SAND2020-2623C

Characterizing model-form uncertainty in an inadequate model of anomalous transport

Teresa Portone, Robert D. Moser

Center for Predictive Engineering and Computational Sciences

March 24, 2020

a National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly ubsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA00035

Motivation/Issues

- All models are approximations of reality.
  - Don't fully understand the modeled phenomena.
  - Can't observe/resolve all relevant aspects of the phenomena.
- Have to use them for prediction anyway, so need to understand their reliability.
- Need model-form uncertainty representations that are physics-informed and predictive.
- Our research group (PECOS) has focused on how to develop such representations.

Target source of uncertainty: multiscale models without sufficient scale separation.

- Macroscopic quantities of interest depend on dynamics at smaller scales.
- Common problem: smaller scales can't be observed or resolved.
- Effect: macroscopic model's dependence on the small scales is significant but uncertain.
- Common to model away the dependence, but can cause errors.

Goal: Develop an uncertainty representation to account for missing dependence on small scales in the context of contaminant transport.

Testbed problem: contaminant transport through heterogeneous porous media.

Isolate model-form uncertainty using a hierarchy of models.

- High-fidelity model that resolves relevant physics, low-fidelity model that does not.
- Discrepancies between the models arises from this missing information.
- Use high-fidelity model to generate data and probe the physics of the problem.

High-fidelity model for field-scale contaminant transport

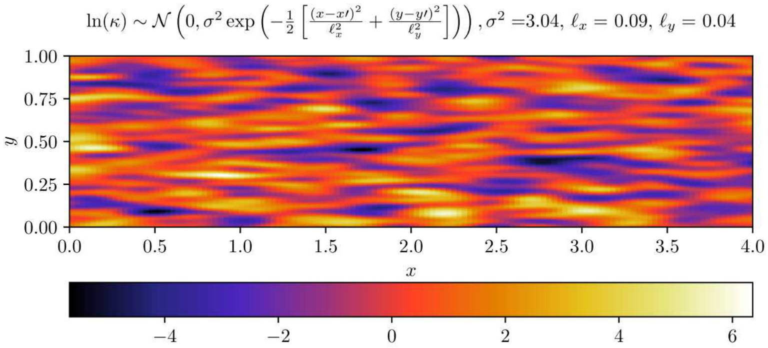
$$\frac{\partial c}{\partial t} + \nabla \cdot (\mathbf{u}c) = \nu_p \Delta c, \quad (x, y) \in [0, L_x] \times [0, L_y]$$
$$\mathbf{u} = -\kappa \nabla p$$
$$\nabla \cdot (\mathbf{u}) = 0$$

Periodic in x, zero Neumann in y

Problem:

- Don't know $\kappa(x, y)$.

Permeability fields are highly heterogeneous, observed to vary over several orders of magnitude.



However, their statistics are homogeneous.

What do we have access to for the low-fidelity model?

- Statistics of κ.
- y-averaged observations of c.

Try to predict average behavior instead.

$$\langle f(x,y)\rangle \equiv \frac{1}{L_y}\int_0^{L_y} \mathbb{E}_\kappa\left[f(x,y)\right] \mathrm{d}y \tag{1}$$

$$f(x,y) = \langle f(x,y)\rangle + f'(x,y) \tag{2}$$

Represent $c, \mathbf{u}$ using (2), apply (1) to the 2D ADE to get

$$\frac{\partial \langle c\rangle}{\partial t} + \langle u\rangle \frac{\partial \langle c\rangle}{\partial x} + \frac{\partial \langle u'c'\rangle}{\partial x} = \nu_p \frac{\partial^2 \langle c\rangle}{\partial x^2}.$$

Can't observe $\langle u'c'\rangle \implies$ dependence uncertain.

Typical closure model for $\langle u'c' \rangle$ is gradient-diffusion:

$$\langle u'c' \rangle \approx -\nu_m \frac{\partial \langle c \rangle}{\partial x}.$$

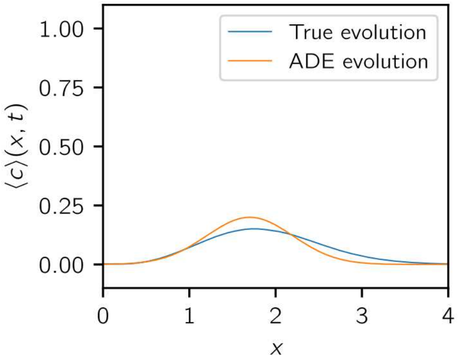
$$\Downarrow$$

$$\frac{\partial \langle c \rangle}{\partial t} + \langle u \rangle \frac{\partial \langle c \rangle}{\partial x} = \nu \frac{\partial^2 \langle c \rangle}{\partial x^2}, \quad \nu = \nu_p + \nu_m,$$

$$\langle c \rangle (0, t) = \langle c \rangle (L_x, t),$$

$$\langle c \rangle (x, 0) = c_0(x).$$

Transport through heterogeneous porous media induces anomalous diffusion in $\langle c \rangle$.



The ADE for $\langle c \rangle$ can dangerously underpredict levels of contaminant downstream.

Goal: characterize the uncertainty in $\langle c \rangle$, given the lack of information about $\langle u'c' \rangle$.

For predictions, a model-form uncertainty representation should (Oliver et al. 2015):

- Perturb the dynamics of the model.
- Accurately extrapolate to prediction scenarios of interest.
- Represent irreducible model uncertainty.

To do this it must:

- Be embedded at the source of the uncertainty.
- Act on the state variable(s).
- Respect physical constraints.
- Be scenario-dependent.
- Be stochastic.

Requirement: state-dependence.

- Represent ϵ_{model} as an operator acting on $\langle c \rangle$.

Requirement: embedded, scenario dependent.

$$\epsilon_{model}\left(\langle c \rangle; \mathbf{s}\right) \equiv -\frac{\partial \langle u' c' \rangle}{\partial x},$$

s scenario parameters.

Model-form uncertainty representation

Including the uncertainty representation, the model for $\langle c \rangle$ is

$$\frac{\partial \langle c \rangle}{\partial t} + \langle u \rangle \frac{\partial \langle c \rangle}{\partial x} = \nu_p \frac{\partial^2 \langle c \rangle}{\partial x^2} + \epsilon_{model}(\langle c \rangle; \mathbf{s}).$$

Uncertainty representation development process:

- Constrain the ϵ_{model}'s structure to reflect prior information (e.g. physical constraints).
- Inspect and encode scenario dependence.
- Characterize remaining uncertainties using probability distributions.
- Update uncertainty representations using data.

$$\frac{\partial \langle c \rangle}{\partial t} + \langle u \rangle \frac{\partial \langle c \rangle}{\partial x} = \nu_p \frac{\partial^2 \langle c \rangle}{\partial x^2} + \epsilon_{model}(\langle c \rangle)$$

Physical constraints

Linearity in $\langle c \rangle$	$\Longrightarrow$	$\epsilon_{model} = \mathcal{L}, \quad \mathcal{L} f_k = \lambda_k f_k.$
Shift invariance	$\Longrightarrow$	$f_k = e^{i2\pi k/L_x x},\ \mathcal{L}\langle c \rangle = \sum_k \langle \hat{c}_k \rangle \lambda_k e^{i2\pi k/L_x x}.$
Conservation of mass	$\Longrightarrow$	$\lambda_0 = 0.$
Solution decays with time	$\Longrightarrow$	$-\nu_p \left(\frac{2\pi k}{L_x}\right)^2 + \Re[\lambda_k] \leq 0.$

- Advection-diffusion induces decay in solution, causing information loss.
- Can only inform the first $\sim$10 eigenvalues.
- Recast the problem to enable observation of eigenvalues directly.

For a given $\mathbf{u}$ define $\widetilde{\mathcal{L}}$ such that

$$\widetilde{\mathcal{L}} \langle c \rangle_y = -\frac{\partial \langle u'c' \rangle_y}{\partial x}.$$

Connection to mean:

$$\mathbb{E}\left[\widetilde{\mathcal{L}} \langle c \rangle_y\right] = \mathbb{E}\left[-\frac{\partial \langle u'c' \rangle_y}{\partial x}\right] = -\frac{\partial \langle u'c' \rangle}{\partial x}.$$

Let $\widetilde{\boldsymbol{\lambda}} = \left[\widetilde{\lambda}_1, \widetilde{\lambda}_2, \dots\right]$ be $\widetilde{\mathcal{L}}$'s eigenvalues. Then

$$\widetilde{\mathcal{L}} \langle c \rangle_y = -\frac{\partial \langle u'c' \rangle_y}{\partial x} \implies \widetilde{\lambda}_k \langle \hat{c}_k \rangle_y = -\left(i^{2\pi k}/L_x\right) \left\langle \widehat{(u'c')}_k \right\rangle_y .$$

- $\langle u'c' \rangle_y$ random $\implies \widetilde{\boldsymbol{\lambda}}$ random.
- If $p(\widetilde{\boldsymbol{\lambda}})$ were known, could compute mean effect of $\langle u'c' \rangle$ exactly.
- Developed a method to compute $\widetilde{\boldsymbol{\lambda}}$ directly instead of infer from observations of $\langle c \rangle$.
- Given an ensemble of $\mathbf{u}$, computed corresponding ensemble of $\widetilde{\boldsymbol{\lambda}}$.
- Can study the statistics of the ensemble to learn about $\widetilde{\mathcal{L}}$.

Computing samples of $\widetilde{\lambda}_k$

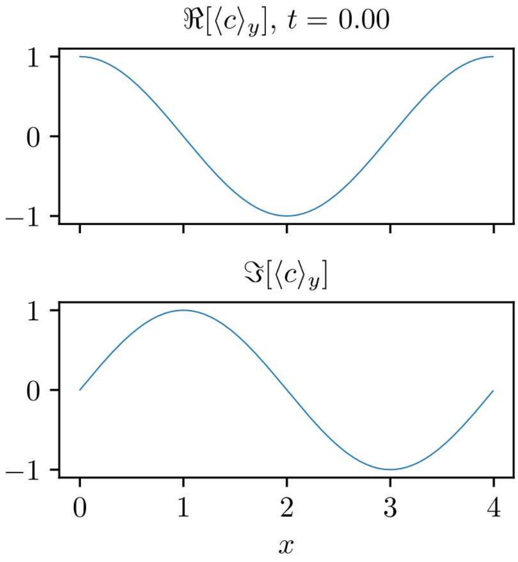
Given $\mathbf{u}$, can determine $\widetilde{\lambda}_k(t)$ using the 2D ADE.

$$\frac{\partial c}{\partial t} + \nabla \cdot (\mathbf{u}c) = \nu_p \Delta c + f_k,$$

$$f_k = \alpha(t) \langle \hat{c}_k \rangle_y \delta(k),$$

$$c(x, y, 0) = \exp\left(i \frac{2\pi k}{L_x} x\right),$$

f_k defined s.t. $| \langle \hat{c}_k \rangle_y | = 1 \, \forall t$.



$$\frac{\partial \langle \hat{c}_k \rangle_y}{\partial t} + i \langle u \rangle \left(\frac{2\pi k}{L_x} \right) \langle \hat{c}_k \rangle_y = -\nu_p \left(\frac{2\pi k}{L_x} \right)^2 \langle \hat{c}_k \rangle_y + \widetilde{\lambda}_k \langle \hat{c}_k \rangle_y + \underbrace{\alpha \langle \hat{c}_k \rangle_y}_{\hat{f}_k}$$

Forcing $\implies \langle \hat{c}_k \rangle_y (t) = e^{i\theta_k(t)}$, so

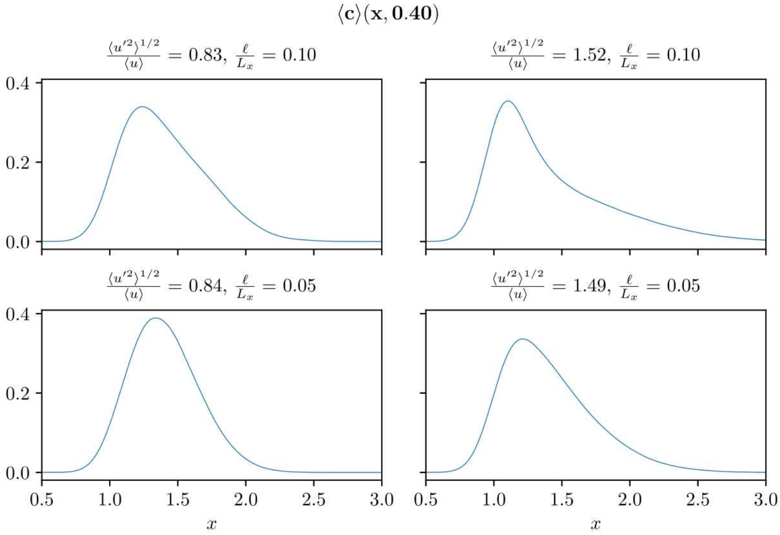
$$e^{i\theta_k(t)} \left[i \frac{\partial \theta_k}{\partial t} + i \langle u \rangle \left(\frac{2\pi k}{L_x} \right) \right] = e^{i\theta_k(t)} \left[-\nu_p \left(\frac{2\pi k}{L_x} \right)^2 + \widetilde{\lambda}_k + \alpha \right]$$

$$\boxed{\begin{aligned} \Re\left[\widetilde{\lambda}_k\right] &= -\alpha - \left(-\nu_p \left(\frac{2\pi k}{L_x} \right)^2 \right) \\ \Im\left[\widetilde{\lambda}_k\right] &= \frac{\partial \theta_k}{\partial t} + \langle u \rangle \left(\frac{2\pi k}{L_x} \right) \end{aligned}}$$

- $p(\widetilde{\boldsymbol{\lambda}})$ depends on stats of $\langle u'c' \rangle$, which aren't available for practical problems.
- Instead studied how $p(\widetilde{\boldsymbol{\lambda}})$ depends on proxy variables that are known *a priori*.
  - $\langle u \rangle$
  - $\langle u'^2 \rangle$
  - $\ell \equiv \int_0^{L_x} \frac{\langle u' u'(x') \rangle}{\langle u'^2 \rangle} \mathrm{d}x'$ integrated autocorrelation length
- Defined nondimensional scenario parameters $\langle u'^2 \rangle^{1/2} / \langle u \rangle$ and ℓ / L_x.
- Computed ensembles of $\mathbf{u}$, $\widetilde{\boldsymbol{\lambda}}$, $\langle c \rangle_y$ over a coarse grid on the 2D scenario space.
- Studied how summary statistics for each ensemble depended on scenario parameters.

First, computed evolution of Gaussian pulse for each scenario.

Anomalous diffusion increases with $\langle u'^2\rangle^{1/2}/\langle u\rangle$ and ℓ/L_x.

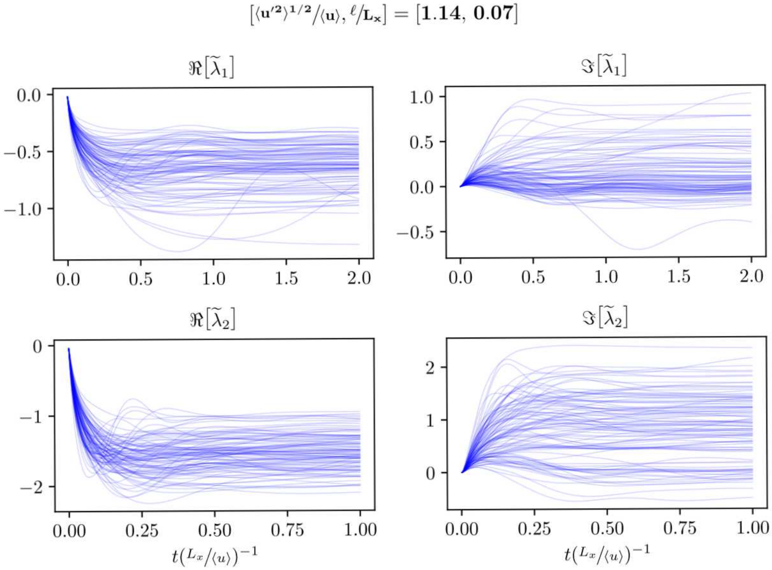


$\widetilde{\lambda}_k$ rapidly become stationary (within one flowthrough time $T_f = {}^{L_x}/\langle u \rangle$).

$\Downarrow$

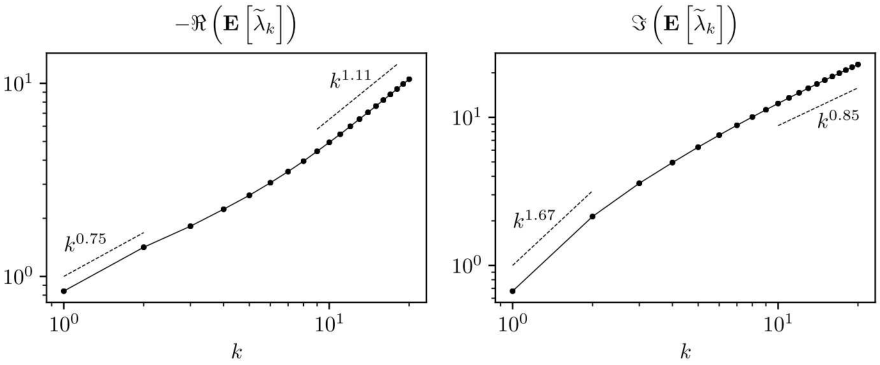
Study stationary values of $\widetilde{\boldsymbol{\lambda}}$.

$\widetilde{\boldsymbol{\lambda}}(t_{final}) \equiv \widetilde{\boldsymbol{\lambda}}$ for remainder of analysis.

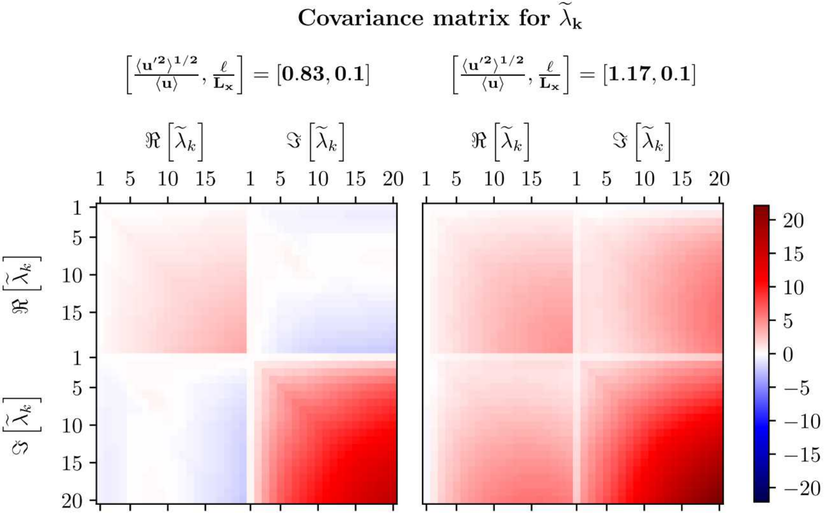


Mean $\Re[\widetilde{\lambda}_k]$ and $\Im[\widetilde{\lambda}_k]$ do not depend on a fixed power of k.

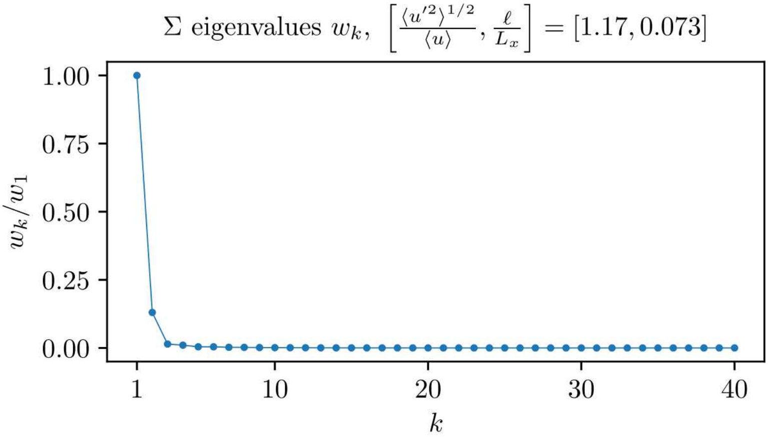
$$\frac{\langle \mathbf{u'^2} \rangle^{1/2}}{\langle \mathbf{u} \rangle} = \mathbf{1.17}, \; \frac{\ell}{\mathbf{L_x}} = \mathbf{0.073}$$



Covariance between $\Re[\widetilde{\lambda}_k]$ and $\Im[\widetilde{\lambda}_k]$, and as a function of k, is significant.



Covariance matrix of the eigenvalues (Σ) admits a low-rank approximation.



$$\mathbb{E}\left[\widetilde{\mathcal{L}}\,\langle c\rangle_y\right] \equiv -\frac{\partial\,\langle u'c'\rangle}{\partial x}$$

- Uncertainty in $\widetilde{\mathcal{L}}$ defined in terms of $p\big(\widetilde{\boldsymbol{\lambda}}\big)$.
- Unresolved dependence on statistics of $\langle u'c'\rangle$ approximated in terms of ${}^{\langle u'^2\rangle^{1/2}}/_{\langle u\rangle}$, ${}^{\ell}/_{L_x}$.
- Express irreducible uncertainty in $\widetilde{\mathcal{L}}$ by defining $p\big(\widetilde{\boldsymbol{\lambda}}\big)$ w.r.t. hyperparameters $\boldsymbol{\xi}$: $p\big(\widetilde{\boldsymbol{\lambda}};\boldsymbol{\xi}\big)$.

Modeling requirements for $p(\widetilde{\boldsymbol{\lambda}}; \boldsymbol{\xi})$

Prior knowledge

- Deterministic constraints: $\widetilde{\lambda}_0 = 0$, $\Re[\widetilde{\lambda}_k] \leq 0$
- Scenario-based constraints: $\langle u'^2 \rangle^{1/2} / \langle u \rangle$, $\ell / L_x \to 0 \implies \mathbb{E}(\widetilde{\boldsymbol{\lambda}}), \mathrm{Var}(\widetilde{\boldsymbol{\lambda}}) \to 0$.

$\widetilde{\boldsymbol{\lambda}}$ ensemble analysis

- $\widetilde{\boldsymbol{\lambda}}$ rapidly become stationary.
- $\mathbb{E}(\Re[\widetilde{\lambda}_k])$ and $\mathbb{E}(\Im[\widetilde{\lambda}_k])$ different functions of k.
- Covariance between $\Re[\widetilde{\lambda}_k]$ and $\Im[\widetilde{\lambda}_k]$ and as a function of k significant.
- Covariance matrix admits rank-2 approximation.

Prototype formulation of $p(\widetilde{\boldsymbol{\lambda}}; \boldsymbol{\xi})$

Assume $\widetilde{\boldsymbol{\lambda}}$ constant in time, i.e. $\widetilde{\boldsymbol{\lambda}} \equiv \widetilde{\boldsymbol{\lambda}}_{stationary}$.

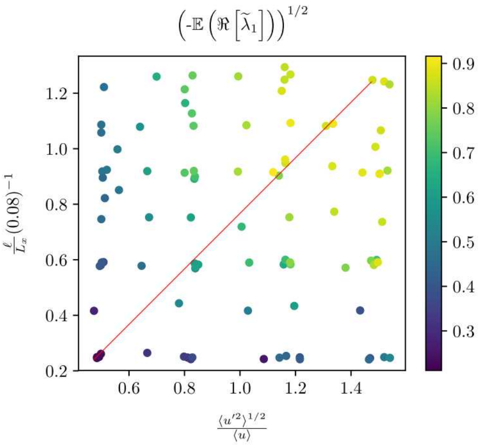
$$p(\widetilde{\boldsymbol{\lambda}}; \boldsymbol{\xi}) = \mathcal{N}\left(\mathbb{E}[\widetilde{\boldsymbol{\lambda}}]_{model}, \Sigma_{model}\right)$$

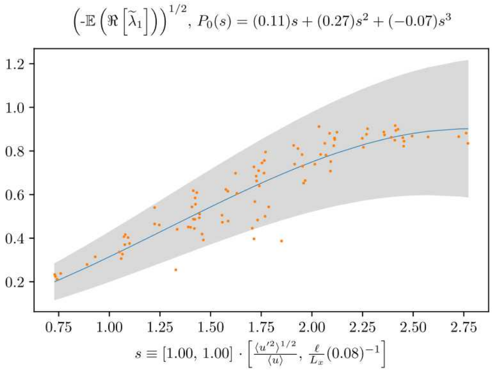
$$\mathbb{E}(\Re[\widetilde{\lambda}_k])_{model} = f_{\boldsymbol{\lambda_R}}(k), \qquad \Sigma_{model} = w_1 \mathbf{v_1}\mathbf{v_1}^T + w_2 \mathbf{v_2}\mathbf{v_2}^T,$$

$$\mathbb{E}(\Im[\widetilde{\lambda}_k])_{model} = f_{\boldsymbol{\lambda_I}}(k). \qquad (\mathbf{v_1})_k = f_{\mathbf{v}_1}(k), \quad (\mathbf{v_2})_k = f_{\mathbf{v}_2}(k).$$

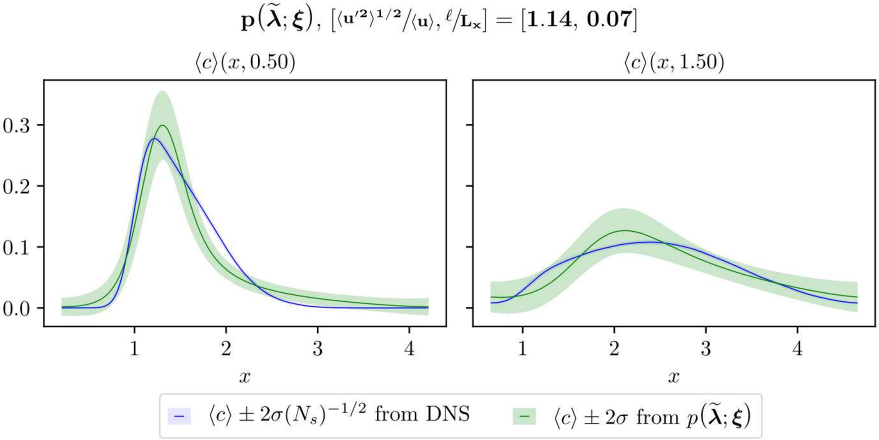
- Approximated $f_{\boldsymbol{\lambda_R}}, f_{\boldsymbol{\lambda_I}}, f_{\mathbf{v}_1}, f_{\mathbf{v}_2}$ as linear functions of k.
- Their slopes and intercepts and w_1, w_2 are the hyperparameters $\boldsymbol{\xi}$.

- $\boldsymbol{\xi}$ depend on $\langle u'^2\rangle^{1/2}/\langle u\rangle$, ℓ/L_x, but dependence is uncertain: $\xi_i \sim \mathcal{N}(m,\sigma)$.
- Computed ξ_i directly for each ensemble generated in the scenario study.
- Made polynomial fits to this data for m, σ as functions of $\langle u'^2\rangle^{1/2}/\langle u\rangle$ and ℓ/L_x.



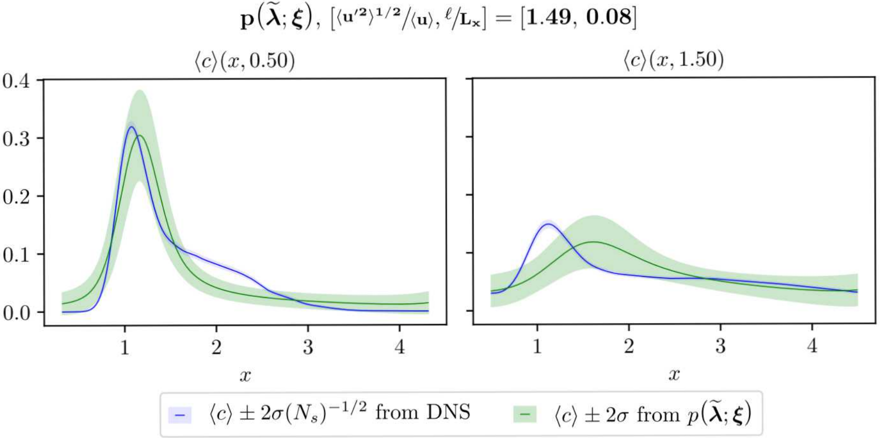


Inspected push-forward of $p(\widetilde{\boldsymbol{\lambda}};\boldsymbol{\xi})$ to mean evolution of a Gaussian across a range of scenarios.



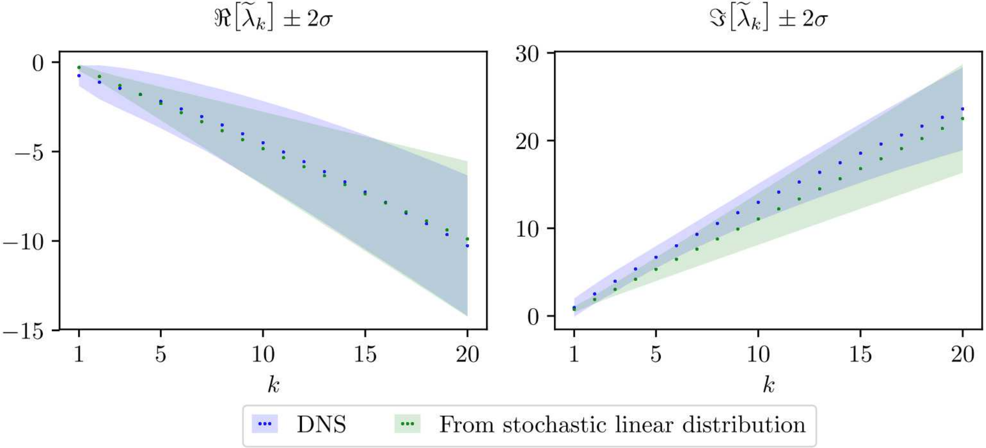
For moderately anomalous cases, the prototype formulation reproduces $\langle c\rangle$ remarkably well.

For an extremely anomalous case, it fails to reproduce important features of $\langle c \rangle$'s evolution.

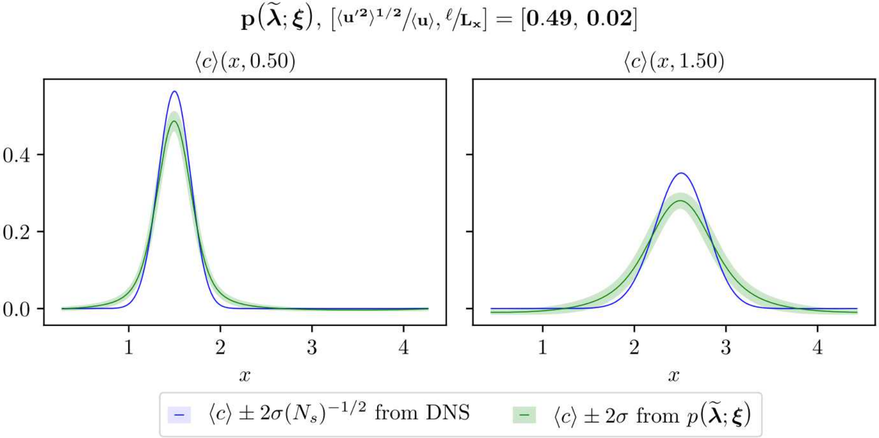


The location of the bulk ot the concentration is not captured at $t = 1.5$.

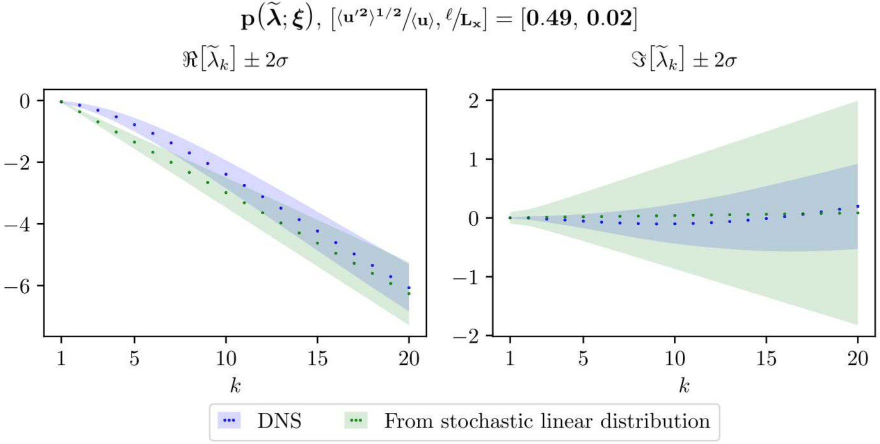
$$\mathbf{p}\left(\widetilde{\boldsymbol{\lambda}};\boldsymbol{\xi}\right),\ [\langle \mathbf{u'^2}\rangle^{1/2}/\langle \mathbf{u}\rangle,\ \ell/\mathbf{L_x}] = [\mathbf{1.49},\ \mathbf{0.08}]$$



For nonanomalous cases, it also overpredicts diffusion.



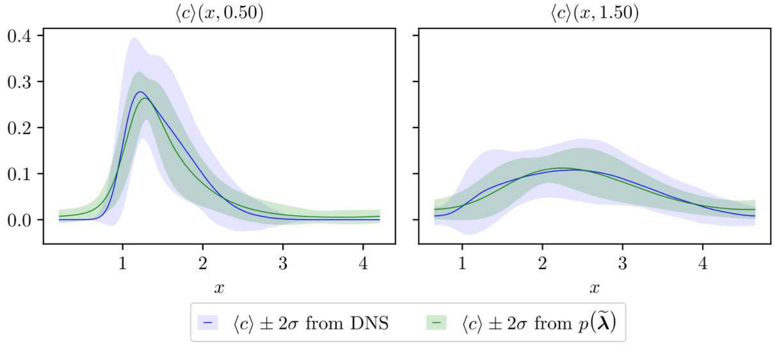
The linear model for $\mathbb{E}(\widetilde{\boldsymbol{\lambda}})$ is invalid for nonanomalous cases with nonlinear dependence on k.



The prototype isn't perfect, but we know why not:

- Assumed linear models for k dependence.
- Made minimal asymptotic arguments for scenario dependence.

Propagating $\widetilde{\boldsymbol{\lambda}}_{stationary}$ from the ensembles encapsulates the mean for all cases.



- Developed a novel method to directly probe the uncertain dependence in a model.
- Used the method to generate observations of the stochastic operator's eigenvalues.
  - Was able to learn about the operator's structure while avoiding an ill-posed inverse problem.
- Used the observations of the eigenvalues to formulate a data-informed representation of their distribution.

"Computational spectroscopy for statistically-invariant systems." Teresa Portone, Robert D. Moser. In preparation.

Contact: tnportone@utexas.edu, tporton@sandia.gov[1]

[1] Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.