ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

May 2020

CHARACTERIZING MODEL-FORM UNCERTAINTY IN AN INADEQUATE MODEL OF ANOMALOUS TRANSPORT

Teresa Portone, Robert D. Moser
Center for Predictive Engineering and Computational Sciences, The University of Texas at Austin

- **Target problem:** multiscale models without sufficient scale separation
  - Macroscopic quantities depend on unobservable dynamics at smaller scales.
  - Macroscopic model's dependence on small scales is significant but uncertain.
  - This dependence is commonly modeled away, but this can cause errors.

- **Application:** homogenized model of contaminant transport through heterogeneous porous media.

We can isolate model-form uncertainty using a hierarchy of models:

- High-fidelity model resolving all relevant physics taken as “truth”, low-fidelity model that does not.
- Discrepancies between models arises from the missing physics.
- Can use high-fidelity model to generate data and probe the physics of the problem.

High-fidelity model for field-scale contaminant transport:

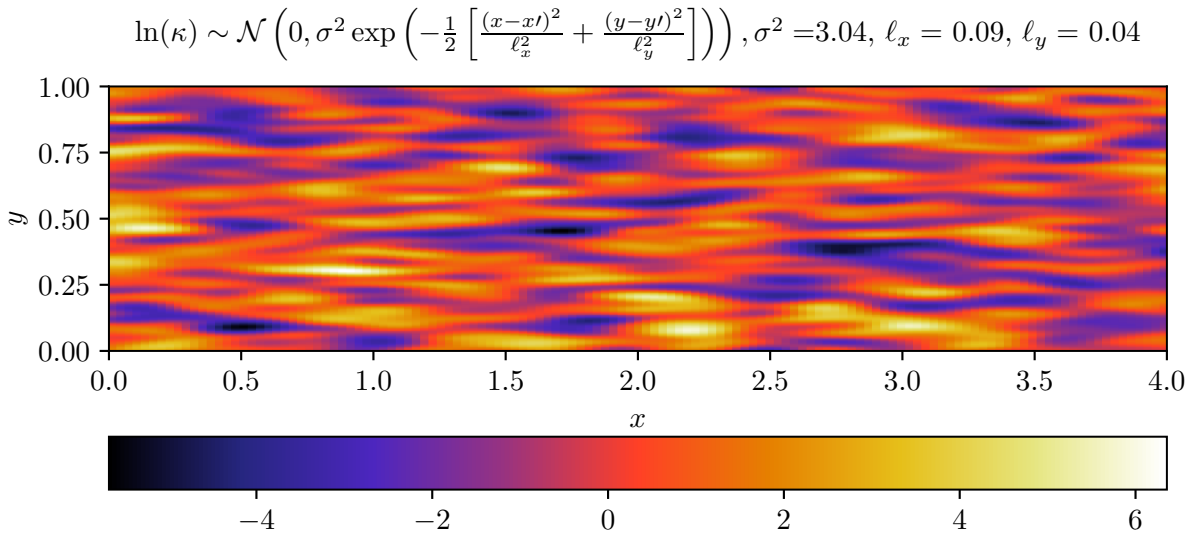
$$\frac{\partial c}{\partial t} + \nabla \cdot (\mathbf{u}c) = \nu_p \Delta c$$

$$\mathbf{u} = -\kappa \nabla p, \quad (x, y) \in \left[0, L_x\right] \times \left[0, L_y\right]$$

$$\nabla \cdot \mathbf{u} = 0$$

Periodic in x, zero Neumann in y

In heterogeneous porous media, permeability fields vary by orders of magnitude across the domain.



Here we will assume their statistics are homogeneous.

For low-fidelity model, have access to statistics of κ and observations of $\langle c \rangle_y$.

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

$$f(x,y) = \langle f(x,y) \rangle + f'(x,y)$$

Apply this decomposition to $\mathbf{u}$, c and average the high-fidelity model to get

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

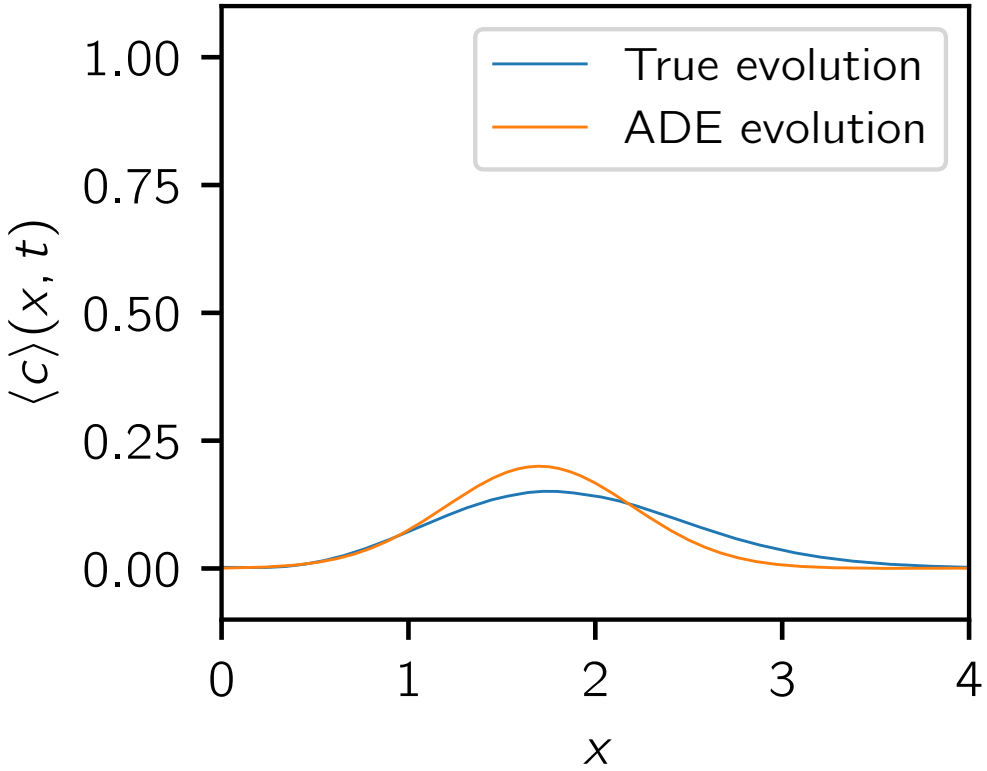
$\langle u'c' \rangle$ depends on unobservable fluctuations at the small scale. Dependence uncertain.

Typical closure model for $\langle u'c' \rangle$ is gradient diffusion: $\langle u'c' \rangle \approx -\nu_m \frac{\partial \langle c \rangle}{\partial x}$

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

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

ADE can dangerously underpredict contaminant levels downstream.



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

- Perturb the dynamics of the model
- Accurately extrapolate to prediction scenarios
- 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

Embedded, state and scenario dependent: represent model-form uncertainty ϵ_{model} as an operator acting on $\langle c \rangle$:

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

Model for $\langle c \rangle$ with uncertainty is

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

Next steps:

- Enforce physical constraints on ϵ_{model}
- Inspect and encode scenario dependence
- Represent remaining uncertainty with probability distributions
- Update uncertainty representations using data

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

Physical constraints:

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

- Want to understand how λ_k change with scenario, but diffusion in the system causes information loss.
- Could only infer the first ~10 eigenvalues of $\mathscr{L}$ from observations of $\langle c \rangle$.
- Recast the problem to observe the eigenvalues directly.

Given a single $\mathbf{u}$ from an ensemble, define $\widetilde{\mathscr{L}}$ such that

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

Note this is only a y average!

Connection to mean:

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

Let $\widetilde{\boldsymbol{\lambda}} = \left[\widetilde{\lambda}_1, \widetilde{\lambda}_2, \ldots\right]$ be the eigenvalues of $\widetilde{\mathscr{L}}$. Then

$$\widetilde{\mathscr{L}}\langle c\rangle_y = -\frac{\partial\langle u'c'\rangle_y}{\partial x} \quad \Longleftrightarrow \quad \widetilde{\lambda}_k\langle\hat{c}_k\rangle_y = -\left(\frac{i2\pi k}{L_x}\left\langle\widehat{(u'c')_k}\right\rangle_y\right)$$

- $\langle u'c'\rangle_y$ random $\Longrightarrow$ $\widetilde{\boldsymbol{\lambda}}$ random
- If we knew $p(\widetilde{\boldsymbol{\lambda}})$, we could compute mean effect of $\langle u'c'\rangle$ by taking expectation
- Developed a method to compute ensemble of $\widetilde{\boldsymbol{\lambda}}$ from ensemble of $\mathbf{u}$
- Can study summary statistics of ensemble to characterize $p(\widetilde{\boldsymbol{\lambda}})$

Scenario dependence

- Want to study how statistics of $\widetilde{\boldsymbol{\lambda}}$ depend on observable properties of $\mathbf{u}$, e.g.
  - $\langle u \rangle$
  - $\langle u'^2 \rangle$
  - ℓ, the integrated autocorrelation length in the streamwise direction.

- Used these to define nondimensional scenario parameters $\dfrac{\langle u'^2 \rangle^{1/2}}{\langle u \rangle}$ and $\dfrac{\ell}{L_x}$.
- Computed ensembles of $\mathbf{u}, \widetilde{\boldsymbol{\lambda}}, \langle c \rangle_y$ on a coarse grid in the 2D scenario space.
- Studied how the summary statistics for each ensemble changed as a function of scenario.

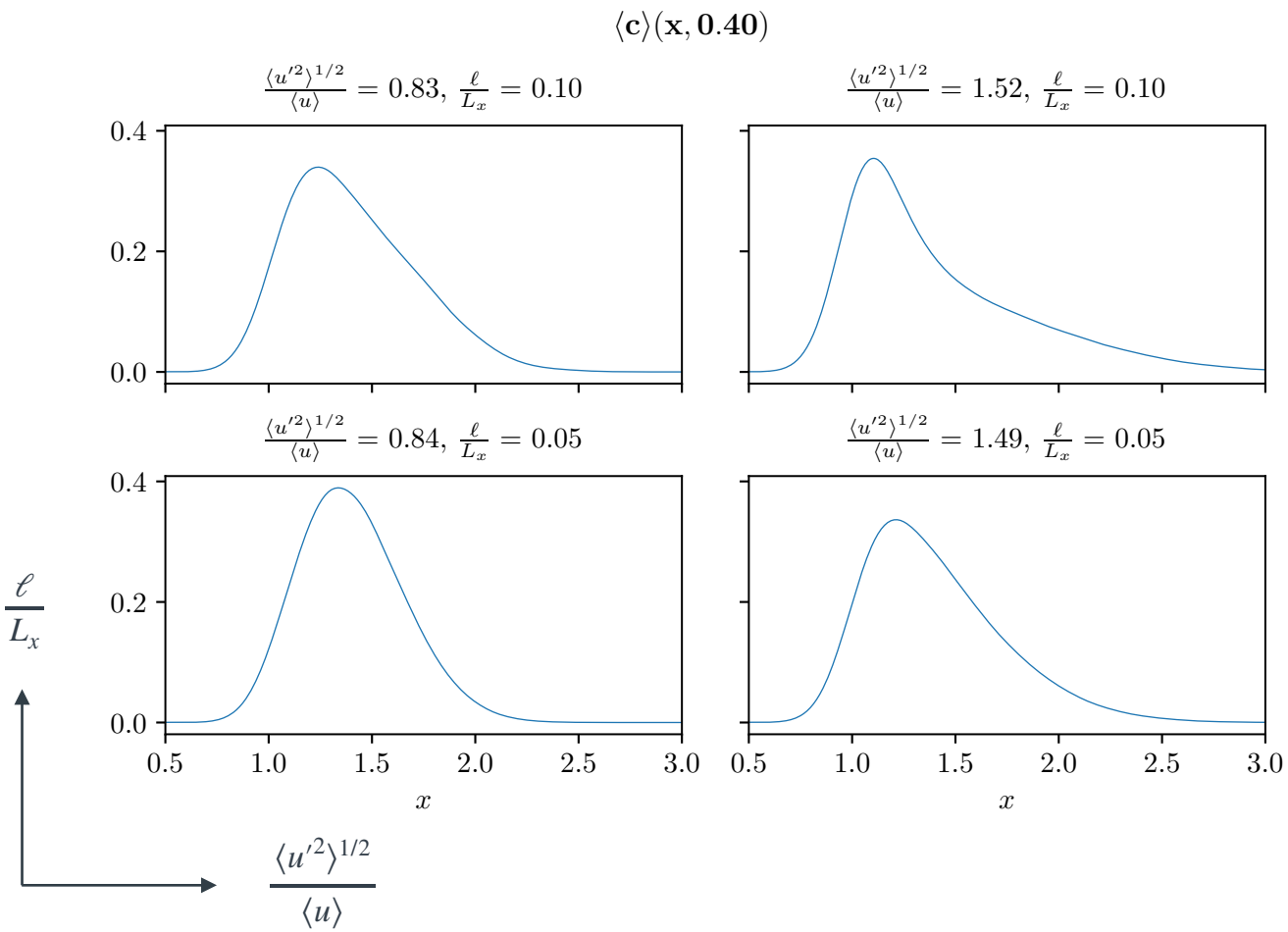
ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

Predictive Engineering & Computational Science

Anomalous diffusion scenario dependence

Evolved a Gaussian IC for each scenario.

Anomalous diffusion increases with both scenario parameters.



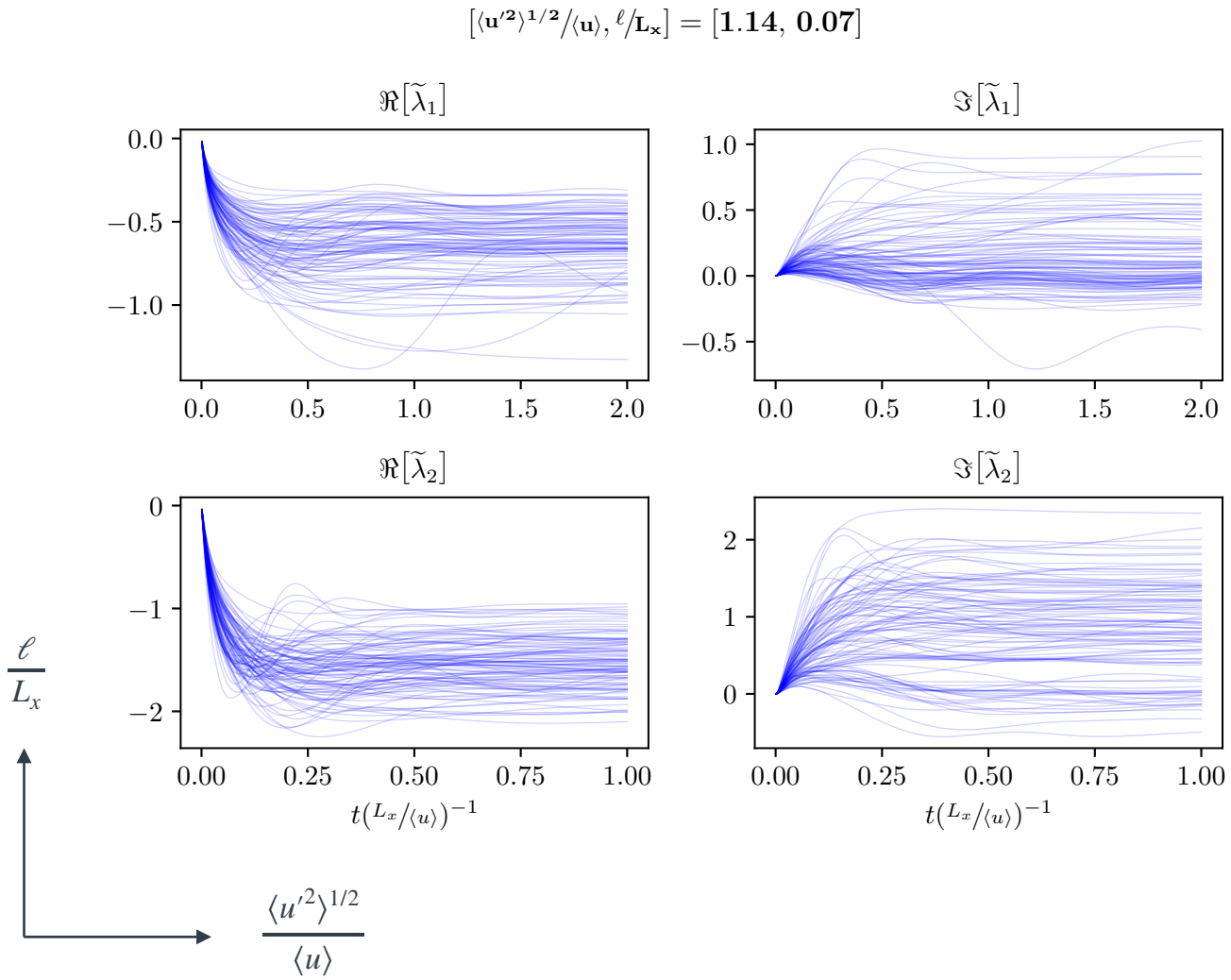
ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

Predictive Engineering & Computational Science

$\widetilde{\lambda}_k$ rapidly become stationary—within one flowthrough time $T_f = \dfrac{L_x}{\langle u \rangle}$.

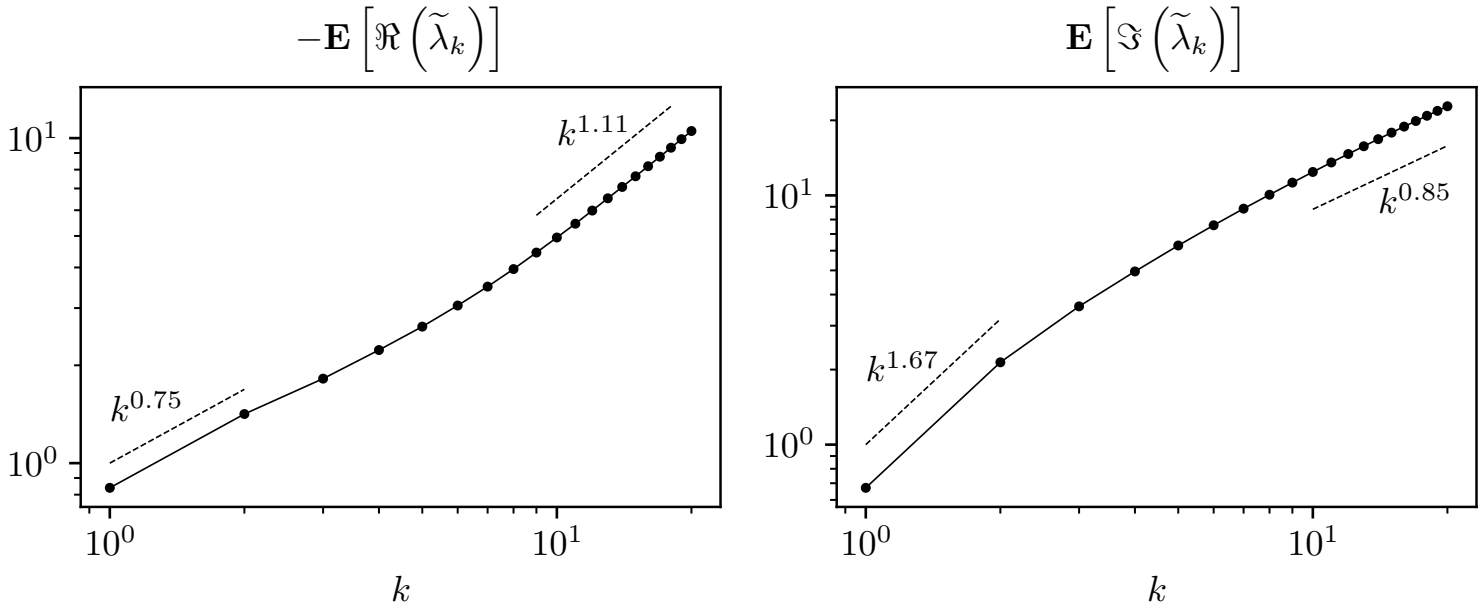
Decided to focus on stationary values of $\widetilde{\boldsymbol{\lambda}}$.

For remainder of talk $\widetilde{\boldsymbol{\lambda}} \equiv \widetilde{\boldsymbol{\lambda}}(t_{final})$.

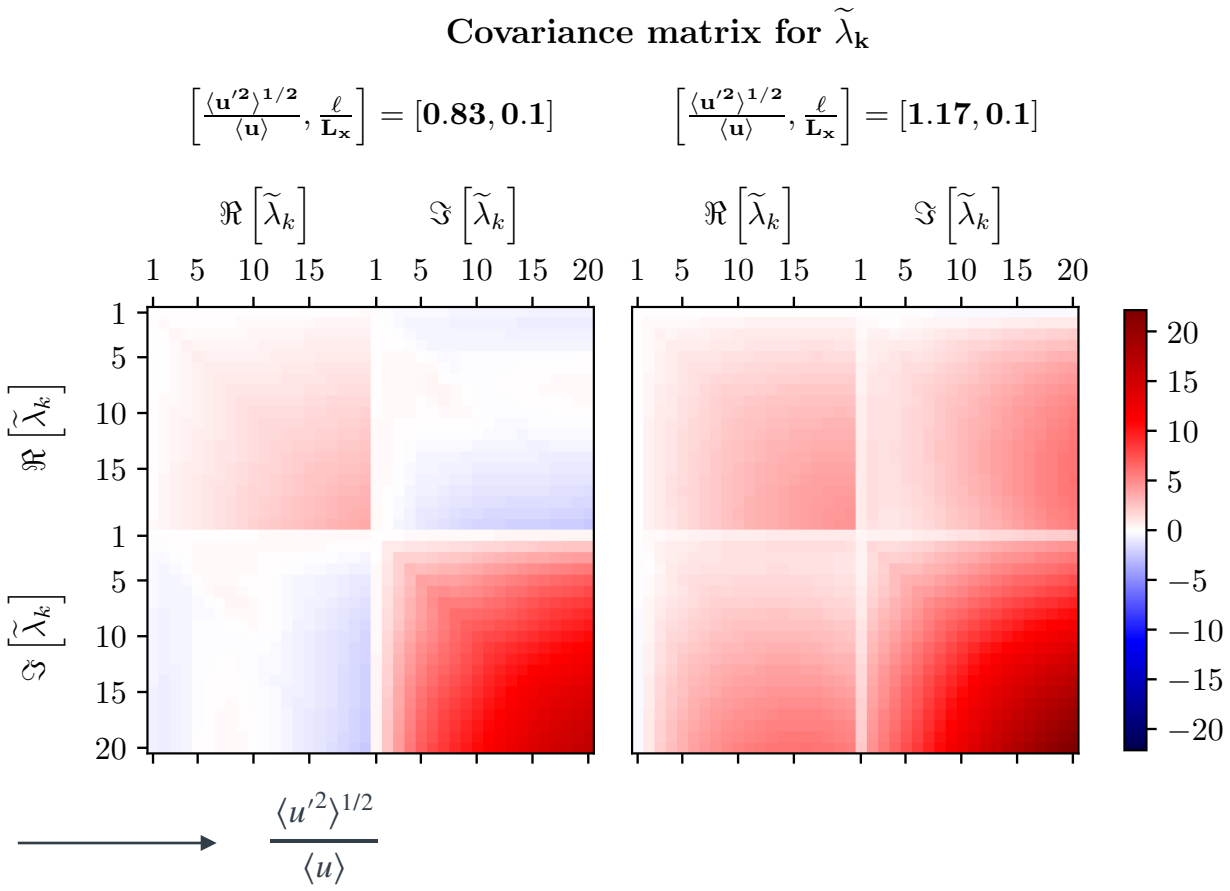


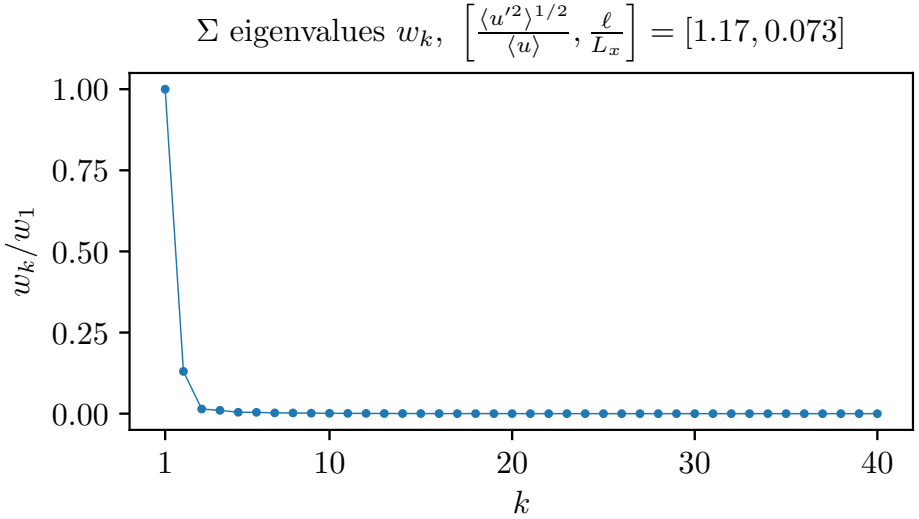
Mean $\mathfrak{R}[\widetilde{\lambda}_k]$ and $\mathfrak{I}[\widetilde{\lambda}_k]$ don't depend on a fixed power of k.

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



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





- Now we're going to build our model-form uncertainty representation.
- Defined in terms of $\widetilde{\mathscr{L}}$ and its eigenvalues.
- Unresolved dependence on $\langle u'c' \rangle$ approximated in terms of $\frac{\langle u'^2 \rangle^{1/2}}{\langle u \rangle}$ and $\frac{\ell}{L_x}$.
- Express irreducible uncertainty in $\widetilde{\mathscr{L}}$ by defining $p(\widetilde{\boldsymbol{\lambda}})$ w.r.t hyper parameters $\boldsymbol{\xi}$, i.e. $p(\widetilde{\boldsymbol{\lambda}}; \boldsymbol{\xi})$.

Prior knowledge

- Deterministic constraints

$\lambda_0 = 0, \Re[\widetilde{\lambda}_k] \leq 0$

- Scenario constraints

$\frac{\langle u'^2 \rangle^{1/2}}{\langle u \rangle}$ and $\frac{\ell}{L_x} \to 0 \implies \mathbb{E}[\widetilde{\boldsymbol{\lambda}}], \mathrm{Var}[\widetilde{\boldsymbol{\lambda}}] \to 0$

$\widetilde{\lambda}$ ensemble analysis

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

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

Assume $\widetilde{\boldsymbol{\lambda}}$ constant in time, and

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

$$\mathbb{E}(\Re[\widetilde{\lambda}_k])_{model} = f_{\lambda_R}(k)$$

$$\mathbb{E}(\Im[\widetilde{\lambda}_k])_{model} = f_{\lambda_I}(k)$$

$$\Sigma_{model} = w_1 \mathbf{v_1} \mathbf{v_1}^T + w_2 \mathbf{v_2} \mathbf{v_2}^T$$

$$(\mathbf{v_1})_k = f_{\mathbf{v_1}}(k) \qquad (\mathbf{v_2})_k = f_{\mathbf{v_2}}(k)$$

- Approximated f_{λ_R}, f_{λ_I}, $f_{\mathbf{v_1}}$, and $f_{\mathbf{v_2}}$ as linear functions of k.
- Their slopes and intercepts and w_1, w_2 are the hyperparameters $\boldsymbol{\xi}$.

- **Prior information** about $\widetilde{\boldsymbol{\lambda}}$ are encoded through $\boldsymbol{\xi}$. e.g.
  - $\Re[\widetilde{\lambda}_k] \leq 0$ enforced by constraining slope, intercept of f_{λ_R} to be negative.

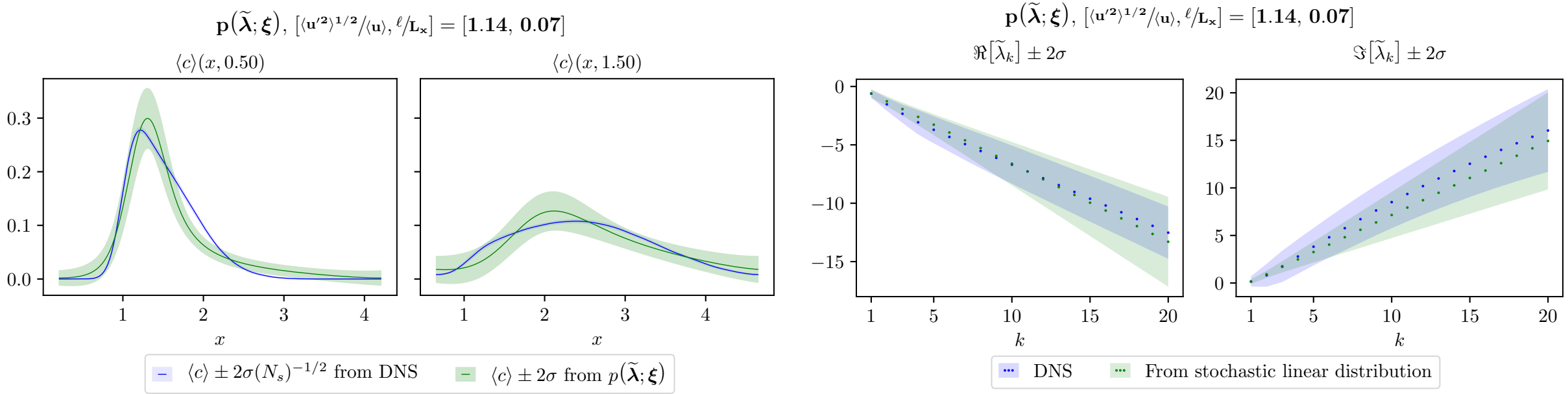
- **Scenario dependence** encoded through $\boldsymbol{\xi}$
  - $\boldsymbol{\xi}$ depend on $\dfrac{\langle u'^2 \rangle^{1/2}}{\langle u \rangle}$ and $\dfrac{\ell}{L_x}$, but dependence is uncertain: $\xi_i \sim \mathcal{N}(m_i, \sigma_i^2)$
  - $m_i = m_i\left(\dfrac{\langle u'^2 \rangle^{1/2}}{\langle u \rangle}, \dfrac{\ell}{L_x}\right), \sigma_i = \sigma_i\left(\dfrac{\langle u'^2 \rangle^{1/2}}{\langle u \rangle}, \dfrac{\ell}{L_x}\right).$
  - As scenario parameters $\to 0$, effect from dispersion $\to 0$, so $m_i(0,0) = 0$, $\sigma_i(0,0) = 0$.

- Least-square polynomial fits for m_i, σ_i as functions of $\dfrac{\langle u'^2 \rangle^{1/2}}{\langle u \rangle}, \dfrac{\ell}{L_x}$ using data from ensemble analysis.

ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

PECOS Predictive Engineering & Computational Science

Computed push-forward of $p(\widetilde{\boldsymbol{\lambda}};\boldsymbol{\xi})$ to $\langle c \rangle$ for Gaussian IC over a range of scenarios

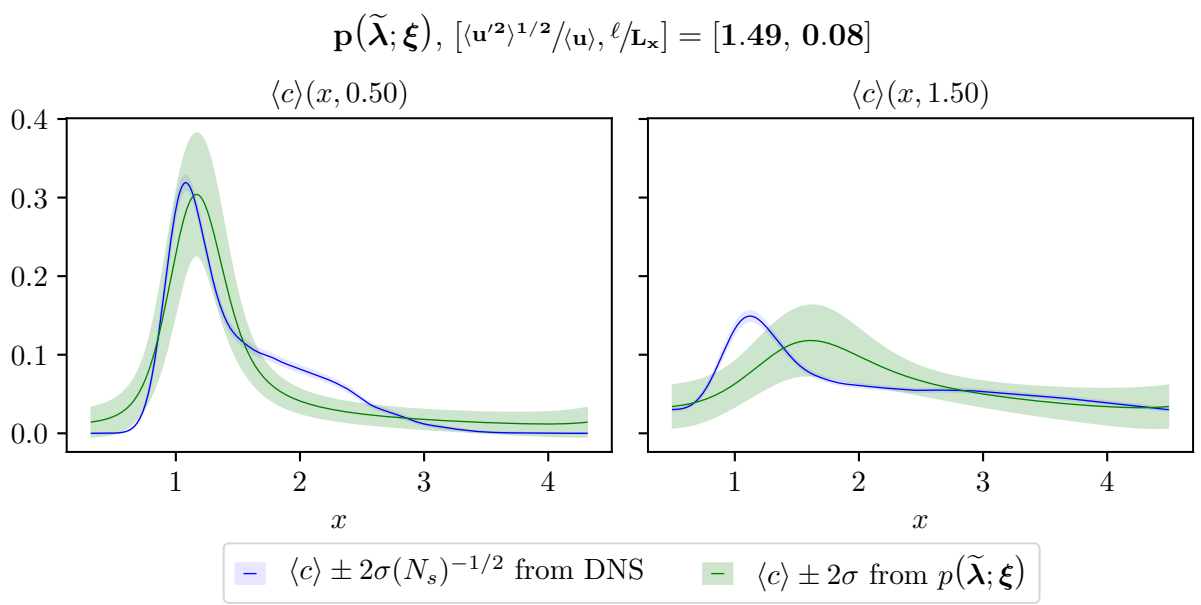


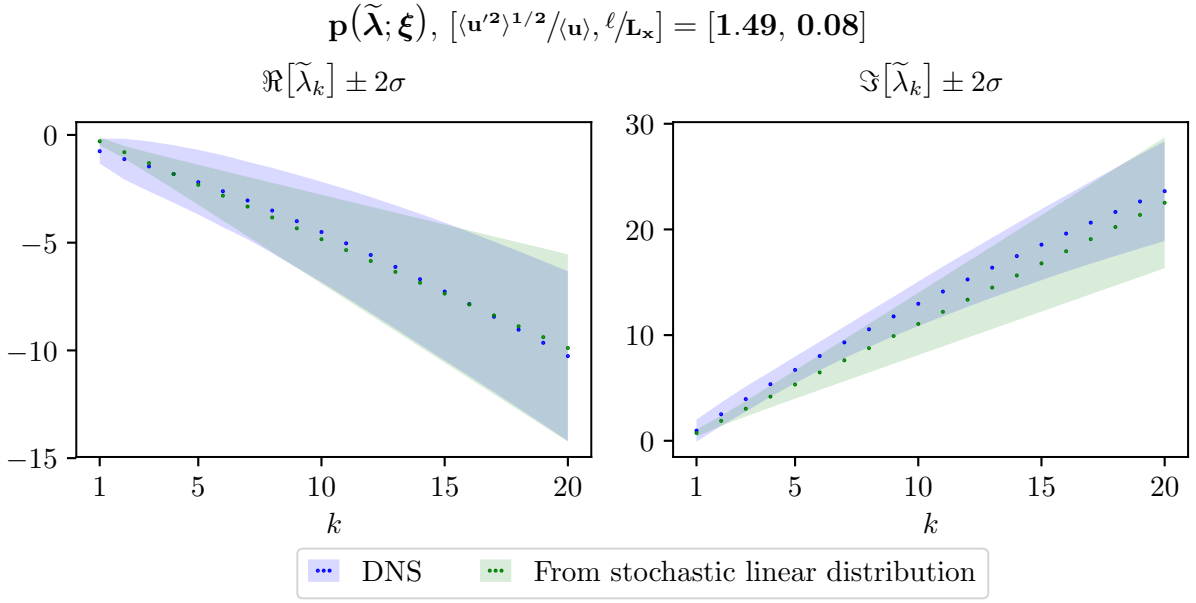
For cases on the interior of the 2D scenario grid (moderately anomalous cases), true $\langle c \rangle$ is mostly captured within uncertainty bounds.

ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

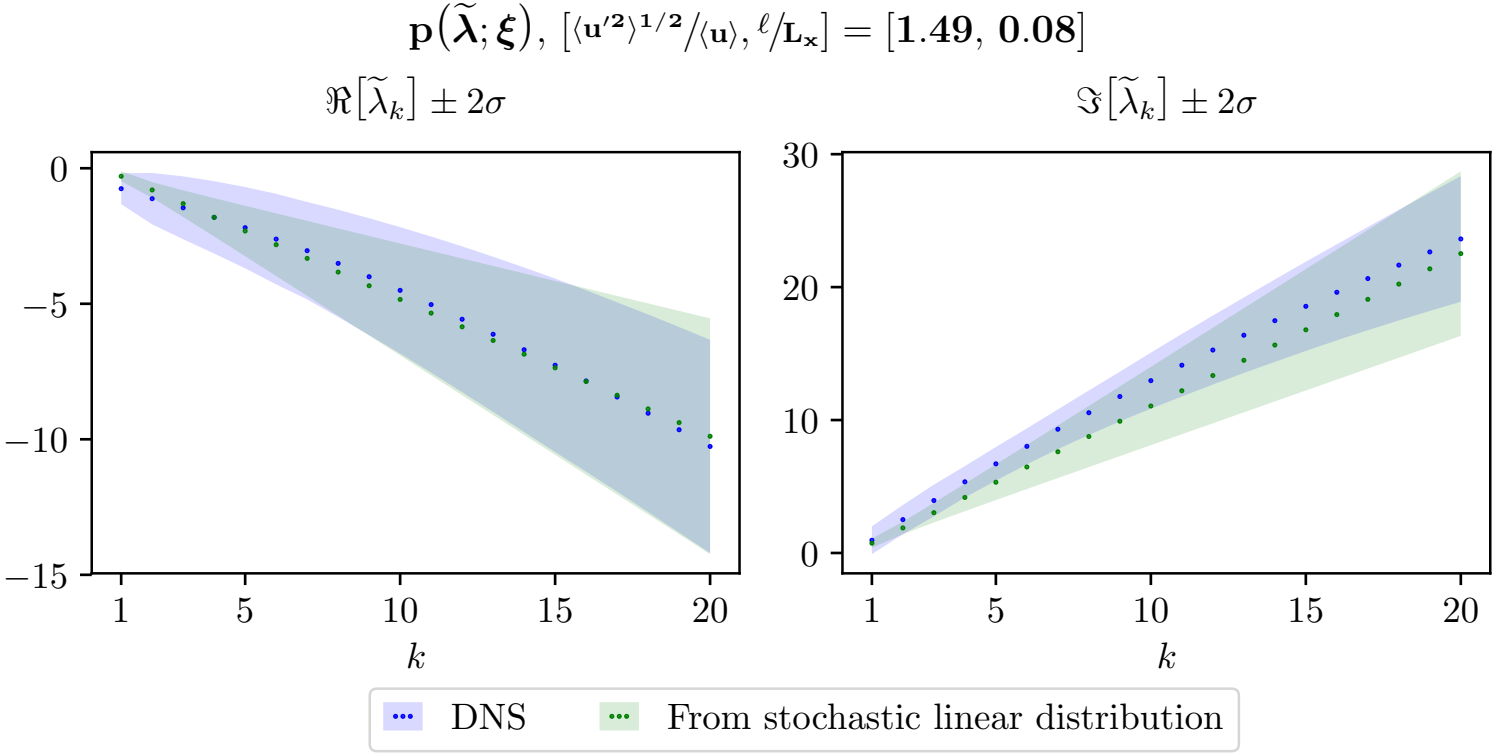
Predictive Engineering & Computational Science

For an extremely anomalous case, it fails to capture the bulk of the concentration.





The mean $\Im[\widetilde{\lambda}_k]$ are underpredicted for all k.

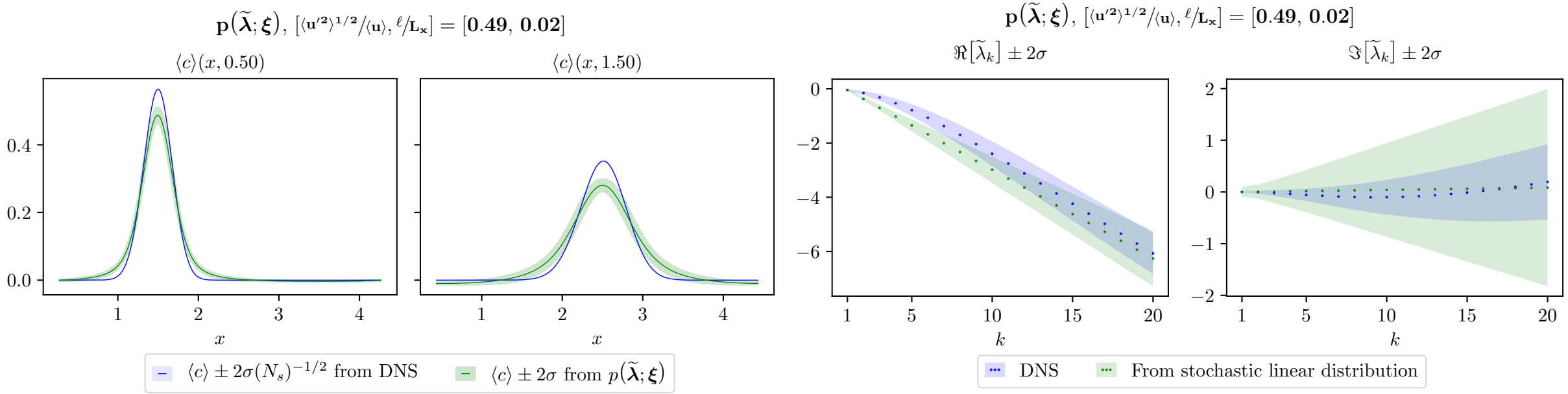


$$\frac{\partial \theta_k}{\partial t} = -\left[\langle u\rangle\left(\frac{2\pi k}{L_x}\right) - \widetilde{\lambda}_k\right],$$ so smaller $\widetilde{\lambda}_k \implies$ faster advection velocity.

ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

Predictive Engineering & Computational Science

For nonanomalous cases it overpredicts diffusion

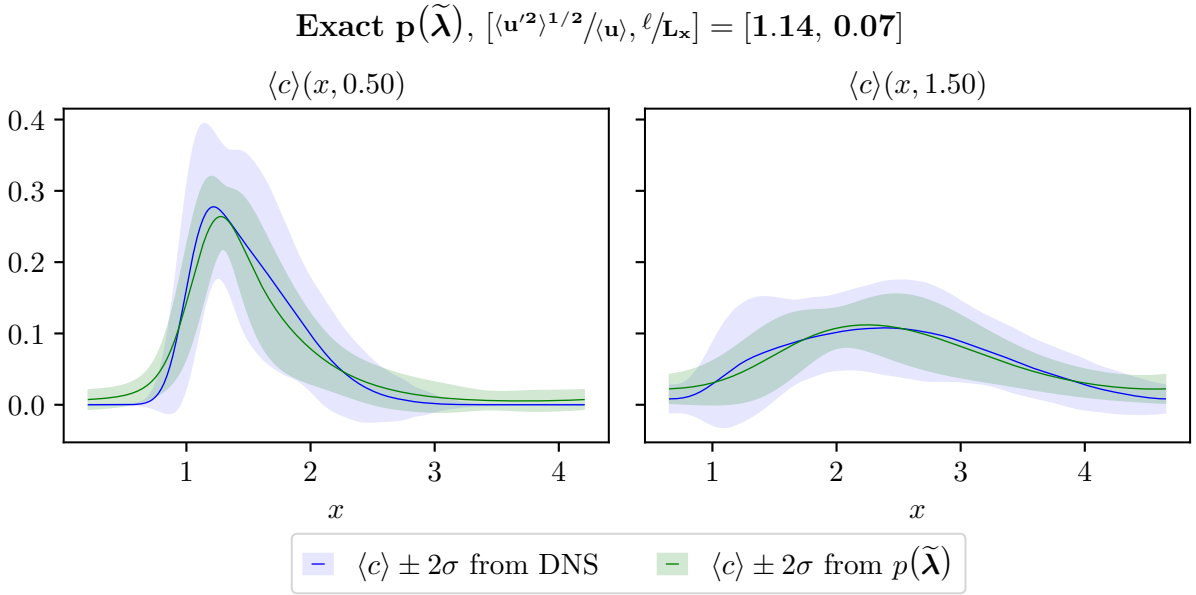


This is consistent with its prediction of larger negative $\Re[\widetilde{\lambda}_k]$.

The prototype $p(\widetilde{\boldsymbol{\lambda}}; \boldsymbol{\xi})$ isn't perfect, but we know why not:

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

Forward-propagating sample $\widetilde{\boldsymbol{\lambda}}$ from ensembles encapsulates the true $\langle c \rangle$ in all scenarios:



ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

Predictive Engineering & Computational Science

- Developed a novel method to directly probe uncertain dependence in a model.

- Used the method to generate observations of inadequacy operator's eigenvalues.
  - Learned about the operator's structure while avoiding an ill-posed inverse problem.

- Use the observations of the eigenvalues to formulate a data-informed representation of model-form uncertainty.
  - Exploited the structure of the problem to develop a relatively low-dimensional representation (40 params).
  - Simplest formulation performs relatively well, and there are clear ways to improve it

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

Contact: tnportone@utexas.edu, tporton@sandia.gov

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.