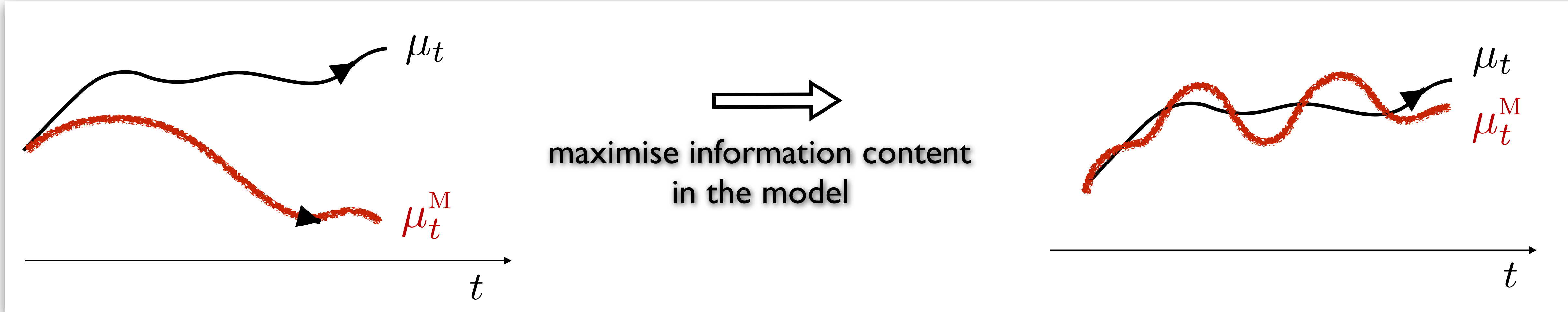
In reality ...

“truth”



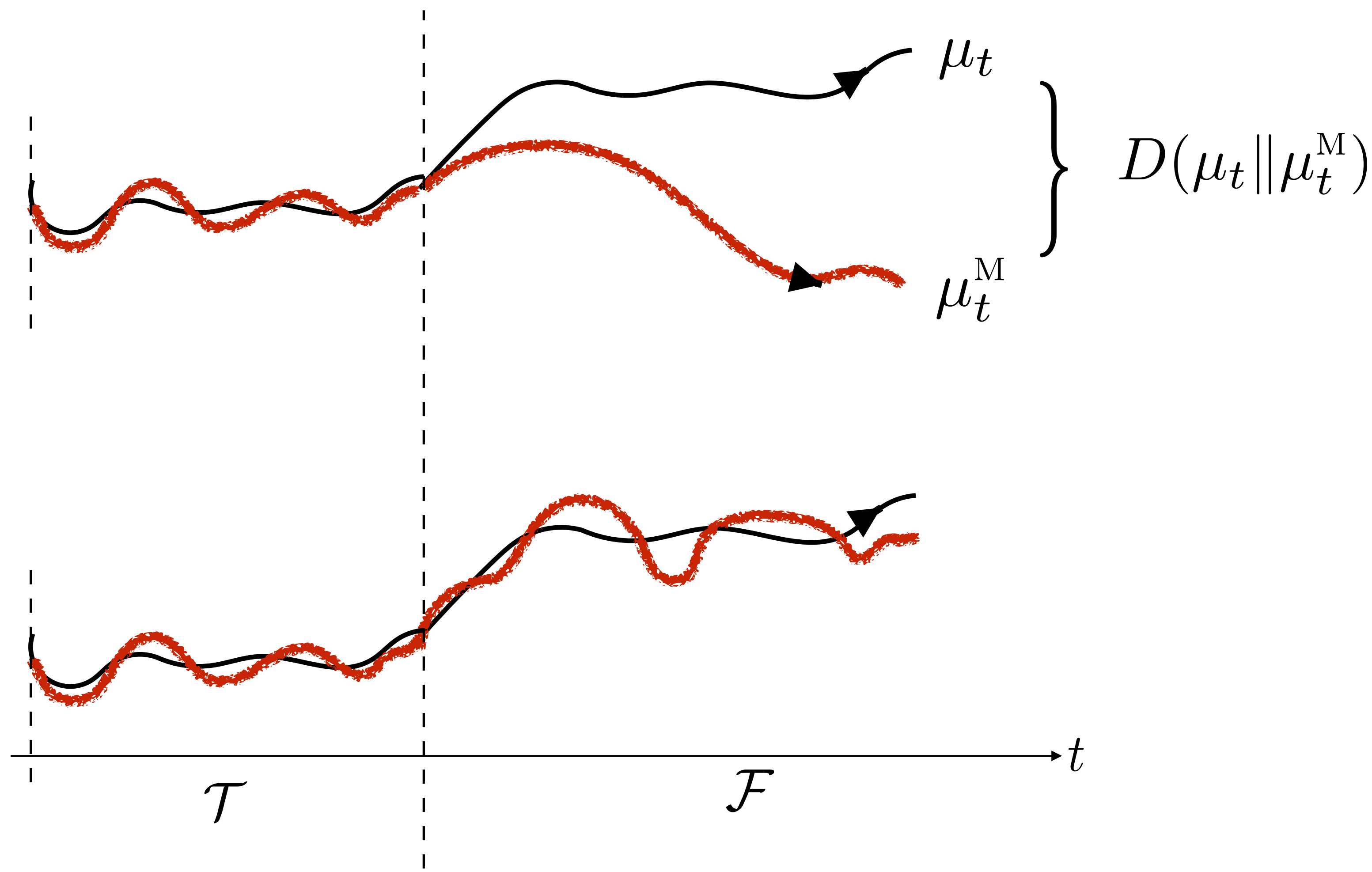
“coarse” finite-dim model



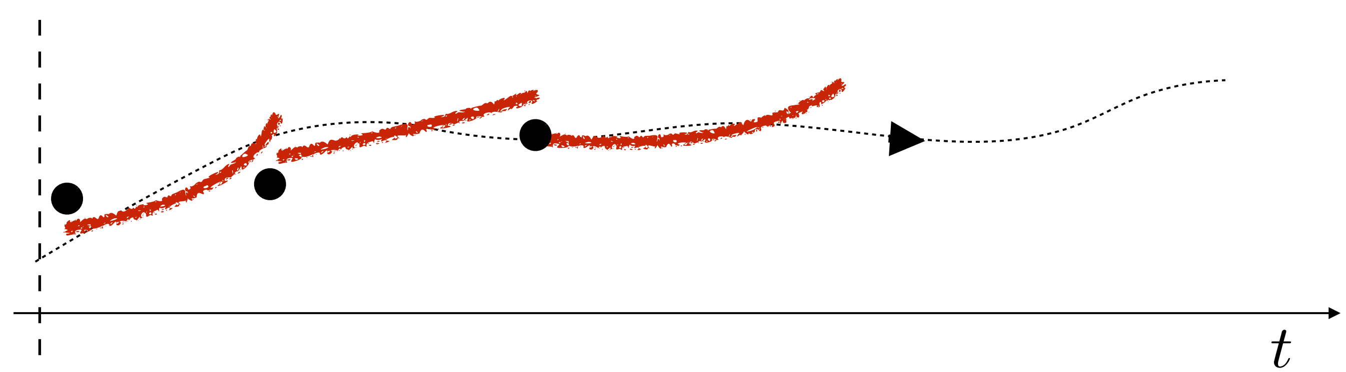


Optimisation of models & their predictions

- **Tuning:** Minimise the lack of information in the imperfect predictions by improving the models in the "*training phase*" (when **lots of data is available**)



- Use **real-time data** in the "*prediction phase*" (**time-sequential data assimilation**)



Optimisation of models & their predictions

Goals:

- "Best-fit" coarse-grained model to "fine-scale" dynamics/data
- UQ for multi-scale problems
- Sensitivity analysis, robustness, parameter identifiability

Key notions:

- Metric $d(\mu_t, \nu_t)$ or a pre-metric $D(\mu_t \| \nu_t)$ on the space of probability measures
- Accuracy/error for observables $\left|\mathbb{E}^{\mu_t}[f] - \mathbb{E}^{\nu_t}[f]\right|$
- Sensitivity under perturbations $\mu_t^\theta \mapsto \mu_t^{\theta+\delta\theta}$ or $\mathbb{E}^{\mu_t^\theta}[f] \mapsto \mathbb{E}^{\mu_t^{\theta+\delta\theta}}[f]$

Optimisation of models & their predictions

Key tools:

- ϕ**-entropies** (or their rates in path-space). In particular, the relative entropy (Kullback-Leibler divergence)

$$D_{KL}(\mu\|\nu) := \int \ln\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\mathrm{d}\mu \qquad \mu \ll \lambda \qquad \nu \ll \lambda$$

- ‘Information’ inequalities for specific observables f, e.g.,

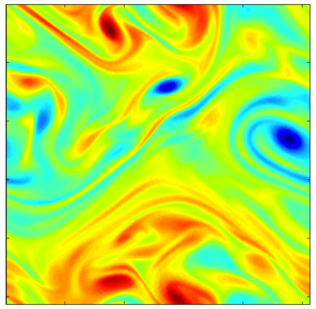
$$\left|\mathbb{E}^{\mu_t}[f] - \mathbb{E}^{\nu_t}[f]\right| \leqslant |f|_\infty \sqrt{2D_{KL}(\mu_t\|\nu_t)} \qquad \text{(Pinsker)}$$

$$\left|\mathbb{E}^{\mu_t}[f] - \mathbb{E}^{\nu_t}[f]\right| \leqslant 2\left(\mathbb{E}^{\mu_t}[f^2] + \mathbb{E}^{\nu_t}[f^2]\right)^{1/2} \sqrt{D_{KL}(\mu_t\|\nu_t)}$$

- Sensitivity analysis (Fisher information, linear response to perturbations)

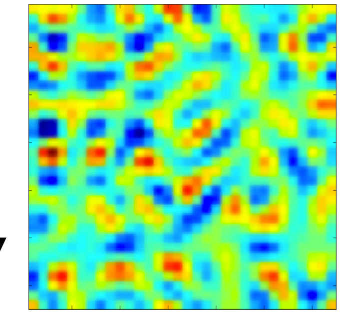
$$D_{KL}(\mu^\theta\|\mu^{\theta+\delta\theta}) = \frac{1}{2}\delta\theta^\dagger F(\mu^\theta)\delta\theta + \mathcal{O}(\delta\theta^3)$$

Measure of the lack of information in models



$\pi(\boldsymbol{u})$ marginal truth density

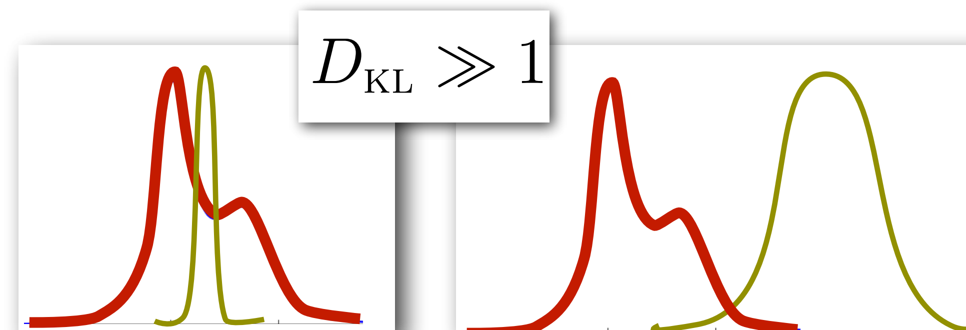
$\pi^{\mathrm{M}}(\boldsymbol{u})$ model density



on $\boldsymbol{u} \in \Omega_{\mathcal{K}}$ (e.g., resolved Fourier modes)

Model error

$$D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}}) = \int \pi \log \frac{\pi}{\pi^{\mathrm{M}}}$$



The **relative entropy** $D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}})$ quantifies the lack of information in π^{M} relative to π

(i) $D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}}) \geqslant 0$, $D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}}) = 0$ iff $\pi = \pi^{\mathrm{M}}$

(ii) $D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}})$ is invariant under nonlinear changes of variables

(iii) $d_H^2(\pi, \pi^{\mathrm{M}}) \leqslant D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}}) \leqslant d_H(\pi, \pi^{\mathrm{M}}) + \frac{1}{2}\chi^2(\pi, \pi^{\mathrm{M}})$

Model error, internal prediction skill and sensitivity

Branicki, Nonlinearity, 2012

- Model error

$$\mathcal{E}(t;t_0) = D_{KL}(\pi_t \| \pi_t^{\mathrm{M}})$$

Lack of information in the imperfect model density compared to the perfect statistical forecast

- Prediction skill

$$\mathcal{S}\mathrm{K}(t;t_0) = D_{KL}(\pi_t \| \pi_{att})$$

- Internal prediction skill

$$\mathcal{S}\mathrm{K}^{\mathrm{M}}(t;t_0) = D_{KL}(\pi_t^{\mathrm{M}} \| \pi_{att}^{\mathrm{M}})$$

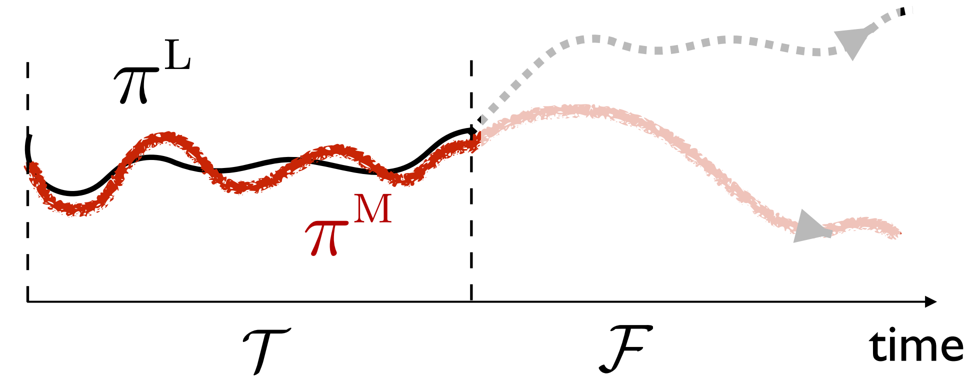
Information beyond the climate in the perfect/imperfect model forecast.

- Model sensitivity

$$\mathcal{S}\mathrm{E}(t;t_0) = D_{KL}(\pi_t^{\delta} \| \pi_{att}) \qquad \mathcal{S}\mathrm{E}^{\mathrm{M}}(t;t_0) = D_{KL}(\pi_t^{\mathrm{M},\delta} \| \pi_{att}^{\mathrm{M}})$$

Lack of information in the perfect/imperfect unperturbed climate relative to the statistical forecast of response to external or internal perturbations.

Tuning imperfect models



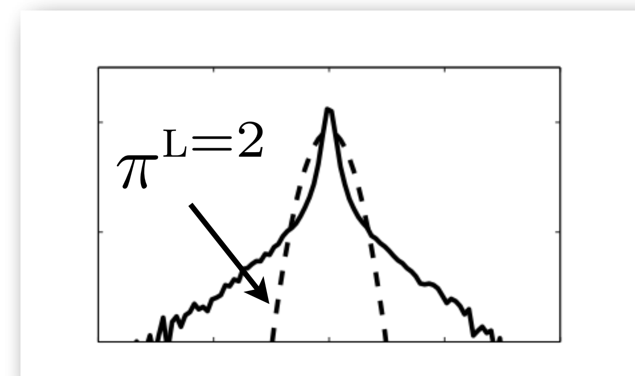
π : true marginal density

$\mathcal{M}$: class of imperfect models with marginal densities π^{M}

- The best model $\mathrm{M}_* \in \mathcal{M}$ minimises the lack of information

$$D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}*}) = \min_{\mathrm{M}\in\mathcal{M}} D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}})$$

$$\pi^{\mathrm{L}} \propto \exp\big(-\boldsymbol{\theta}(\bar{\boldsymbol{E}}_t)\cdot\boldsymbol{E}(\boldsymbol{u})\big)$$



- For π^{L} the max-entropy approximation of π based on L moment constraints

$$D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}}) \leqslant D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{L}}) + D_{\mathrm{KL}}(\pi^{\mathrm{L}} \,\|\, \pi^{\mathrm{M}})$$

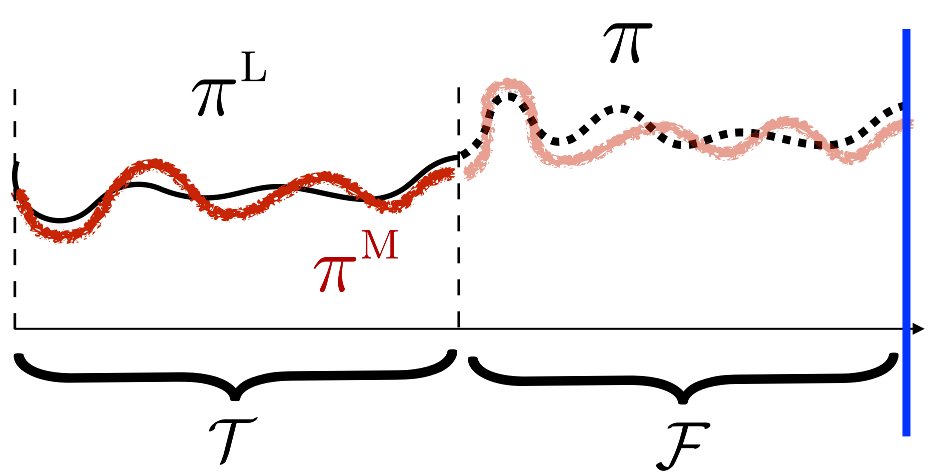
optimised model

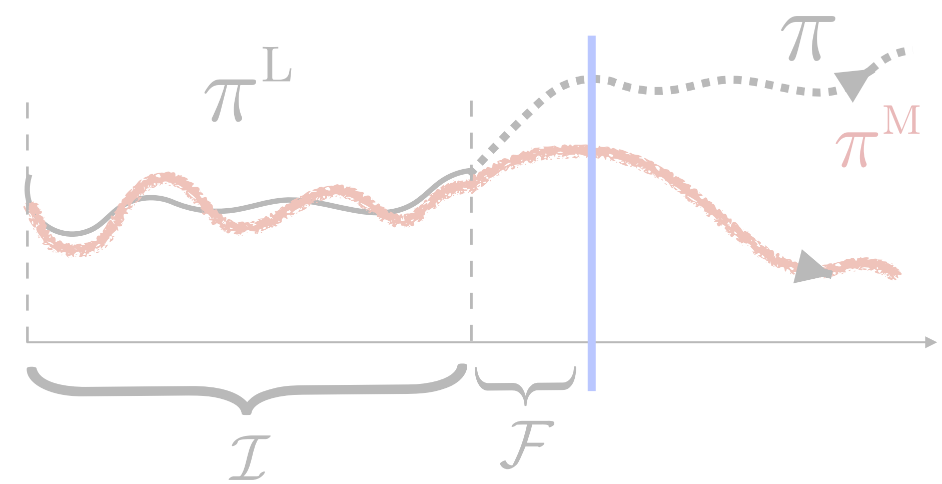
"information" barrier

$$D_{\mathrm{KL}}(\pi^{\mathrm{L}} \,\|\, \pi^{\mathrm{M}*}) = \min_{\mathrm{M}\in\mathcal{M}} D_{\mathrm{KL}}(\pi^{\mathrm{L}} \,\|\, \pi^{\mathrm{M}})$$

Improving imperfect predictions via tuning attractor fidelity

Branicki, Enc. Appl. Math, 2015





$$D^{\mathcal{F}}_{\mathrm{KL}}(\pi^{\delta}\,\|\,\pi^{\mathrm{M},\delta}) \leqslant D^{\mathcal{F}}_{\mathrm{KL}}(\pi^{\delta}\,\|\,\pi^{\mathrm{L},\delta}) + D^{\mathcal{F}}_{\mathrm{KL}}(\pi^{\mathrm{L},\delta}\,\|\,\pi^{\mathrm{M},\delta})$$

FACT: Improving attractor fidelity of model improves predictions

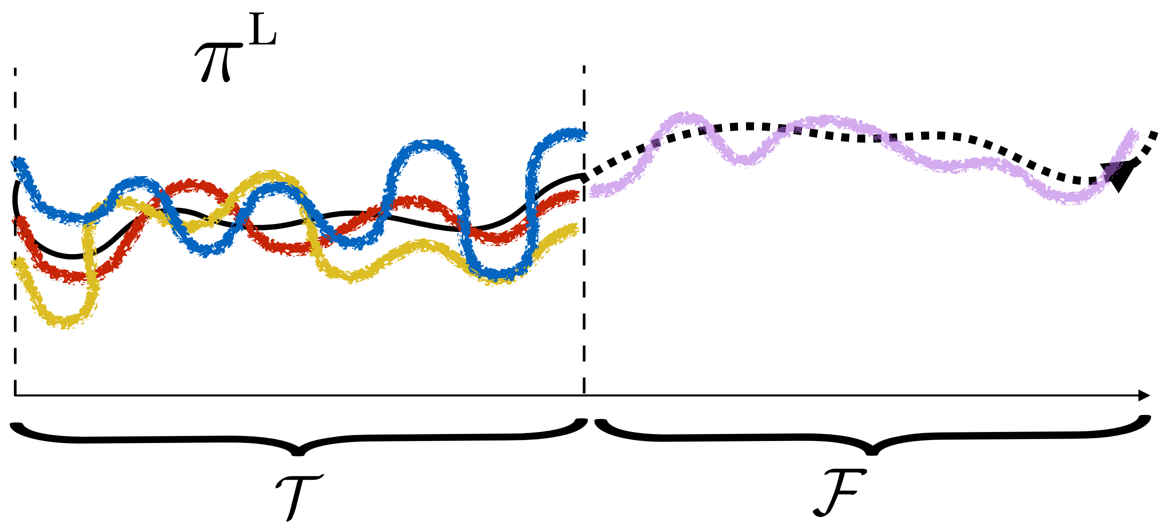
$$D^{\mathcal{F}}_{\mathrm{KL}}(\pi^{\mathrm{L},\delta}\,\|\,\pi^{\mathrm{M},\delta}) \leqslant \left\|\boldsymbol{\theta}^{\mathrm{M}} - \boldsymbol{\theta}\right\|^{1/2}_{L^2(\mathcal{T})}\left\|\bar{\boldsymbol{E}}^{\delta}\right\|^{1/2}_{L^2(\mathcal{F})} + \mathcal{O}\left((\delta\bar{\boldsymbol{E}})^2\right)$$

$$\bar{\boldsymbol{E}}^{\delta} = \bar{\boldsymbol{E}} + \delta\bar{\boldsymbol{E}} \qquad \bar{\boldsymbol{E}}^{\delta} \equiv \int_{\Omega}\boldsymbol{E}(\boldsymbol{u})\pi^{\delta}(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$$

$$\pi^{\delta} \propto \exp\left(-\,\boldsymbol{\theta}^{\delta}(\bar{\boldsymbol{E}}^{\delta})\cdot\boldsymbol{E}(\boldsymbol{u})\right)$$

Improving probabilistic predictions by tuning attractor fidelity

Branicki, Enc. Appl. Math, 2015



$$\pi^{\mathrm{MME}}_{\alpha,t} = \sum_i \alpha_i\, \pi^{\mathrm{M}_i}_t$$

$$D^{\mathcal{F}}_{\mathrm{KL}}(\pi^{\delta} \,\|\, \pi^{\mathrm{MME},\delta}_{\alpha}) \leqslant D^{\mathcal{F}}_{\mathrm{KL}}(\pi^{\delta} \,\|\, \pi^{\mathrm{L},\delta}) + D^{\mathcal{F}}_{\mathrm{KL}}(\pi^{\mathrm{L},\delta} \,\|\, \pi^{\mathrm{MME},\delta}_{\alpha})$$

FACT: Improving attractor fidelity of MME improves predictions

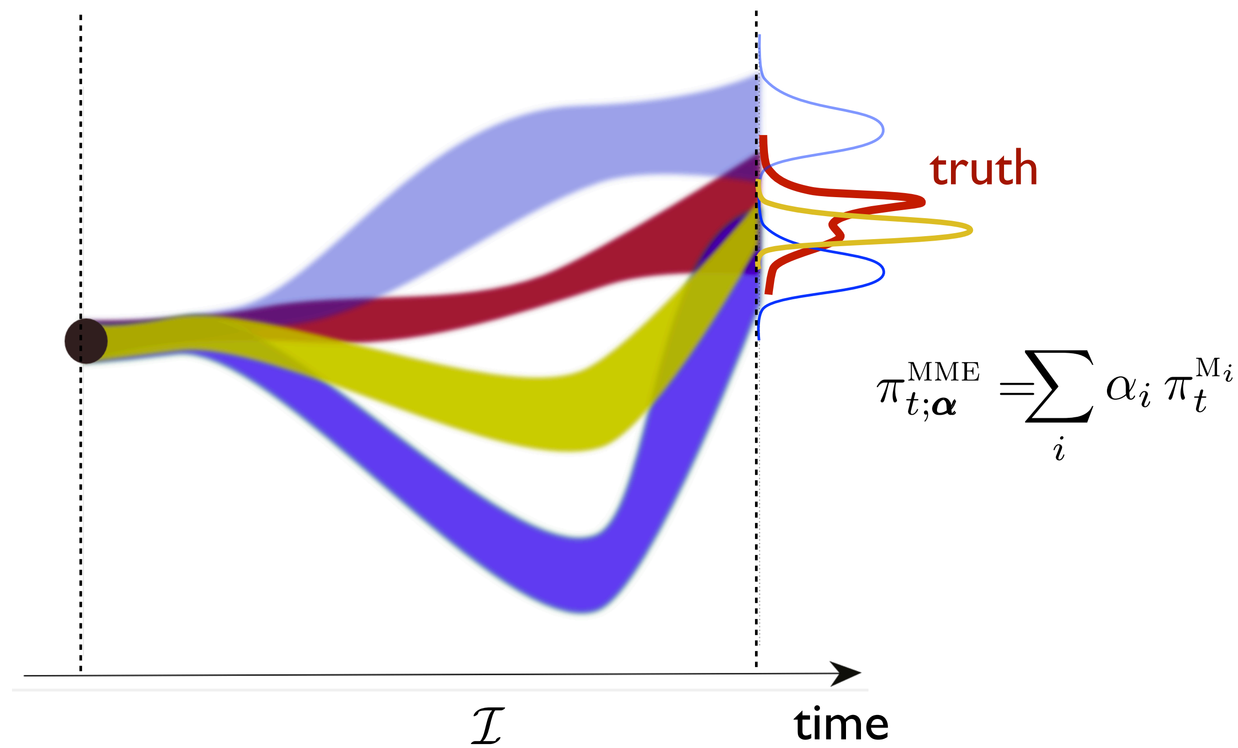
$$D^{\mathcal{F}}_{\mathrm{KL}}(\pi^{\mathrm{L},\delta} \,\|\, \pi^{\mathrm{MME},\delta}_{\alpha}) \leqslant \Big\| \sum \alpha_i \boldsymbol{\theta}^{\mathrm{M}}_i - \boldsymbol{\theta} \Big\|^{1/2}_{L^2(\mathcal{T})} \big\| \bar{\boldsymbol{E}}^{\delta} \big\|^{1/2}_{L^2(\mathcal{F})} + \mathcal{O}\left((\delta \bar{\boldsymbol{E}})^2\right)$$

$$\bar{\boldsymbol{E}}^{\delta} = \bar{\boldsymbol{E}} + \delta \bar{\boldsymbol{E}} \qquad \bar{\boldsymbol{E}}^{\delta} \equiv \int_{\Omega} \boldsymbol{E}(\boldsymbol{u}) \pi^{\delta}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$$

$$\pi^{\delta} \propto \exp\left(-\, \boldsymbol{\theta}^{\delta}(\bar{\boldsymbol{E}}^{\delta}) \cdot \boldsymbol{E}(\boldsymbol{u})\right)$$

Multi-model Ensemble (MME) predictions

- Use a single model or a mixture of models for best predictions ?



- Useful properties: Convexity of the relative entropy in the second argument & the ‘triangle’ inequality

$$D^{\mathcal{I}}_{\mathrm{KL}}\Big(\pi \,\|\, \sum_i \alpha_i \pi^{\mathrm{M}_i}\Big) \leqslant \sum_i \alpha_i D^{\mathcal{I}}_{\mathrm{KL}}\Big(\pi \,\|\, \pi^{\mathrm{M}_i}\Big), \qquad \alpha_i \geqslant 0\,, \quad \sum_i \alpha_i = 1\ .$$

$$D^{\mathcal{I}}_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}_i}) \leqslant D^{\mathcal{I}}_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{L}}) + D^{\mathcal{I}}_{\mathrm{KL}}(\pi^{\mathrm{L}} \,\|\, \pi^{\mathrm{M}_i})$$

Multi-model Ensemble predictions

- Why relative entropy?

$$D_{\mathrm{KL}}\Big(\pi \,\|\, \sum_i \alpha_i \pi^{\mathrm{M}_i}\Big) \leqslant \sum_i \alpha_i D_{\mathrm{KL}}\Big(\pi \,\|\, \pi^{\mathrm{M}_i}\Big), \qquad D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}_i}) \leqslant D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{L}}) + D_{\mathrm{KL}}(\pi^{\mathrm{L}} \,\|\, \pi^{\mathrm{M}_i})$$

Gives good bounds on predictive skill for attractor perturbations.

- Use a mixture of models instead of a single model when (necessary cond.)

$$D^{\mathcal{I}}_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{MME}}_{\alpha}) - D^{\mathcal{I}}_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}_\diamond}) < 0 \qquad \pi^{\mathrm{MME}}_t = \sum_i \alpha_i \; \pi^{\mathrm{M}_i}_t$$

- Sufficient condition for using MME given only the error of individual models

$$D^{\mathcal{I}}_{\mathrm{KL}}(\pi^{\mathrm{L}} \| \, \pi^{\mathrm{M}_\diamond}) > \sum_{i \neq \diamond} \beta_i \, D^{\mathcal{I}}_{\mathrm{KL}}(\pi^{\mathrm{L}} \| \, \pi^{\mathrm{M}_i}) \qquad \beta_i = \frac{\alpha_i}{1-\alpha_\diamond}$$

- Weaker criterion for prediction improvement using MME given error of individual models.

$$D^{\mathcal{I}}_{\mathrm{KL}}(\pi^{\mathrm{L}} \| \, \pi^{\mathrm{M}_\diamond}) + \Delta > \sum_{i \neq \diamond} \beta_i \, D^{\mathcal{I}}_{\mathrm{KL}}(\pi^{\mathrm{L}} \| \, \pi^{\mathrm{M}_i}) \qquad \beta_i = \frac{\alpha_i}{1-\alpha_\diamond}$$

Improving imperfect predictions via the MME approach

General formulation (see Branicki & Majda, J. Nonlin. Sci. 2015)

Theorem. The sufficient condition improvement of imperfect predictions via the MME approach can be expressed in terms of the least-biased densities as

$$\mathscr{A}_{\boldsymbol{\alpha}}\Big(\pi_t^{\mathrm{L1}}, \{\pi_t^{\mathrm{M}_i,\mathrm{L2}}/\pi_t^{\mathrm{M}_i}\}\Big) + \mathscr{B}_{\boldsymbol{\alpha}}\Big(\{\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}\}\Big) + \mathscr{C}_{\boldsymbol{\alpha}}\Big(\overline{\boldsymbol{E}}_t, \{\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}\}\Big) > 0,$$

where

$$\mathscr{A}_{\boldsymbol{\alpha}} = \int \mathrm{d}\boldsymbol{u}\, \pi_t^{\mathrm{L1}}(\boldsymbol{u})\, \mathfrak{M}(\boldsymbol{u}), \qquad \mathfrak{M}(\boldsymbol{u}) = \sum_{i\neq\diamond} \beta_i \left[\log \frac{\pi_t^{\mathrm{M}_i,\mathrm{L2}}(\boldsymbol{u})}{\pi_t^{\mathrm{M}_i}(\boldsymbol{u})} - \log \frac{\pi^{\mathrm{M}_\diamond,\mathrm{L2}}(\boldsymbol{u})}{\pi_t^{\mathrm{M}_\diamond}(\boldsymbol{u})}\right],$$

is non-zero only when some of the model densities are not in the least-biased form, i.e., $\pi_t^{\mathrm{M}_i,\mathrm{L2}} \neq \pi_t^{\mathrm{M}_i}$ for some i, and

$$\mathscr{B}_{\boldsymbol{\alpha}} = \sum_{i\neq\diamond} \beta_i \left[\log C^{\mathrm{M}_\diamond,\mathrm{L2}}\big(\overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond}\big) - \log C_t^{\mathrm{M}_i,\mathrm{L2}}\big(\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}\big)\right], \qquad \mathscr{C}_{\boldsymbol{\alpha}} = \sum_{i\neq\diamond} \beta_i \left[\Big(\boldsymbol{\theta}_t^{\mathrm{M}_\diamond} - \boldsymbol{\theta}_t^{\mathrm{M}_i}\Big)\cdot \bar{\boldsymbol{E}}_t\right],$$

where π_t^{L1} and $\pi_t^{\mathrm{M},\mathrm{L2}}$ are the least biased densities

$$\pi_t^{\mathrm{L1}} = C_t^{-1} \exp\Big(-\sum_{i=1}^{\mathrm{L1}} \theta_i(t) E_i(\boldsymbol{u})\Big), \qquad \pi_t^{\mathrm{M},\mathrm{L2}} = \big(C_t^{\mathrm{M}}\big)^{-1} \exp\Big(-\sum_{i=1}^{\mathrm{L2}} \theta_i^{\mathrm{M}}(t) E_i(\boldsymbol{u})\Big)$$

maximising the Shannon entropy

$$\mathcal{S} = -\int \pi^{\mathrm{L}} \ln \pi^{\mathrm{L}} \quad \text{with} \quad \int \pi^{\mathrm{L}}(\boldsymbol{u}) E_i(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} = \int \pi(\boldsymbol{u}) E_i(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}, \quad i = 1, \ldots, \mathrm{L},$$

Condition for improving forced response predictions via MME approach

FACT 7. The sufficient condition (13) for improvement of the imperfect predictions via the MME approach can be expressed in terms of the least-biased approximations of the true density evolving from the initial density characterized by $\overline{\boldsymbol{E}}_0,\ \boldsymbol{\theta}_0$ as

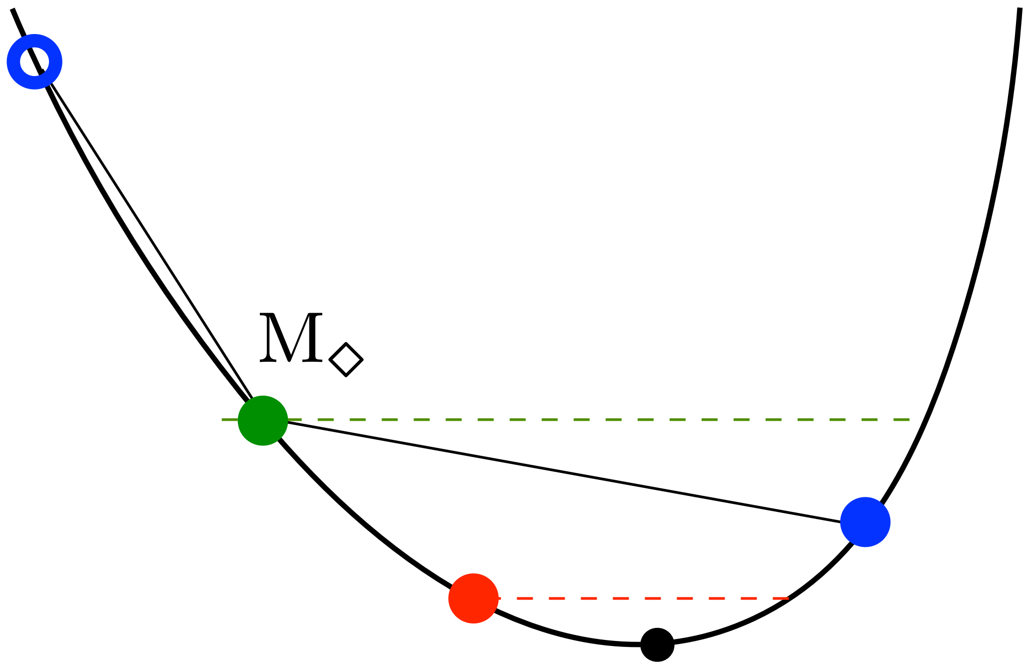
$$\boxed{\mathscr{A}_{\boldsymbol{\alpha}}\left(\pi_t^{\mathrm{L}1}, \{\pi_t^{\mathrm{M}_i,\mathrm{L}2}/\pi_t^{\mathrm{M}_i}\}\right) + \widetilde{\mathscr{B}}_{\boldsymbol{\alpha}}\left(\{\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}\}; \overline{\boldsymbol{E}}_0\right) + \widetilde{\mathscr{C}}_{\boldsymbol{\alpha}}\left(\{\overline{\boldsymbol{E}}_t^{\mathrm{M}_i}\}, \tilde{\overline{\boldsymbol{E}}}_t\right) > 0,} \tag{20}$$

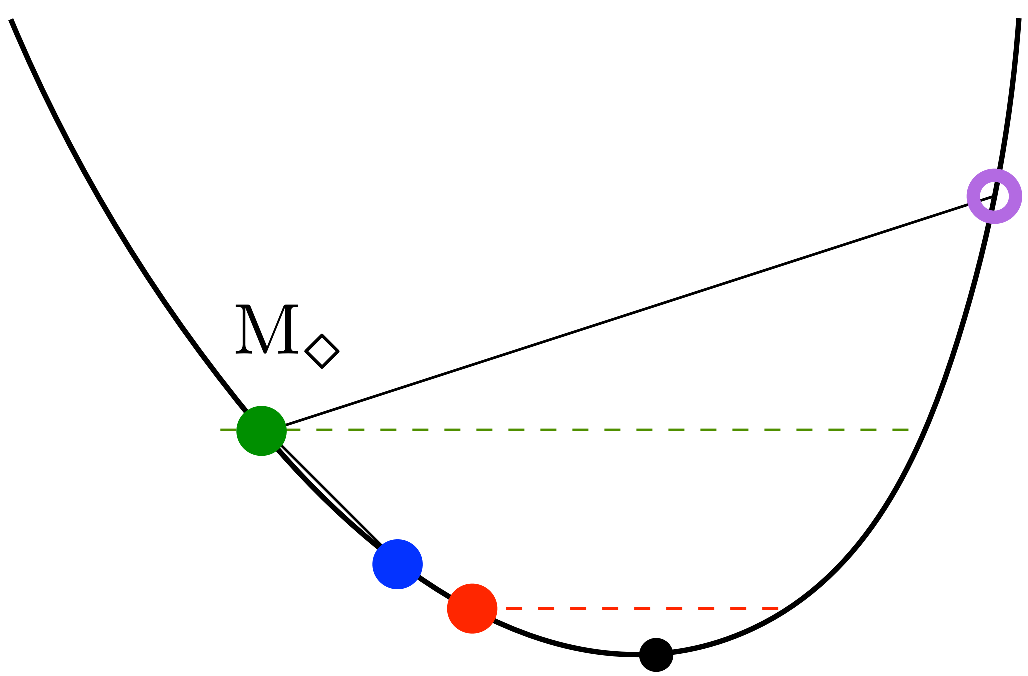
where $\mathscr{A}_{\boldsymbol{\alpha}}$ is defined in (14) and

$$\widetilde{\mathscr{B}}_{\boldsymbol{\alpha}} = \sum_{i\neq\diamond}\beta_i\left[\mathcal{P}(\pi_0^{\mathrm{L}1}, \pi_t^{\mathrm{M}_\diamond,\mathrm{L}2}) - \mathcal{P}(\pi_0^{\mathrm{L}1}, \pi_t^{\mathrm{M}_i,\mathrm{L}2})\right], \qquad \widetilde{\mathscr{C}}_{\boldsymbol{\alpha}} = \sum_{i\neq\diamond}\beta_i\left[\left(\boldsymbol{\theta}_t^{\mathrm{M}_\diamond}(\overline{\boldsymbol{E}}_t^{\mathrm{M}_\diamond}) - \boldsymbol{\theta}_t^{\mathrm{M}_i}(\overline{\boldsymbol{E}}_t^{\mathrm{M}_i})\right)\cdot\delta\tilde{\overline{\boldsymbol{E}}}_t\right],$$

where the weights β_i are defined in (9) and the initial least-biased density of the truth is given by $\pi_0^{\mathrm{L}1}(\boldsymbol{u}) = C_0^{-1}\exp(-\boldsymbol{\theta}_0\cdot\boldsymbol{E}(\boldsymbol{u}))$.

Improving imperfect predictions via the MME approach

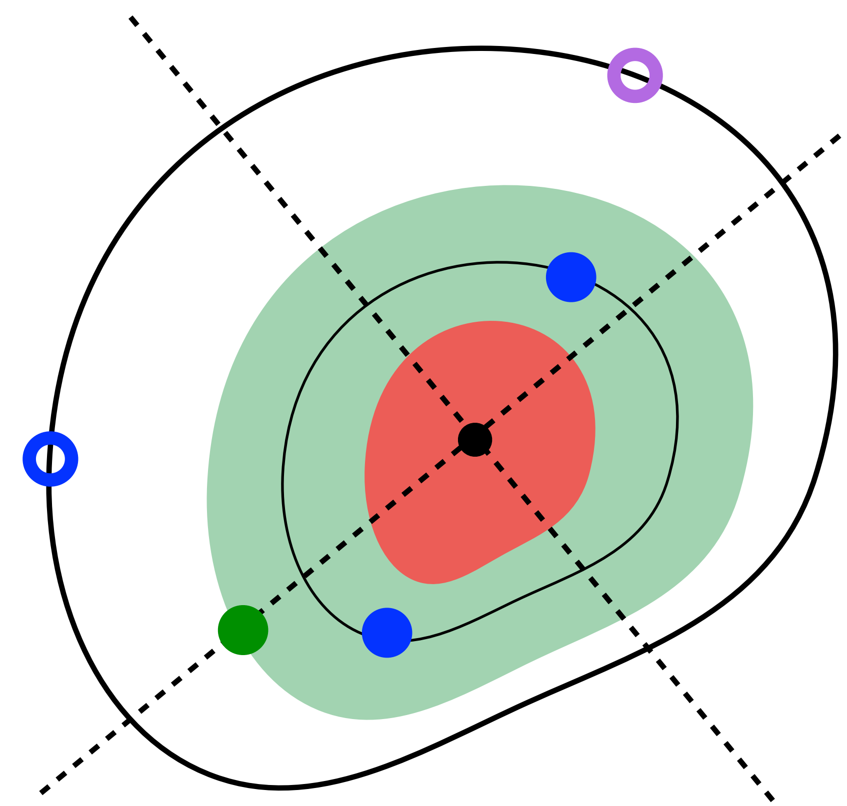




Necessary condition:

MME better than $\mathrm{M}_\diamond$ (●) when

$$D^{\mathcal{I}}_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{MME}}_{\alpha}) - D^{\mathcal{I}}_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}_\diamond}) < 0$$



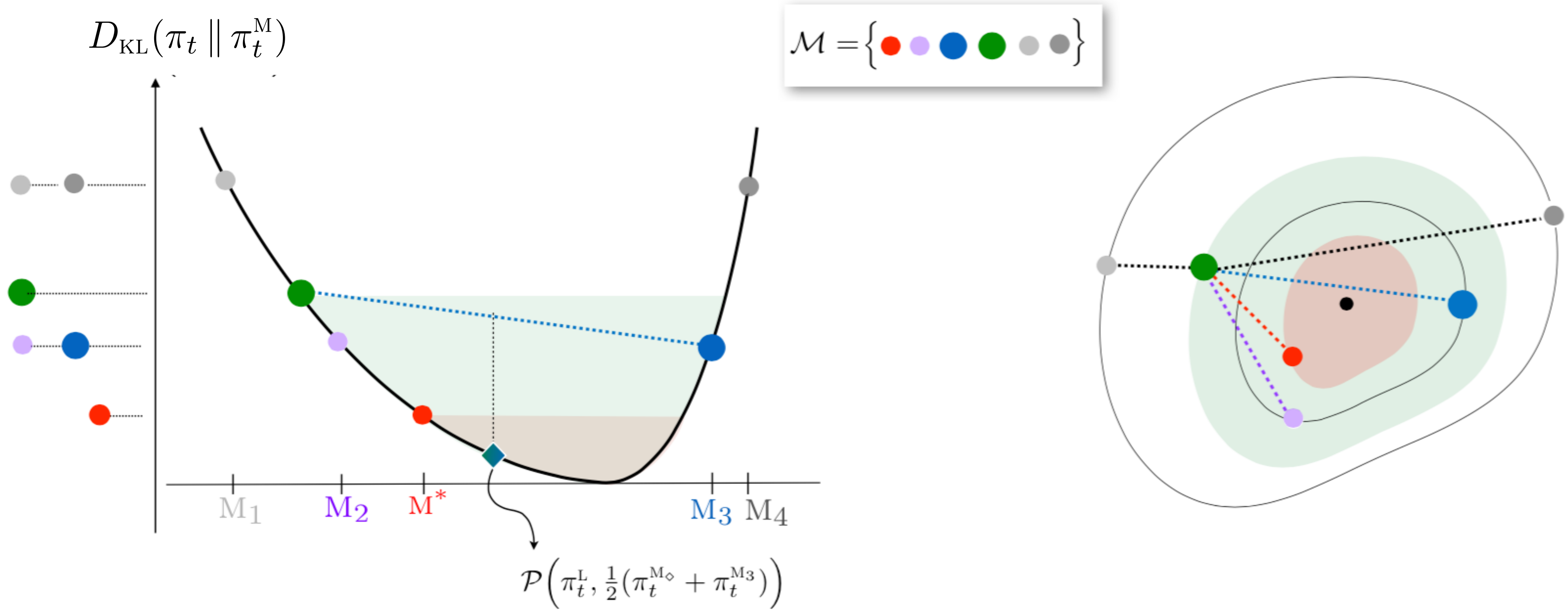
“finite-dim” sketch

Improving imperfect predictions via the MME approach

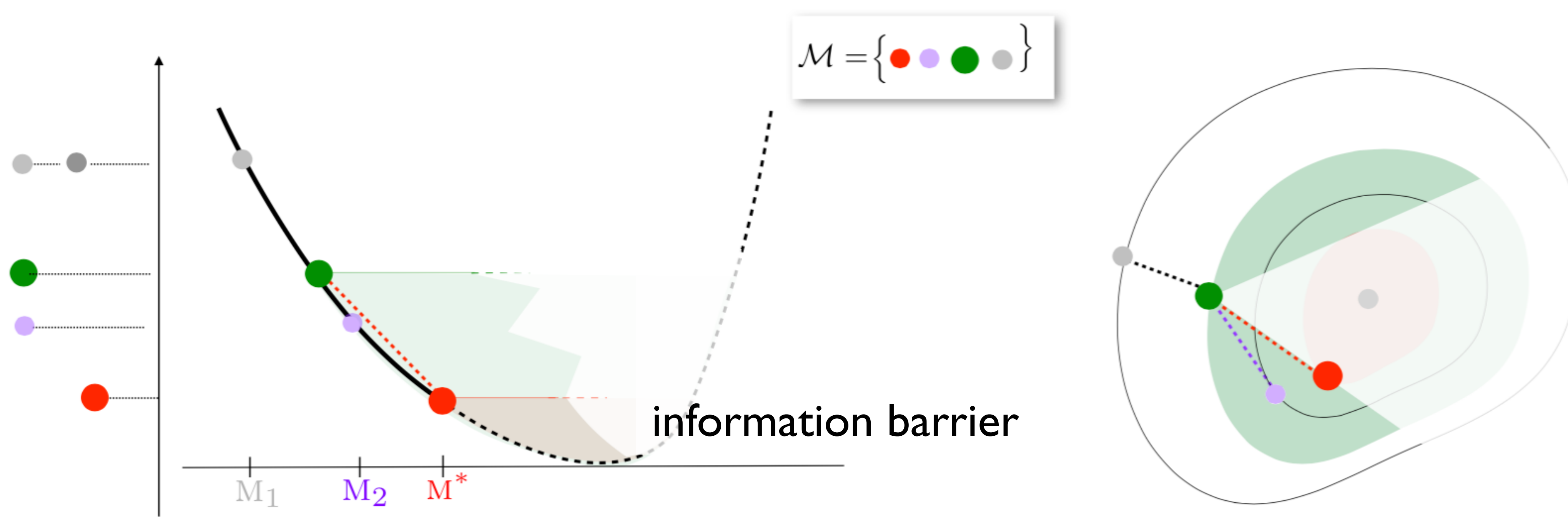
Use a single model or a mixture of models for best predictions ?

$$D_{\mathrm{KL}}^{\mathcal{I}}(\pi^{\mathrm{L}} \| \pi^{\mathrm{M}_\diamond}) > \sum_{i \neq \diamond} \beta_i \, D_{\mathrm{KL}}^{\mathcal{I}}(\pi \| \pi^{\mathrm{M}_i})$$

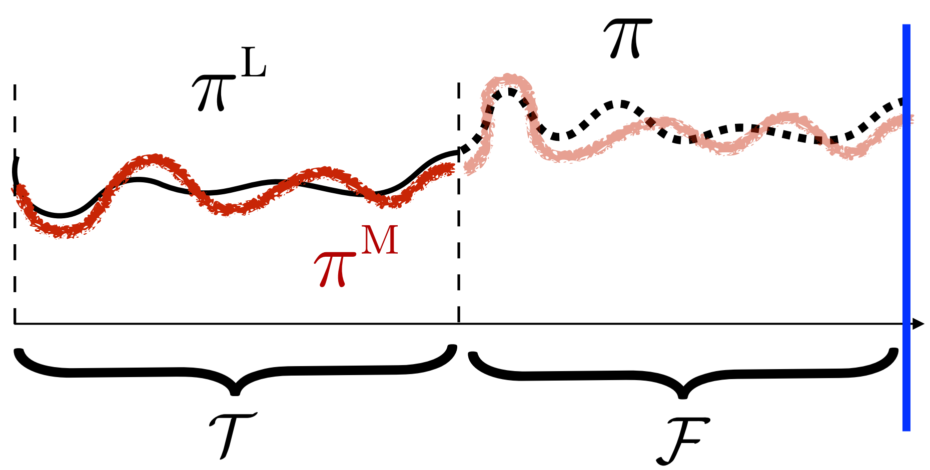
use mixture

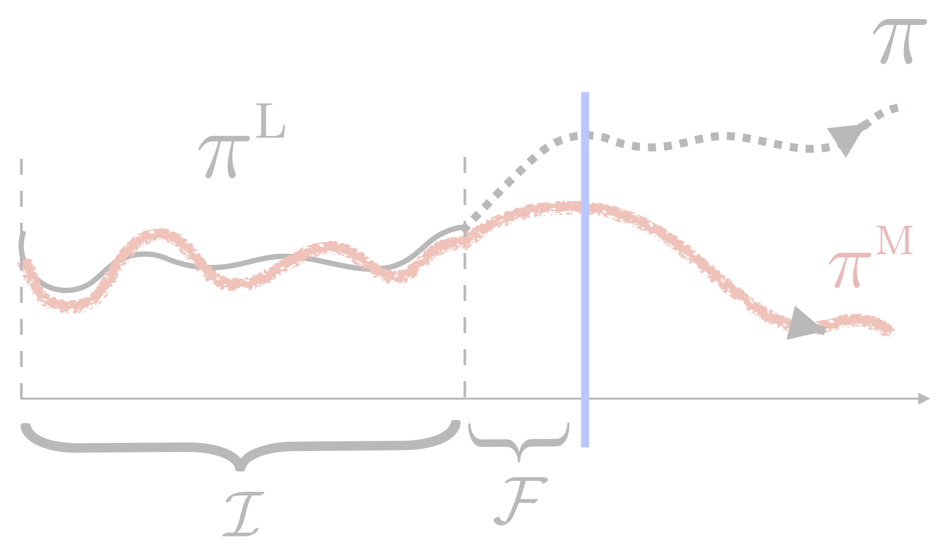


use single model



Improving imperfect predictions via improving model fidelity





FACT: Improving attractor fidelity of model improves predictions
(simplest case for least-biased/max-entropy models)

$$D_{\mathrm{KL}}^{\mathcal{F}}(\pi^{\mathrm{L},\delta} \,\|\, \pi^{\mathrm{M},\delta}) \leqslant \left\|\boldsymbol{\theta}^{\mathrm{M}} - \boldsymbol{\theta}\right\|_{L^2(\mathcal{T})}^{1/2} \left\|\bar{\boldsymbol{E}}^{\delta}\right\|_{L^2(\mathcal{F})}^{1/2} + \mathcal{O}\left((\delta\bar{\boldsymbol{E}})^2\right)$$

$$\bar{\boldsymbol{E}}^{\delta} = \bar{\boldsymbol{E}} + \delta\bar{\boldsymbol{E}} \qquad \bar{\boldsymbol{E}}^{\delta} \equiv \int_{\Omega} \boldsymbol{E}(\boldsymbol{u})\pi^{\delta}(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$$

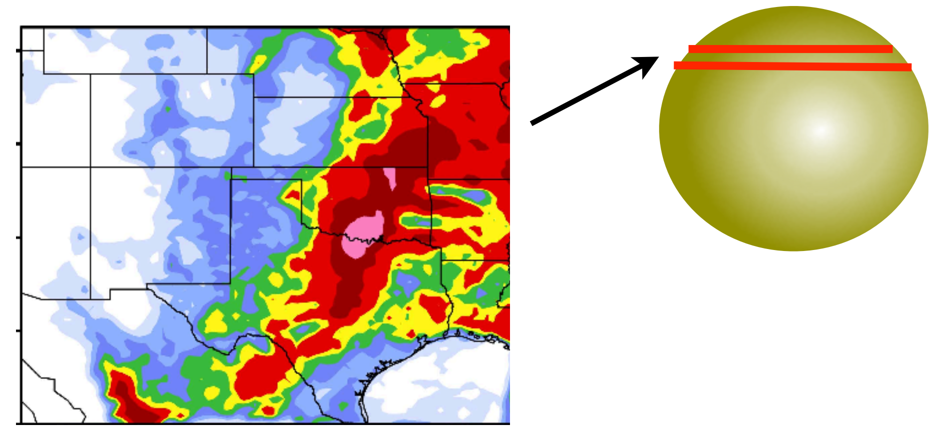
$$\pi^{\delta} \propto \exp\left(-\boldsymbol{\theta}^{\delta}(\bar{\boldsymbol{E}}^{\delta})\cdot\boldsymbol{E}(\boldsymbol{u})\right)$$

More details:

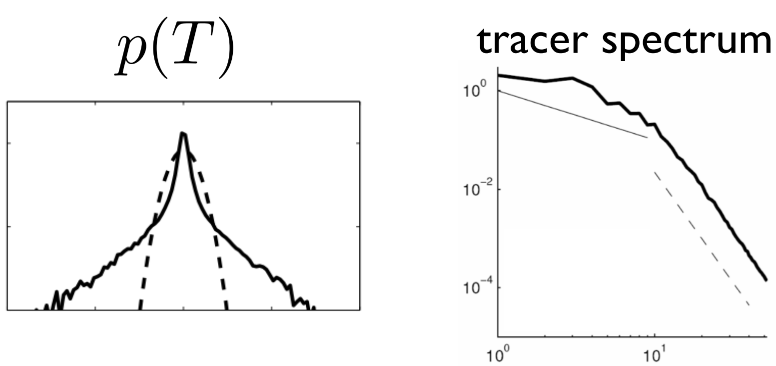
Branicki, Enc. Appl. Math, 2015

Exactly solvable test models for turbulent tracer with realistic features

$$\frac{\partial T}{\partial t} + \boldsymbol{v}(\boldsymbol{x},t)\cdot\nabla T = \kappa \Delta T$$

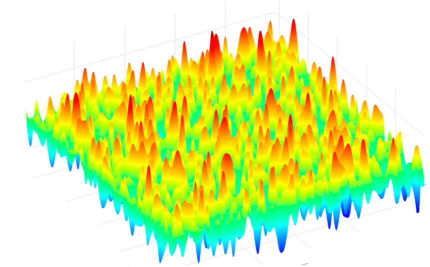


Non-Gaussian passive tracer with mean gradient $T = \alpha y + T'(x,t)$



Reduced-order model & stochastic parameterisation

$$\frac{\partial T^M}{\partial t} + \bar{\boldsymbol{v}}^M\cdot\nabla T^M = (\kappa + \kappa_{eddy})\ \Delta T^M + \sigma_T \dot{W}$$



Model improvement

$$D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}*}) = \min_{\mathrm{M}\in\mathcal{M}} D_{\mathrm{KL}}(\pi \,\|\, \pi^{\mathrm{M}})$$

Majda & Branicki, DCDS 2012

Exactly solvable test models for turbulent tracer with realistic features

$$\partial_t T + \boldsymbol{v}(\boldsymbol{x},t)\cdot\nabla T = \kappa\Delta T$$

Majda & Gershgorin, Proc. Roy.Soc. 2011

Majda & Branicki DCDS 2012

- Physical space

$$\frac{\partial T'}{\partial t} = Q\left(\frac{\partial}{\partial x}, U(t)\right)T' - \alpha v(x,t) + \kappa\frac{\partial^2 T}{\partial x^2},$$

$$\frac{\partial U}{\partial t} = -d_U U + f(t) + \sigma_U \dot{W}_U(t),$$

$$\frac{\partial v}{\partial t} = P\left(\frac{\partial}{\partial x}, U(t)\right)v + f_v(t) + b_v(x,t) + \sigma_v(x)\dot{W}_v(t),$$

- Fourier domain

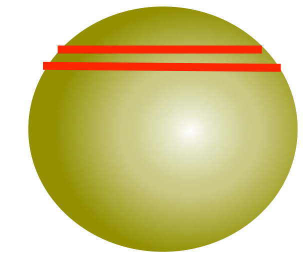
$$\dot{T}'_k(t) = (-d_{T_k} + i\omega_{T_k}(t))T'_k(t) - \alpha v_k(t),$$

$$\dot{U}(t) = -d_U U(t) + f_U(t) + \sigma_U \dot{W}_U(t),$$

$$\dot{v}_k(t) = (-d_{v_k} - \gamma_{v_k}(t) + i\omega_{v_k}(t))v_k(t) + b_{v_k}(t) + f_{v_k}(t) + \sigma_{v_k}\dot{W}_{v_k}(t),$$

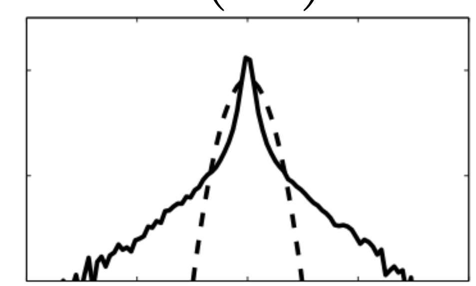
$$\dot{\gamma}_{v_k}(t) = -d_{\gamma_{v_k}}\gamma_{v_k}(t) + \sigma_{\gamma_{v_k}}\dot{W}_{\gamma_{v_k}}(t),$$

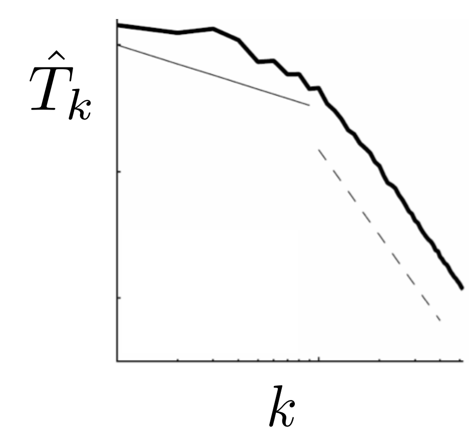
$$\dot{b}_{v_k}(t) = (-d_{b_{v_k}} + i\omega_{b_{v_k}})b_{v_k}(t) + \sigma_{b_{v_k}}\dot{W}_{b_{v_k}}(t),$$



$$T = \alpha y + T'(x,t)$$

$$\pi(T')$$





- Identification of mechanisms for intermittency
- Rigorous justification/critique of various turbulent closures
- Non-local effects due to mean flow - fluctuation interactions

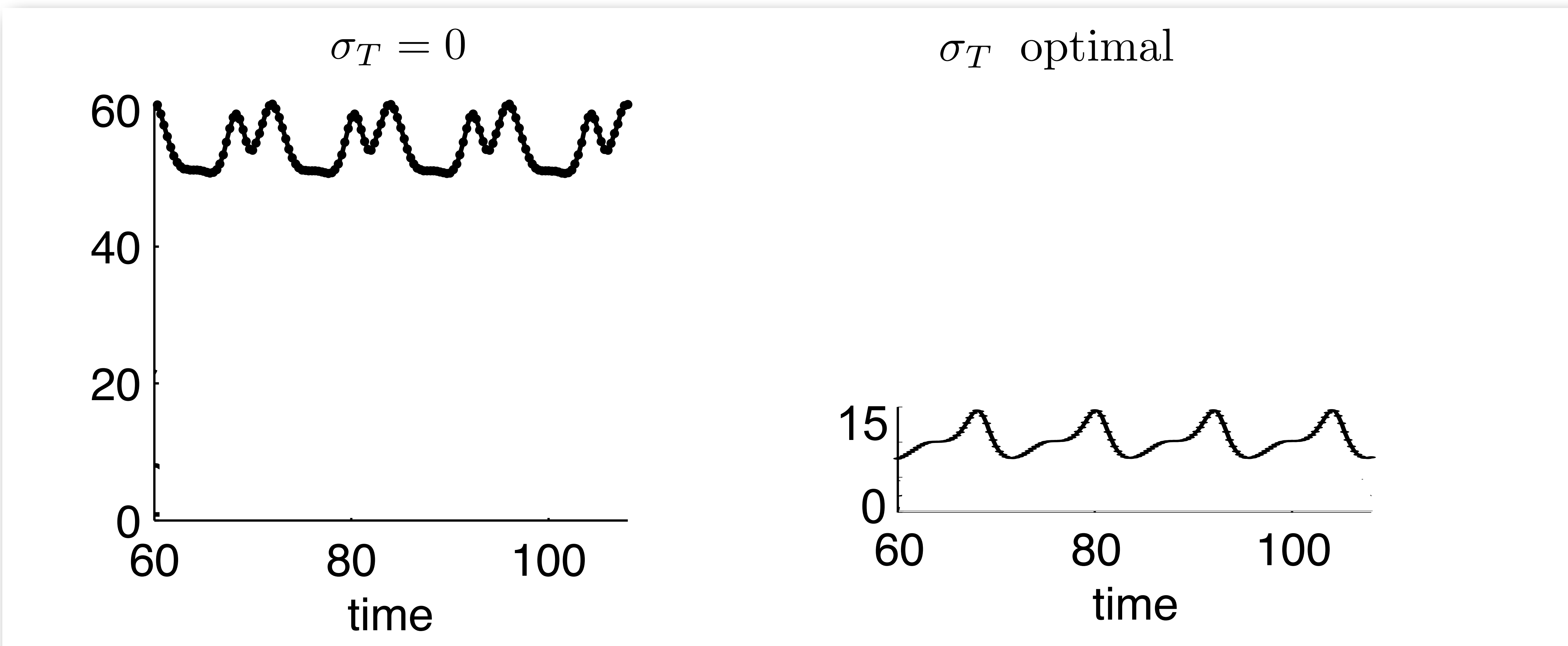
Improving reduced-order models for turbulent tracer

Baby configuration: **model improvement on attractor by simple noise inflation**

$$\partial_t T + \boldsymbol{v} \cdot \nabla T = \kappa \Delta T$$

$$\partial_t T^{\mathrm{M}} + \boldsymbol{v}^{\mathrm{M}} \cdot \nabla T^{\mathrm{M}} = \tilde{\kappa} \Delta T^{\mathrm{M}} + \sigma_T^* \dot{\tilde{W}}$$

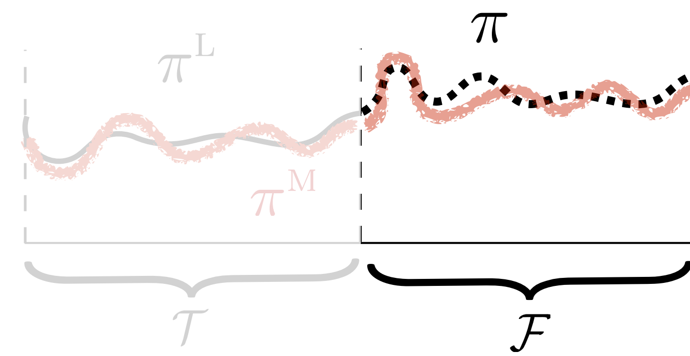
$$D_{\mathrm{KL}}^{\mathcal{I}}(\pi_{att} \,\|\, \pi_{att}^{\mathrm{M}})$$

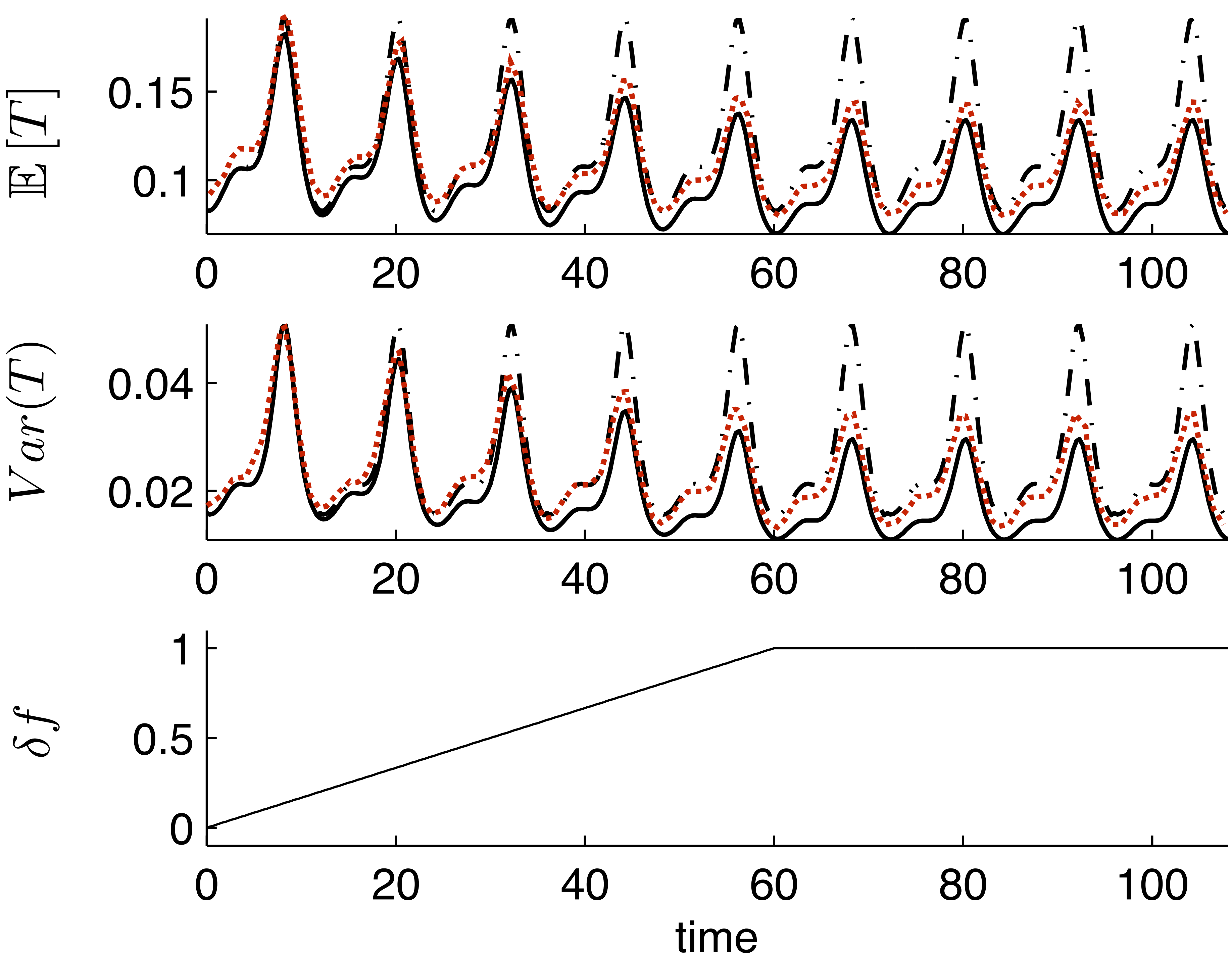


Model **error on attractor** for models **with optimised noise** is **greatly reduced**

Improving reduced-order models for turbulent tracer

Forced response for attractor-tuned **model with optimal noise**

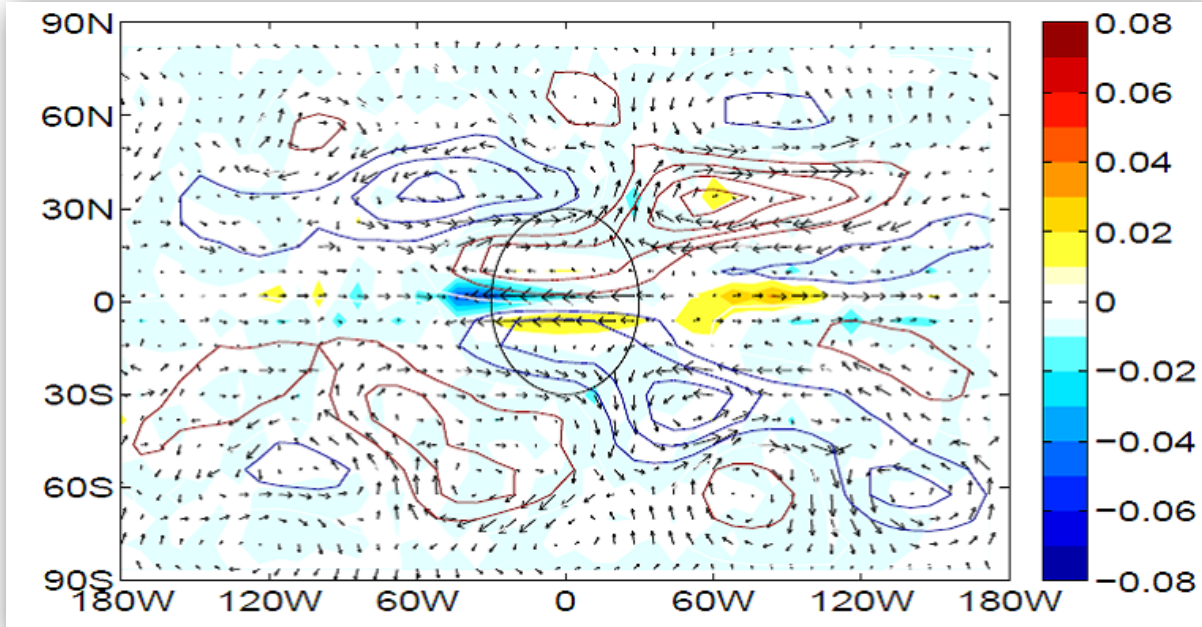




Information-theoretic improvement of predictive skill of GCMs

joint work with R. Leung, S. Hagos, G. Lin & A. Majda

(stalled for now …)



$$D_{KL}(\pi_t^{\delta} \| \pi_t^{\mathrm{M},\delta}) = H\left({}^{\{\mathrm{G}\}}\pi_t^{\delta}\right) - H\left(\pi_t^{\delta}\right) + \frac{1}{2\sigma^2}\left(\int_0^t (\mathcal{R}_{\bar{u}}(t-s) - \mathcal{R}_{\bar{u}}^{\mathrm{M}}(t-s))\delta f(s)\mathrm{d}s\right)^2 + \frac{1}{4\sigma^4}\left(\int_0^t (\mathcal{R}_{\sigma^2}(t-s) - \mathcal{R}_{\sigma^2}^{\mathrm{M}}(t-s))\delta f(s)\mathrm{d}s\right)^2 + \mathcal{O}(\delta^3)$$

"Climate change" error

FDT

FDT (fluctuation-dissipation relationships) ⟶ $\delta\bar{\boldsymbol{u}} = \int_{t_0}^{t} \mathcal{R}_{\bar{\boldsymbol{u}}}(t-s)\delta f(t)\mathrm{d}s \qquad \delta R = \int_{t_0}^{t} \mathcal{R}_R(t-s)\delta f(t)\mathrm{d}s$

Summary:

- **Natural synergy between the information theoretic framework and empirical data**
- **Systematic framework for dimensionality reduction and 'information retainment'** depending **on amount/quality of available data** and computational cost
- **Information-theoretic framework is useful for UQ on reduced subspaces** of dynamical variables
  - The framework naturally suited to deal with **model error** and **partial observability** of the true dynamics
  - **Information-theoretic optimization** of imperfect models requires **simultaneous tuning of statistical moments** and can significantly improve prediction skill and sensitivity of imperfect models
- If correctly implemented, the **MME framework** is **useful for improving forced response** of the unknown truth dynamics based solely on the **information from its statistical equilibrium**
- **Sufficient condition** for **improving imperfect predictions via MME** approach obtained within the information-theoretic framework
- This formulation **can be extended to MME prediction** with **filtering/data assimilation** algorithms
- **Pathspace framework in development**, including more detailed measures of predictive fidelity

References:

Branicki, Information theory in prediction of complex systems, Enc. Applied Math, 2015

Branicki & Majda, An information-theoretic framework for improving Multi-Model Ensemble forecasts, J. Nonlin. Sci., 2015

Branicki & Majda, Quantifying Bayesian filter performance for turbulent dynamical systems via Information theory, Comm. Math. Sci, 2014

Branicki & Majda, Quantifying uncertainty for predictions with model errors in non-Gaussian models with intermittency, Nonlinearity, 2012

Majda & Branicki, Lessons in UQ for Turbulent Dynamical Systems, DCDS 2012

Branicki, Chen & Majda, Non-Gaussian test models for prediction and state estimation with model errors, Chineese Ann. Math 2013

Majda & Gershgorin, The Link Between Statistical Equilibrium Fidelity and Forecasting Skill for Complex Systems with Model Error, PNAS 2011

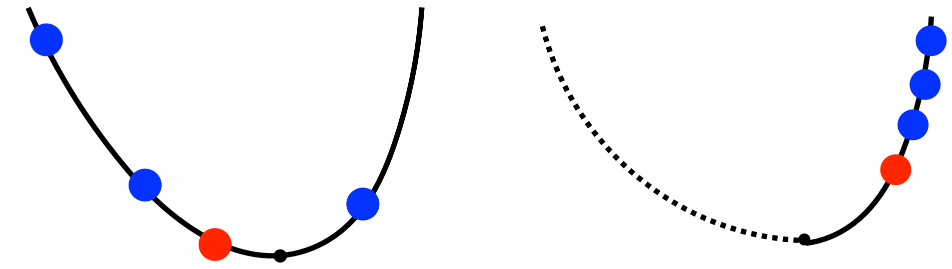
Majda & Gershgorin, Improving Model Fidelity and Sensitivity for Complex Systems through Empirical Information Theory, PNAS 2011

- Extras:

  - Model error reduction, tuning and information barriers:
    A simple example linear Gaussian example

  - Improving "climate change" predictions by tuning on attractor:
    Linear response theory & fluctuation-dissipation constraints.

Model error reduction, tuning and information barriers: A simple example linear Gaussian example



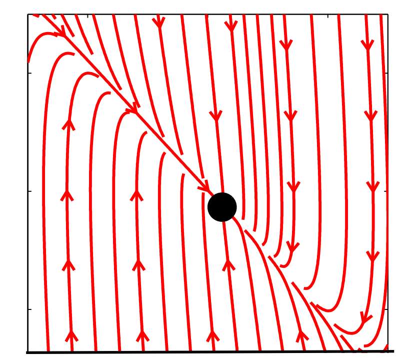
- Perfect model

$$\dot{u} = au + v + F \qquad \text{‘resolved’ dynamics}$$

$$\dot{v} = qu + Av + \sigma \dot{W} \qquad \text{‘unresolved’ dynamics}$$

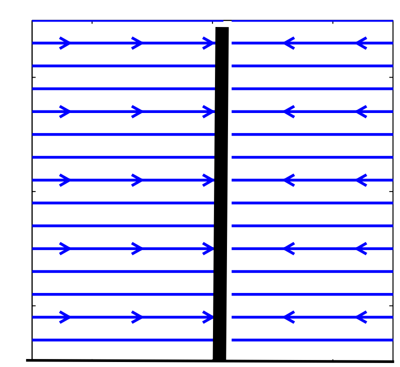
- Gaussian equilibrium if

$$a + A < 0, \quad aA - q > 0.$$



- Imperfect model: Mean Stochastic Model

$$\dot{u}_{\mathrm{M}} = -\gamma_{\mathrm{M}} u_{\mathrm{M}} + F_{\mathrm{M}} + \sigma_{\mathrm{M}} \dot{W}_{\mathrm{M}}$$



Tuning the marginal statistics on attractor

- The imperfect model

$$\dot{u}_M = -\gamma_M u_M + F_M + \sigma_M \dot{W}_M$$

- Tuning the imperfect model equilibrium statistics

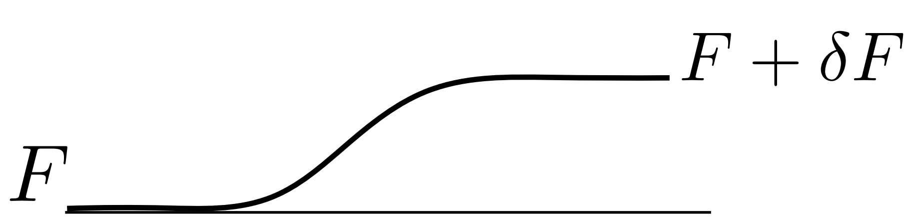
$$\bar{u}_M \quad \Big| \quad \frac{F_{M*}}{\gamma_M} = -\frac{AF}{aA-q}$$

$$Var[u_M] \quad \Big| \quad \frac{\sigma^2_{M*}}{2\,\gamma_M} = -\frac{\sigma^2}{2(aA-q)(a+A)}$$

(F_{M*}, σ_{M*}) fixed

γ_M free

- Infinite-time response to change in forcing

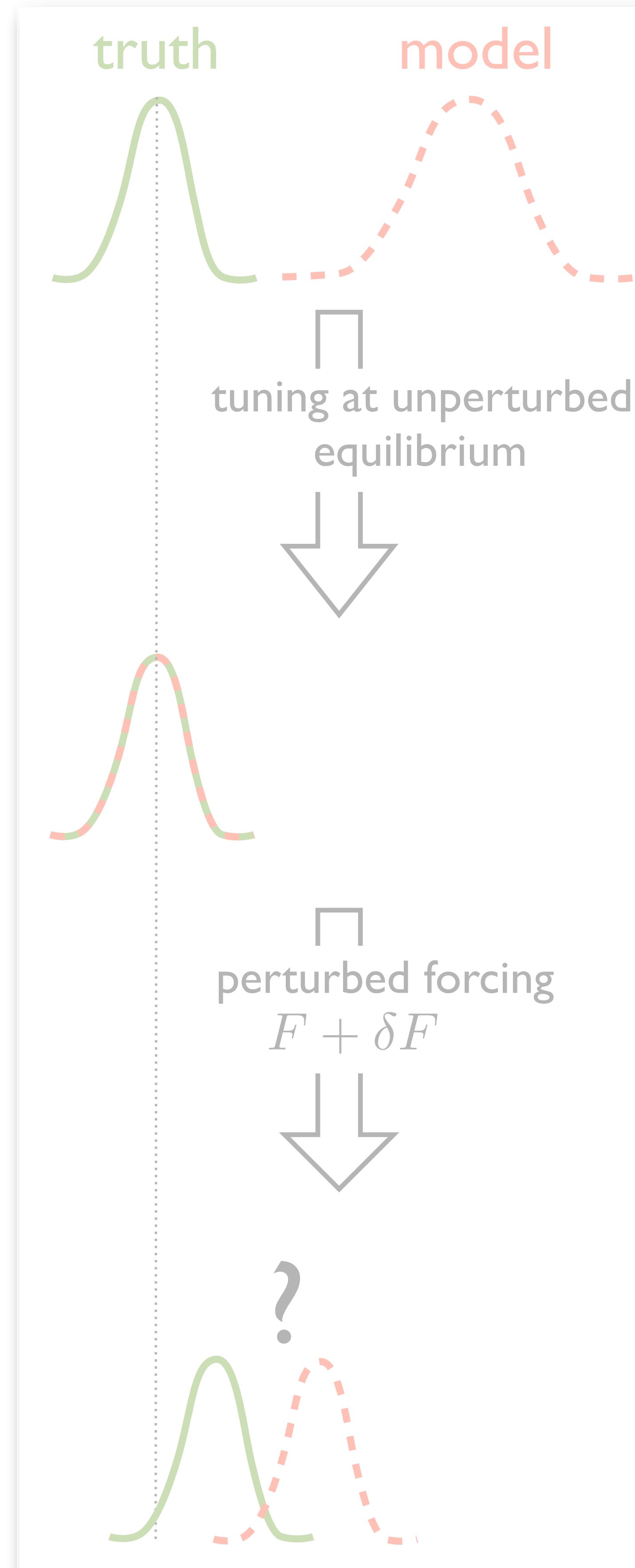


Perfect model

$$\delta \bar{u}^{\infty} = -\frac{A}{aA-q}\delta F$$

Imperfect model

$$\delta \bar{u}^{\infty}_{M*} = \frac{1}{\gamma_M}\delta F$$



Model error & information barriers

$$\dot{u} = au + v + F$$
$$\dot{v} = qu + Av + \sigma\dot{W}$$

$$\dot{u}_{\mathrm{M}} = -\gamma_{\mathrm{M}} u_{\mathrm{M}} + F_{\mathrm{M}*} + \sigma_{\mathrm{M}*}\dot{W}_{\mathrm{M}}$$

Model error on the perturbed attractor

$$\mathcal{P}(\pi_{\delta F}, \pi_{\delta F}^{\mathrm{M}*}) \propto \left| \frac{A}{aA - q} + \frac{1}{\gamma_{\mathrm{M}}} \right|^2 |\delta F|^2$$

$$aA - q > 0$$
$$\gamma_{\mathrm{M}} > 0$$

- $A > 0$: **Intrinsic barrier to improving sensitivity**

  No minimum of $\mathcal{P}$ for finite $\gamma_{\mathrm{M}} > 0$

- $A < 0$: **Perturbed attractor fidelity and sensitivity captured for**

$$\gamma^{\mathrm{M}_{prf}} = -A^{-1}(aA - q)$$

More details in:

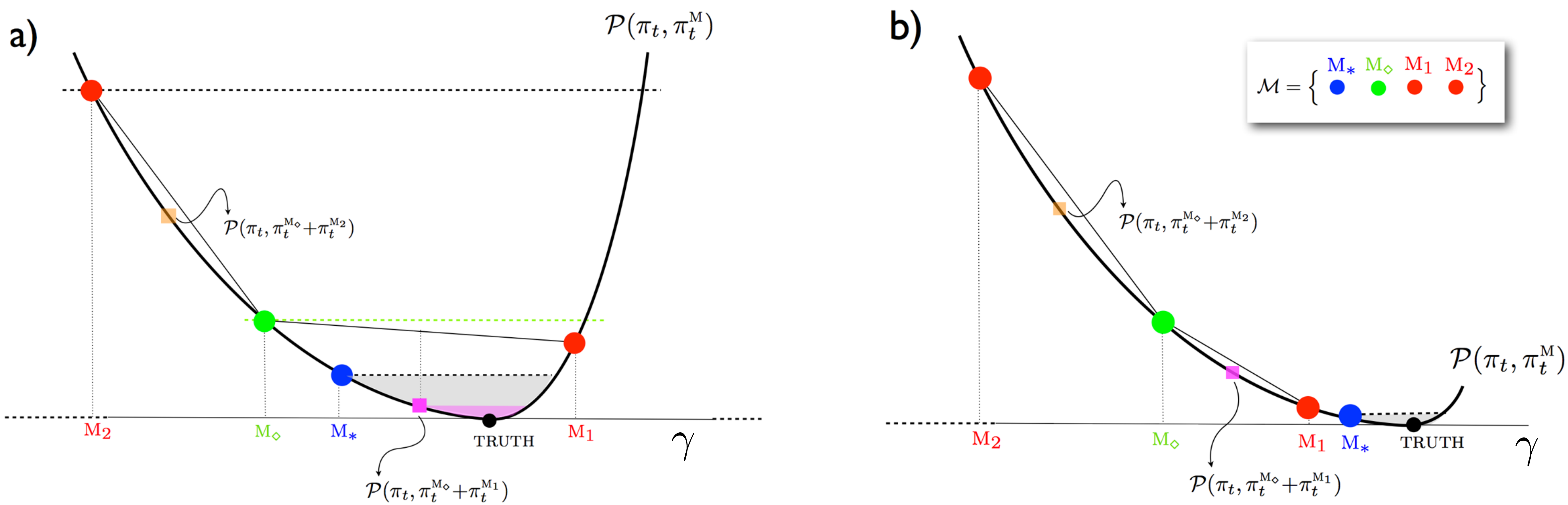
Majda & Branicki, Lessons in Uncertainty Quantification for Turbulent Dynamical Systems, DCDS 2012

Branicki & Majda, Quantifying uncertainty for predictions with model errors in non-Gaussian models with intermittency, Nonlinearity, 2012

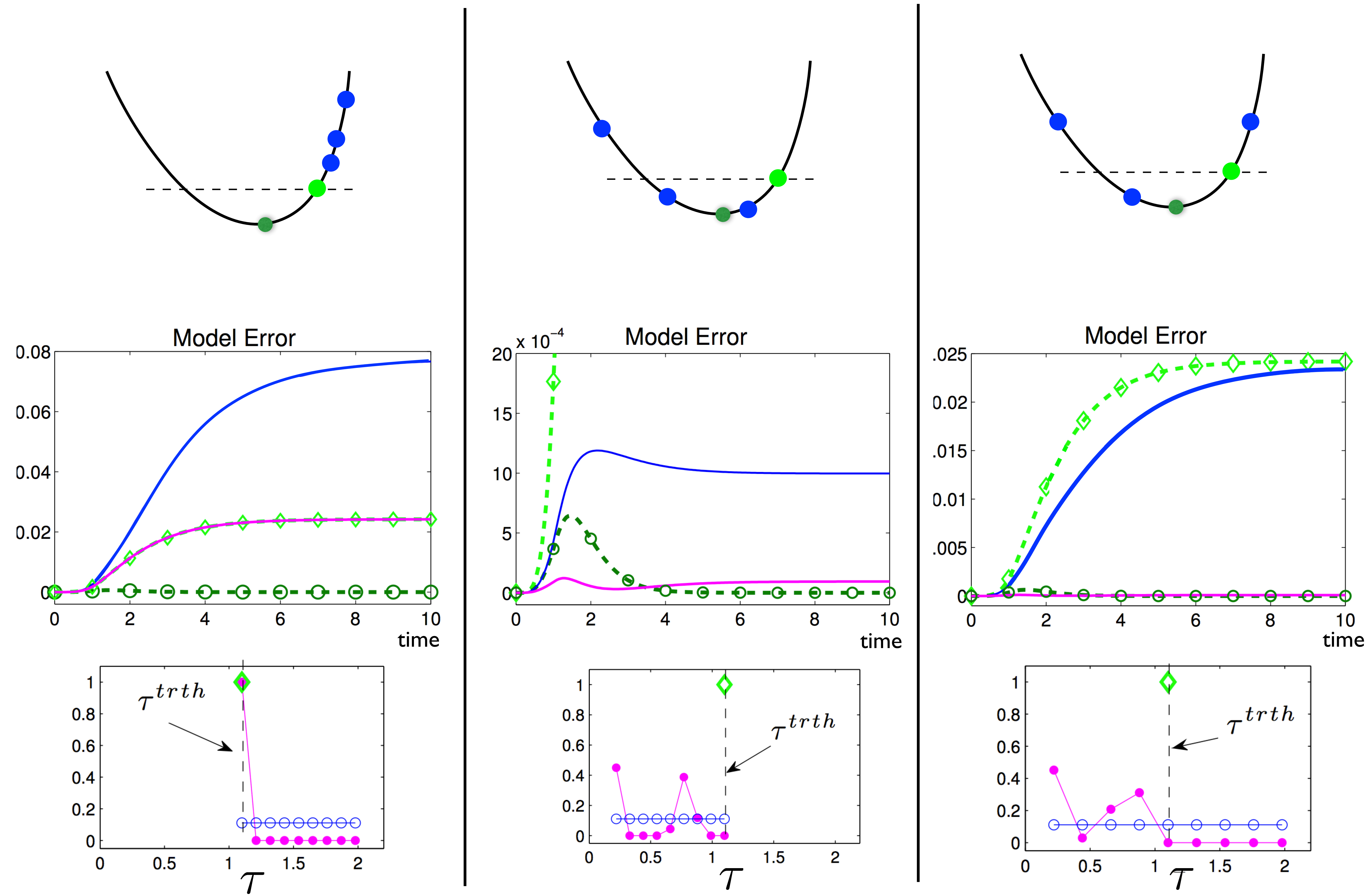
Model error & information barriers in MME prediction

$$\mathcal{P}(\pi_t^{\mathrm{L}}, \pi_t^{\mathrm{M}_\diamond}) > \sum_{i \neq \diamond} \beta_i \, \mathcal{P}(\pi_t^{\mathrm{L}}, \pi_t^{\mathrm{M}_i})$$

$$\frac{|\delta F|^2}{2E} \sum_{i \neq \diamond} \beta_i \left[\left(\frac{A}{aA - q} + \frac{1}{\gamma^{\mathrm{M}_\diamond}} \right)^2 - \left(\frac{A}{aA - q} + \frac{1}{\gamma^{\mathrm{M}_i}} \right)^2 \right] > 0$$



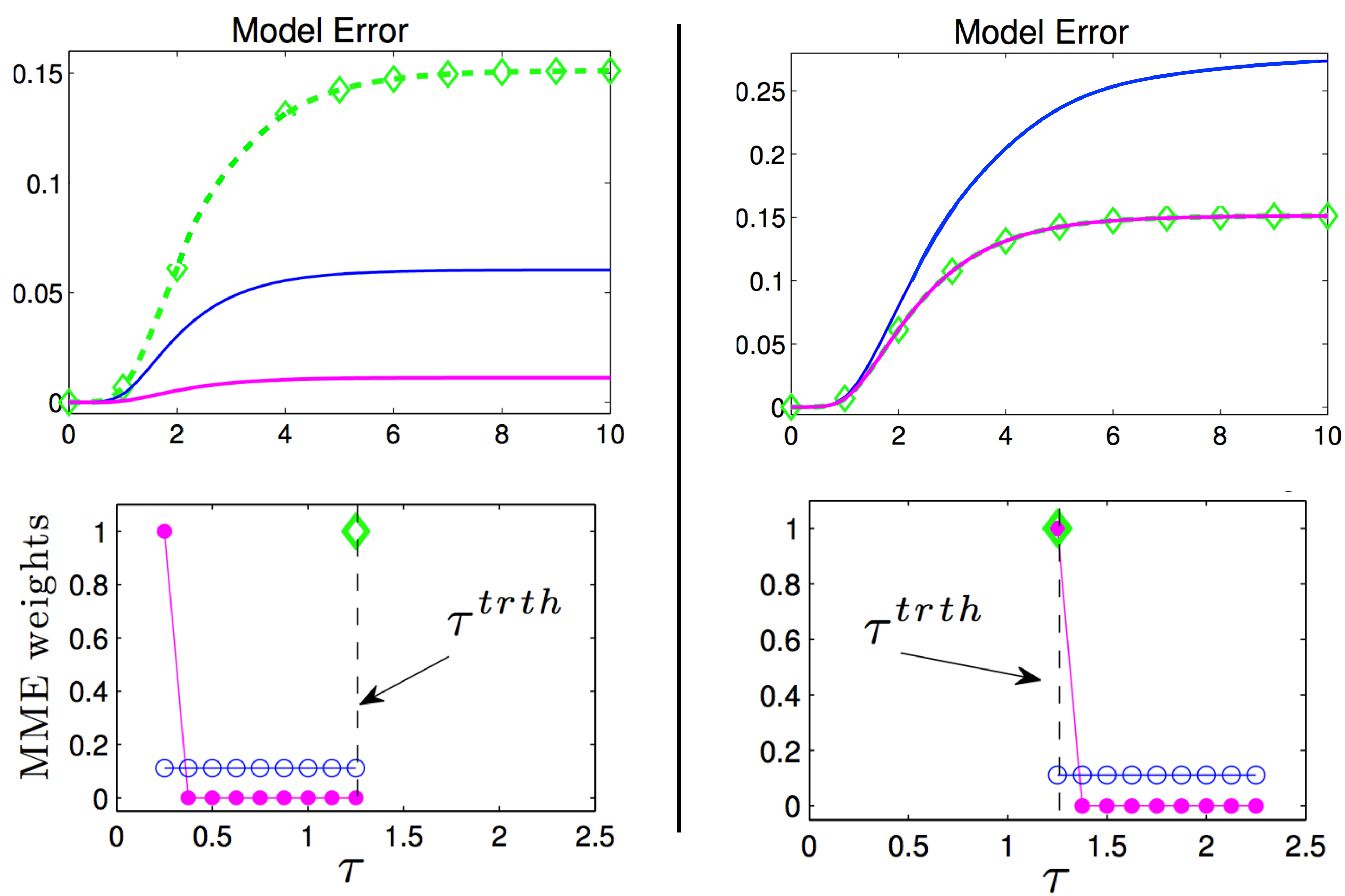
MME prediction with no information barrier



MME prediction with information barrier $A > 0$

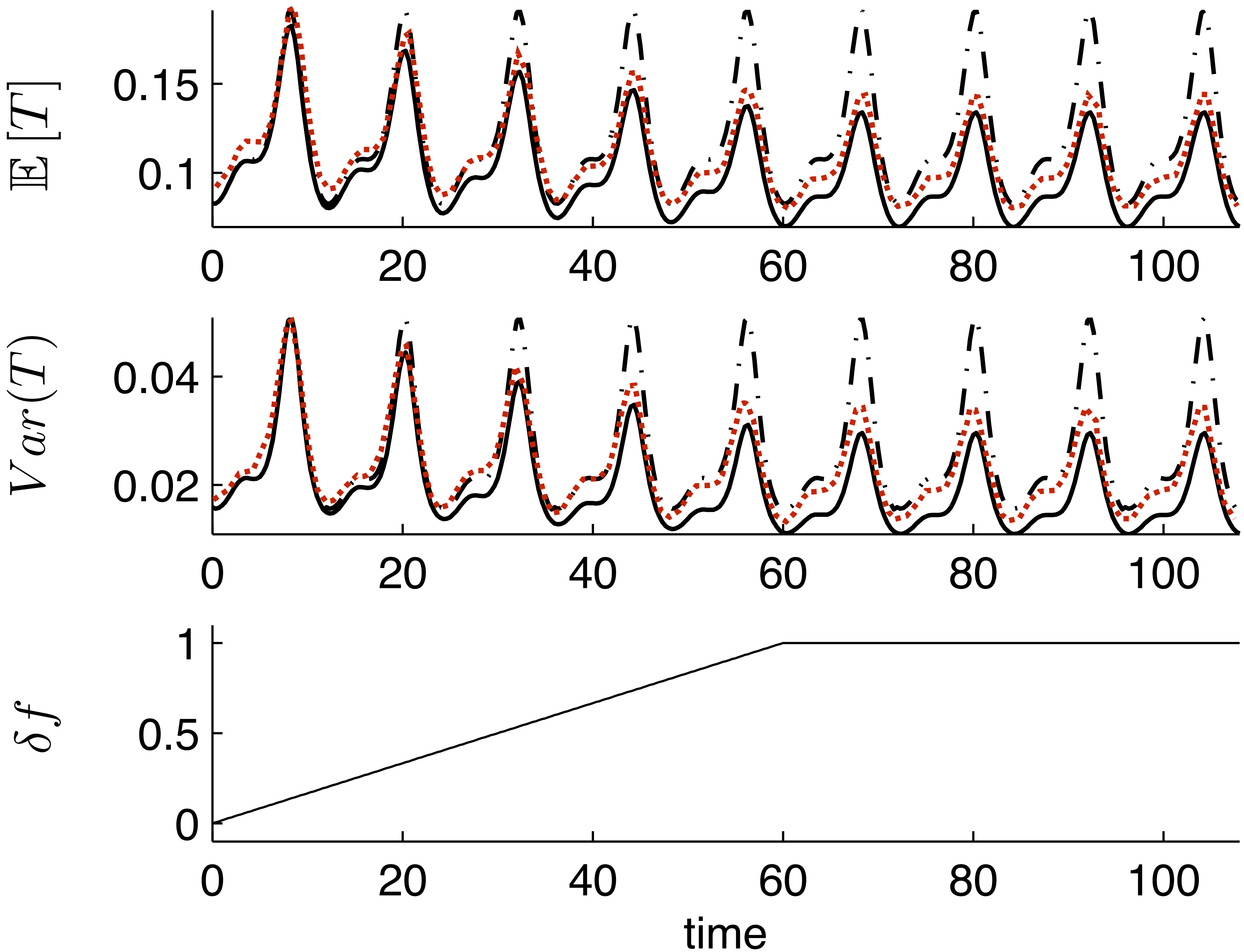
$$\dot{u} = au + v + F$$
$$\dot{v} = qu + Av + \sigma\dot{W}$$

$$\dot{u}_M = -\gamma_M u_M + F_{M*} + \sigma_{M*}\dot{W}_M$$



- The MME prediction does not reduce the information barrier
- The infinite time response can be improved for any overdamped ensemble

Improving "climate change" predictions by tuning on attractor: Linear response theory & fluctuation-dissipation constraints.



Essentials of FDT

(e.g., Majda, Abramov, Grote 2005)

- Original system $\dot{\boldsymbol{v}} = \boldsymbol{f}(\boldsymbol{v}, t) + \sigma(\boldsymbol{v})\dot{W}(t)$ $\quad \boldsymbol{v} \in \mathbb{R}^K$
  - Invariant measure $\mathcal{L}_{\text{FP}}\, p_{eq}(\boldsymbol{v}) = 0$
  - Expected value of A $\overline{A(\boldsymbol{u})} \equiv \int A(\boldsymbol{u}) p_{eq}(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}$ $\quad \boldsymbol{u} \in \mathbb{R}^N \subset \mathbb{R}^K$

- Perturbed system $\boldsymbol{f} \to \boldsymbol{f} + \delta \boldsymbol{f} \qquad \sigma \to \sigma + \delta\sigma$
  - Invariant measure $\mathcal{L}^{\delta}_{\text{FP}}\, p^{\delta}_{eq}(\boldsymbol{v}) = 0$
  - Expected value of A $\overline{A(\boldsymbol{u})}^{\delta}$

$\delta\overline{A(\boldsymbol{u})} = \overline{A(\boldsymbol{u})}^{\delta} - \overline{A(\boldsymbol{u})}$? Yes, if $p^{\delta}_{eq}(\boldsymbol{v})$ differentiable at $\delta = 0$

Formal generalizations to dissipative systems (Hairer & Majda 2010) with time-periodic attractors (Majda & Wang 2010, Gershgorin & Majda 2009)

Essentials of FDT

(e.g., Majda, Abramov, Grote 2005)

- Expected change of $\overline{\boldsymbol{A}(\boldsymbol{u})}$ on a subset $\boldsymbol{u} \in \mathbb{R}^N \subset \mathbb{R}^K$

$$\delta\overline{A(\boldsymbol{u})} = \overline{A(\boldsymbol{u})}^{\delta} - \overline{A(\boldsymbol{u})} = \int_{t_0}^{t} \mathcal{R}(t-s)\delta f(s)\mathrm{d}s$$

$$\mathcal{R}(\tau) = \overline{\boldsymbol{A}(\boldsymbol{u}(\tau))B(\boldsymbol{v}(0))}$$

$$B(\boldsymbol{v}(\tau)) = -\frac{\mathrm{div}(\boldsymbol{h}p_{eq})}{p_{eq}}$$

FDT

$\mathcal{R}(\tau)$ can be computed through a correlation function in the unperturbed attractor

Approximate FDT algorithms

$$\delta\overline{A(\boldsymbol{u})} = \overline{A(\boldsymbol{u})}^{\delta} - \overline{A(\boldsymbol{u})} = \int_{t_0}^{t} \mathcal{R}(t-s)\delta f(s)\mathrm{d}s$$

$$\mathcal{R}(\tau) = \overline{\boldsymbol{A}(\boldsymbol{u}(\tau))B(\boldsymbol{v}(0))}$$

$$B(\boldsymbol{v}(\tau)) = -\frac{\mathrm{div}(\boldsymbol{h}p_{eq})}{p_{eq}}$$

- Quasi-Gaussian FDT

$$\mathcal{R}_A^G(\tau) = \overline{\boldsymbol{A}(\boldsymbol{u}(\tau))B^G(\boldsymbol{u}(0))}$$

$$B^G(\boldsymbol{v}) = -\frac{\mathrm{div}(\boldsymbol{h}p_{eq}^G)}{p_{eq}^G}$$

- Kicked response

$$\mathcal{R}_A(t)\cdot\delta\boldsymbol{x}^0 = \overline{A(\boldsymbol{u}^{\delta\boldsymbol{x}^o})} - \overline{A(\boldsymbol{u})}$$

- Blended response FDT Abramov & Majda, Nonlinearity 2007

Practical algorithms for computing the linear response via FDT

- **Kicked response FDT**

Perturb the initial data for the perfect/imperfect models in the direction $\delta\boldsymbol{x}^0$ in a statistical fashion generating solutions of the unperturbed perfect and imperfect models with perturbed initial conditions

$$\partial_t p = \mathcal{L}_{FP}\, p \qquad p\Big|_{t=t_0} = p_{eq}(\boldsymbol{v} + \delta\boldsymbol{x}^0)$$

$$\mathcal{L}_{FP}\, p_{eq} = 0$$

$$\dot{\boldsymbol{v}}^\delta = \boldsymbol{f}(\boldsymbol{v}^\delta) + \delta\boldsymbol{x}^0\tilde{\delta}(t) + \sigma(\boldsymbol{v}^\delta)\dot{W}(t)$$

Derive the linear response by monitoring relaxation from the "kick" $\delta\boldsymbol{f} = \delta\boldsymbol{x}^0\,\tilde{\delta}(t)$

$$\mathcal{R}_A(t)\cdot\delta\boldsymbol{x}^0 = \int_{t_0}^{t}\mathcal{R}_A(t-s)\cdot\delta\boldsymbol{x}^0\tilde{\delta}(s)\mathrm{d}s = \overline{A(\boldsymbol{u}^\delta)} - \overline{A(\boldsymbol{u})}$$

FDT & time-periodic attractors

$$\frac{\partial p_{att}}{\partial s} + \nabla_{\boldsymbol{v}}[p_{att}\,\boldsymbol{f}(\boldsymbol{v},s)] - \frac{1}{2}\nabla_{\boldsymbol{v}}\cdot\nabla_{\boldsymbol{v}}[\sigma\sigma^T p_{att}] = 0.$$

$$\tilde{\langle} A \tilde{\rangle} = \frac{1}{T_0}\int_0^{T_0}\int_{\mathbb{R}^P} A(\boldsymbol{v},s)p_{att}(\boldsymbol{v},s)\mathrm{d}\boldsymbol{v}\mathrm{d}s. \qquad \frac{1}{T_0}\int_0^{T_0}\int_{\mathbb{R}^P} p_{att}(\boldsymbol{v},s)\mathrm{d}\boldsymbol{v}\mathrm{d}s = 1$$

$$\begin{aligned}\pi^M(\boldsymbol{u},s,t) &= \pi^M_{att}(\boldsymbol{u},s) + \delta\pi^M(\boldsymbol{u},s,t),\\ \pi(\boldsymbol{u},s,t) &= p_{att}(\boldsymbol{u},s) + \delta\pi(\boldsymbol{u},s,t).\end{aligned}$$

$$\delta\,\tilde{\langle} A \tilde{\rangle}\,(t) = \int_0^t R_A(t-s)\delta f(s)\mathrm{d}s,$$

$$R(t) = \tilde{\langle} A(\boldsymbol{v}(t+s), t+s) \otimes B(\boldsymbol{v}(s), s)\tilde{\rangle}.$$

Essence of fluctuation-dissipation theorem for forced dissipative systems

$$\frac{\partial}{\partial t} p_t = \mathcal{L}_{\text{FP}}\, p_t, \qquad p_t(\mathbf{v})|_{t=0} = p_0(\mathbf{v}) \qquad\qquad d\mathbf{v} = \boldsymbol{F}(\mathbf{v})dt + \sigma(\mathbf{v})d\mathbf{W}(t)$$

$$\mathcal{L}_{\text{FP}} = -\nabla\cdot\left[\boldsymbol{F}(\mathbf{v})\,\cdot\,\right] + \tfrac{1}{2}\nabla\cdot\nabla\left[Q(\mathbf{v})\,\cdot\,\right] \qquad\qquad Q = \sigma \otimes \sigma^T$$

Marginal density: $\pi_t(\boldsymbol{u}) = \int p_t(\boldsymbol{u}, \boldsymbol{v})d\boldsymbol{v}$

Perturbation:

$$\pi_t^\delta = \pi_{eq} + \delta\tilde{\pi}_t$$

$$\boldsymbol{F}_t^\delta = \boldsymbol{F}_t^\delta + \delta\tilde{\boldsymbol{F}}$$

$$\delta\tilde{\boldsymbol{F}}(\mathbf{v}, t) = \delta\widehat{\boldsymbol{F}}(\mathbf{v})f(t)$$

Changes in the statistical moments ...

$$\overline{\boldsymbol{E}}_t^\delta = \overline{\boldsymbol{E}}_0 + \delta\tilde{\overline{\boldsymbol{E}}}_t, \qquad \overline{\boldsymbol{E}}_t^{\text{M}\delta} = \overline{\boldsymbol{E}}_0^{\text{M}} + \delta\tilde{\overline{\boldsymbol{E}}}_t^{\text{M}}, \qquad \tilde{\overline{\boldsymbol{E}}}_0 = \tilde{\overline{\boldsymbol{E}}}_0^{\text{M}} = 0.$$

$$\delta\overline{E}_t \equiv \mathbb{E}^{\pi_t^\delta}\left[E(\boldsymbol{u})\right] - \mathbb{E}^{\pi_{eq}}\left[E(\boldsymbol{u})\right] = \int E(\boldsymbol{u})\tilde{\pi}_t(\boldsymbol{u})d\boldsymbol{u}$$

Essence of fluctuation-dissipation theorem for forced dissipative systems

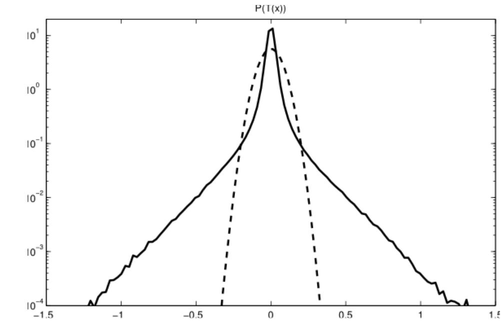
$$\delta\overline{E}_t \equiv \mathbb{E}^{\pi_t^\delta}\Big[E(\boldsymbol{u})\Big] - \mathbb{E}^{\pi_{eq}}\Big[E(\boldsymbol{u})\Big] = \int E(\boldsymbol{u})\tilde{\pi}_t(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$$

Changes in the statistical moments can be computed via appropriate correlation functions in the unperturbed equilibrium

$$\delta\overline{E}_t = \int_0^t \mathrm{d}t' f(t') \int \mathrm{d}\boldsymbol{u} \int \mathrm{d}\boldsymbol{v}\, E(\boldsymbol{u}) e^{(t-t')\mathcal{L}_{\text{FP}}} \mathcal{L}_\delta\, p_{eq}(\mathbf{v}) = \int_0^t \mathrm{d}t'\, \mathcal{R}_E(t-t') f(t')$$

$$\mathcal{R}_E(t) = \mathbb{E}^{p_{eq}}\Big[E(\boldsymbol{u}(t+\tau))B(\mathbf{v}(\tau))\Big] \qquad B(\mathbf{v}) = \frac{\mathcal{L}_\delta\, p_{eq}(\mathbf{v})}{p_{eq}(\mathbf{v})}$$

Poor sensitivity for Non-Gaussian tracer via qG-FDT



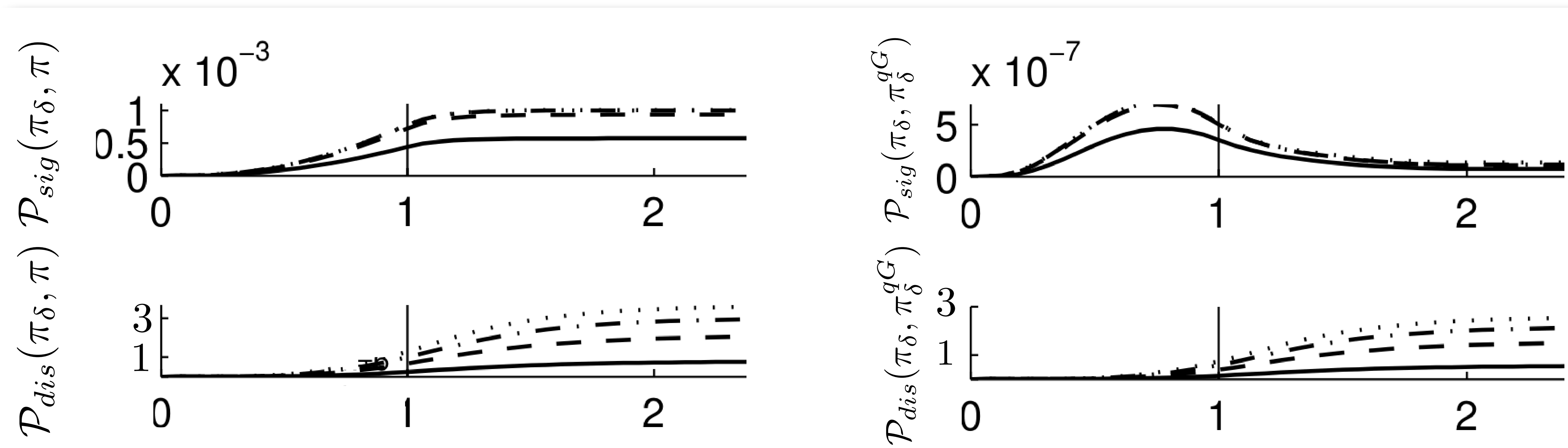
non-Gaussian tracer

- Quasi-Gaussian FDT

Expected response of a functional *A* to forcing perturbation

$$\delta\overline{A(\boldsymbol{u})} = \int_{t_0}^{t} \mathcal{R}_A^G(t-s)\delta f(s)\mathrm{d}s$$

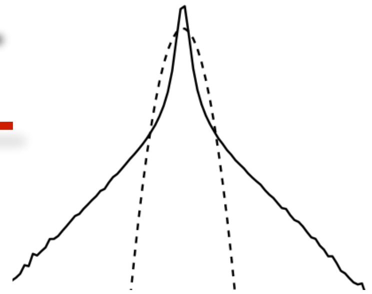
$$\mathcal{R}_A^G(\tau) = \overline{\boldsymbol{A}(\boldsymbol{u}(\tau))B^G(\boldsymbol{u}(0))} \qquad B^G(\boldsymbol{u}) = -\frac{\mathrm{div}(\boldsymbol{h}p_{eq}^G)}{p_{eq}^G}$$



- Good skill from qG-FDT for the mean
- No skill from qG-FDT for the variance

Majda & Gershgorin, PNAS 2011

High prediction skill for the tracer statistics via kicked FDT

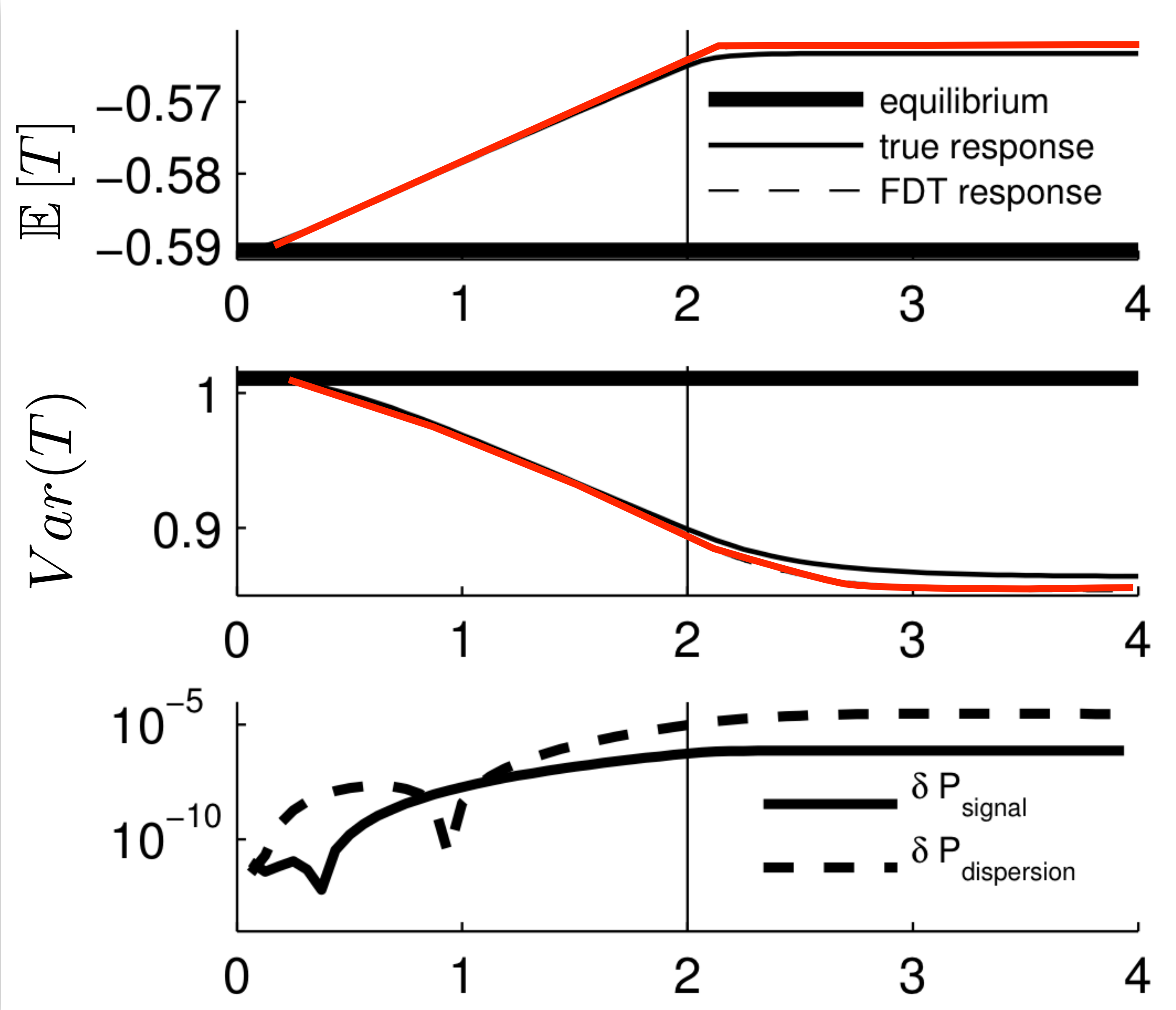


non-Gaussian tracer

- Kicked-response FDT

$$\overline{\delta A(\boldsymbol{u})} = \int_{t_0}^{t} \mathcal{R}_A^{kck}(t-s)\delta f(s)\mathrm{d}s$$

$\mathcal{R}_A^{kck}$ estimated from monitoring the system relaxation to equilibrium after a kick



- High predictive skill from kicked-FDT for the mean & variance

Majda & Gershgorin, PNAS 2011