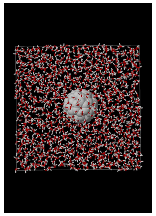
Some Basic Statistical Modeling Issues in Molecular and Ocean Dynamics

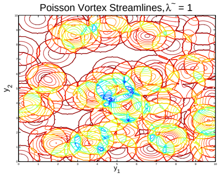
Peter R. Kramer
Department of Mathematical Sciences
Rensselaer Polytechnic Institute, Troy, NY
and
Institute for Mathematics and Its Applications, Minneapolis, MN
Supported by NSF CAREER Grant DMS-0449717
and CMG OCE-0620956

Overview

- Stochastic Drift-Diffusion Parameterization of Water Dynamics near Solute
  - with Adnan Khan and Shekhar Garde



- Statistical mesoscale modeling for oceanic flows
  - with Banu Baydil and Shafer Smith (Courant, CAOS)



Biological Disclaimer

(www.molecularium.com, S. Garde et al)

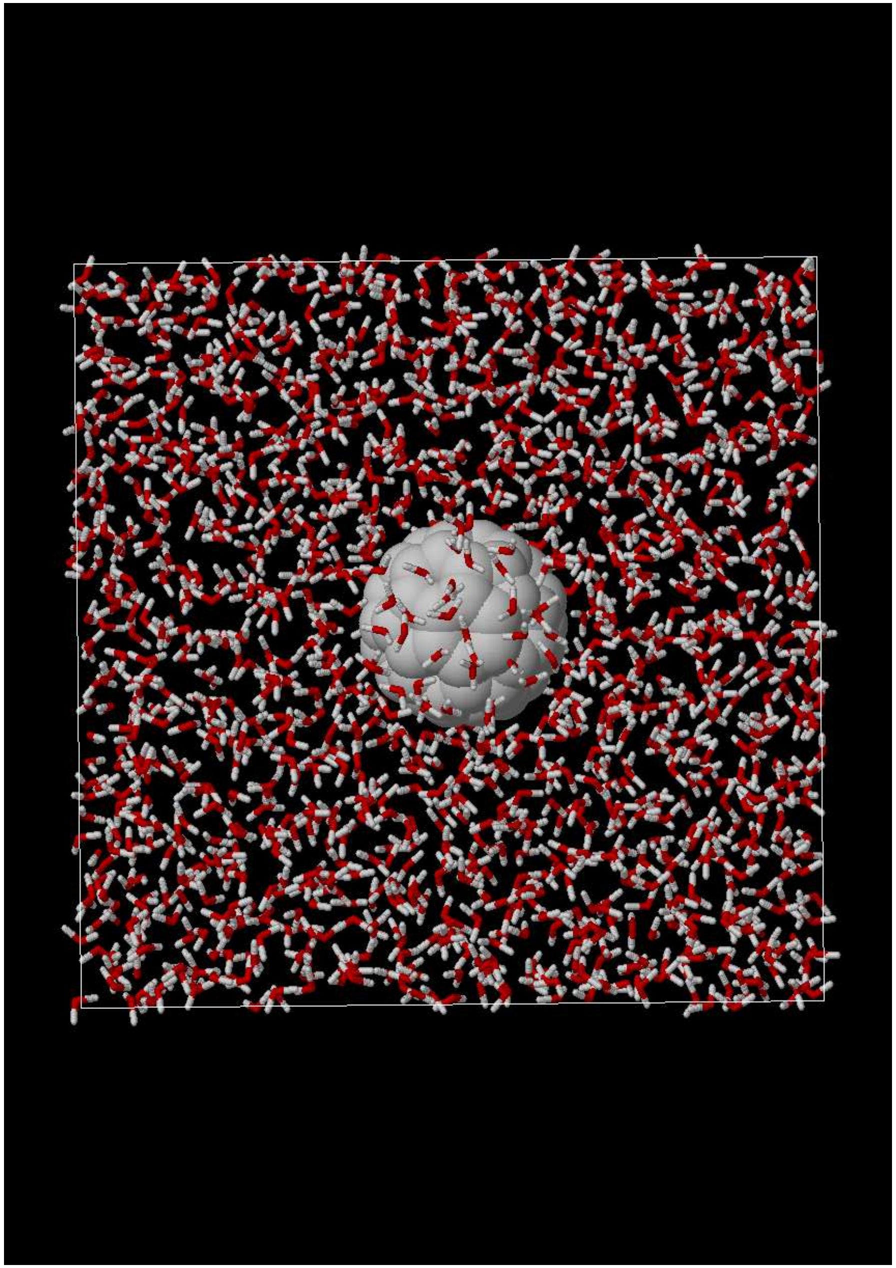
Stochastic Parameterization of Water Dynamics near Solute

Simplified statistical description of water dynamics as possible basis for implicit solvent method to accelerate molecular dynamics simulations for proteins, etc. (with Adnan Khan (Lahore) and Shekhar Garde (Biochemical Engineering))

As a first step, we explore stochastic parameterization of water near C_{60} buckyball molecule.

- isotropic, chemically simple

Molecular Dynamics Snapshot of Buckyball Surrounded by Water



Statistical dynamics encoded in biophysical literature in terms of a diffusion coefficient:

(Makarov et al, 1998; Lounnas et al, 1994,...)

$$D_B(\boldsymbol{r}) \equiv \left\langle \frac{|\boldsymbol{X}(t+2\tau) - \boldsymbol{X}(t)|^2}{6\tau} \middle| \boldsymbol{X}(t) = \boldsymbol{r} \right\rangle - \left\langle \frac{|\boldsymbol{X}(t+\tau) - \boldsymbol{X}(t)|^2}{6\tau} \middle| \boldsymbol{X}(t) = \boldsymbol{r} \right\rangle$$

But this seems to mix together inhomogeneities in mean and random motion.

Drift-Diffusion Framework

We explore capacity of models of the form

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{U}(\boldsymbol{X}(t))\,\mathrm{d}t + \Sigma(\boldsymbol{X}(t))\,\mathrm{d}\boldsymbol{W}(t),$$

for water molecule center-of-mass position $\boldsymbol{X}(t)$.

- drift vector coefficient $\boldsymbol{U}(\boldsymbol{r})$
- diffusion tensor coefficient $\mathsf{D}(\boldsymbol{r}) = \frac{1}{2}\Sigma(\boldsymbol{r})\Sigma^{\dagger}(\boldsymbol{r})$

For isometric solute (buckyball):

- $\boldsymbol{U}(\boldsymbol{r}) = U_{\parallel}(|\boldsymbol{r}|)\hat{\boldsymbol{r}}$,
- $\mathsf{D}(\boldsymbol{r}) = D_{\parallel}(|\boldsymbol{r}|)\hat{\boldsymbol{r}} \otimes \hat{\boldsymbol{r}} + D_{\perp}(|\boldsymbol{r}|)(\mathsf{I} - \hat{\boldsymbol{r}} \otimes \hat{\boldsymbol{r}})$,

for position $\boldsymbol{r} = |\boldsymbol{r}|\hat{\boldsymbol{r}}$ relative to center of symmetry.

Physically Inspired (DD-I) Model

In analogy to Brownian dynamics simulations, take

$$\boldsymbol{U}(\boldsymbol{r}) = -\gamma^{-1}\boldsymbol{\nabla}\phi(\boldsymbol{r}),$$
$$\mathsf{D}(\boldsymbol{r}) = D_0\mathsf{I}.$$

Physically Inspired (DD-I) Model

In analogy to Brownian dynamics simulations, take

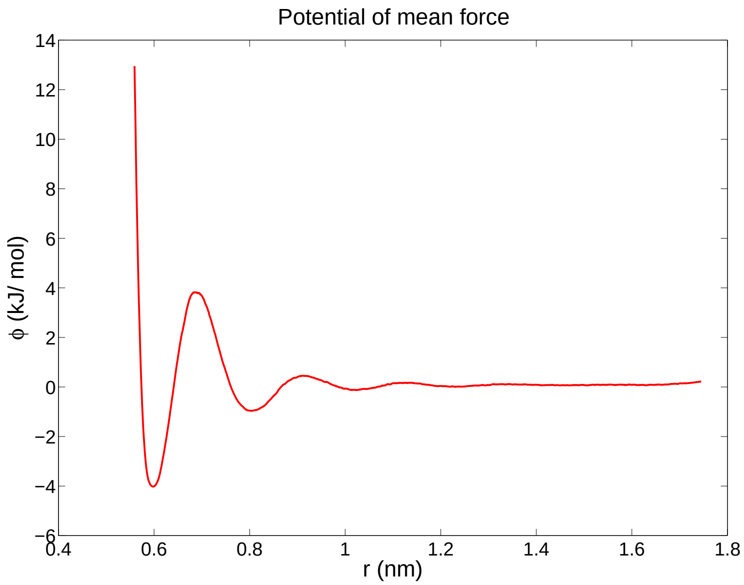
$$\boldsymbol{U}(\boldsymbol{r}) = -\gamma^{-1}\boldsymbol{\nabla}\phi(\boldsymbol{r}),$$
$$\mathsf{D}(\boldsymbol{r}) = D_0\mathsf{I}.$$

- Potential of mean force obtained from measuring concentration $c(\boldsymbol{r})$ and Boltzmann distribution $c(\boldsymbol{r}) \propto \exp(-\phi(\boldsymbol{r})/k_B T)$.
- Diffusivity unchanged from bulk value.
- Friction coefficient from Einstein relation $\gamma = k_B T/D_0$.

Physically Inspired (DD-I) Model

In analogy to Brownian dynamics simulations, take

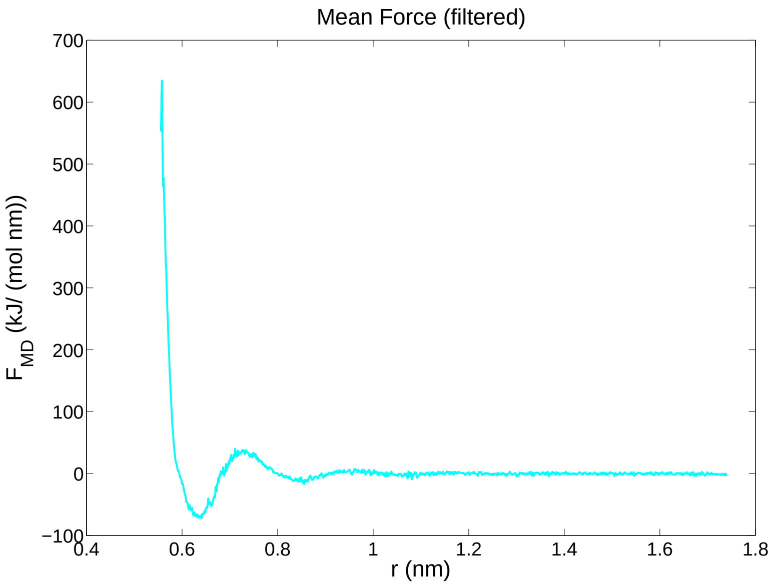
$$\boldsymbol{U}(\boldsymbol{r}) = -\gamma^{-1}\boldsymbol{\nabla}\phi(\boldsymbol{r}),$$
$$\mathsf{D}(\boldsymbol{r}) = D_0\mathsf{I}.$$



Physically Inspired (DD-I) Model

In analogy to Brownian dynamics simulations, take

$$\boldsymbol{U}(\boldsymbol{r}) = -\gamma^{-1}\boldsymbol{\nabla}\phi(\boldsymbol{r}),$$
$$\mathsf{D}(\boldsymbol{r}) = D_0\mathsf{I}.$$



Systematic, Data-Driven Parameterization (DD-II) Model

Parametrize drift and diffusion functions from mathematical definitions:

$$U_{\|}(|\boldsymbol{r}|) = \lim_{\tau \downarrow 0} \left\langle \frac{\boldsymbol{X}(t+\tau) - \boldsymbol{X}(t)}{\tau} \cdot \hat{\boldsymbol{r}} \middle| \boldsymbol{X}(t) = \boldsymbol{r} \right\rangle ,$$

$$D_{\|}(|\boldsymbol{r}|) = \lim_{\tau \downarrow 0} \left\langle \frac{\left|(\boldsymbol{X}(t+\tau) - \boldsymbol{X}(t)) \cdot \hat{\boldsymbol{r}} - U_{\|}(\boldsymbol{r})\tau\right|^2}{2\tau} \middle| \boldsymbol{X}(t) = \boldsymbol{r} \right\rangle ,$$

Systematic, Data-Driven Parameterization (DD-II) Model

Parametrize drift and diffusion functions from mathematical definitions:

$$D_{\perp}(|\boldsymbol{r}|) = \lim_{\tau \downarrow 0} \left\langle \frac{1}{4\tau} \left|(\boldsymbol{X}(t+\tau) - \boldsymbol{X}(t)) \cdot (\mathsf{I} - \hat{\boldsymbol{r}} \otimes \hat{\boldsymbol{r}})\right|^2 \middle| \boldsymbol{X}(t) = \boldsymbol{r} \right\rangle .$$

Obtain statistical data from MD simulations.

Time Difference τ must be chosen carefully

Taking $\tau = \Delta t$ (time step of MD simulation) may not be appropriate

- Limit $\tau \downarrow 0$ implicitly refers to times large enough for drift-diffusion approximation to be valid.

Must choose $T_v \ll \tau \ll T_x$, where:

- T_v is time scale of momentum.
- T_x is time scale of position.

See also Pavliotis and Stuart (2007) about need to undersample.
How choose τ in practice?

Explore simple Ornstein-Uhlenbeck (OU) model

$$\begin{aligned} \mathrm{d}\boldsymbol{X} &= \boldsymbol{V}\,\mathrm{d}t, \\ m\mathrm{d}\boldsymbol{V} &= -\gamma\boldsymbol{V}\,\mathrm{d}t - \alpha\boldsymbol{X}\,\mathrm{d}t + \sqrt{2k_BT\gamma}\,\mathrm{d}\boldsymbol{W}(t) \end{aligned}$$

Explore simple Ornstein-Uhlenbeck (OU) model

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{V}\,\mathrm{d}t,$$
$$m\mathrm{d}\boldsymbol{V} = -\gamma\boldsymbol{V}\,\mathrm{d}t - \alpha\boldsymbol{X}\,\mathrm{d}t + \sqrt{2k_BT\gamma}\,\mathrm{d}\boldsymbol{W}(t)$$

Forces: friction, potential, and thermal.

Explore simple Ornstein-Uhlenbeck (OU) model

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{V}\,\mathrm{d}t,$$
$$m\mathrm{d}\boldsymbol{V} = -\gamma\boldsymbol{V}\,\mathrm{d}t - \alpha\boldsymbol{X}\,\mathrm{d}t + \sqrt{2k_B T\gamma}\,\mathrm{d}\boldsymbol{W}(t)$$

Nondimensionalize:

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{V}\,\mathrm{d}t,$$
$$\mathrm{d}\boldsymbol{V} = -a\boldsymbol{V}\,\mathrm{d}t - a\boldsymbol{X}\,\mathrm{d}t + a\,\mathrm{d}\boldsymbol{W}(t)$$

Explore simple Ornstein-Uhlenbeck (OU) model

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{V}\,\mathrm{d}t,$$
$$m\mathrm{d}\boldsymbol{V} = -\gamma\boldsymbol{V}\,\mathrm{d}t - \alpha\boldsymbol{X}\,\mathrm{d}t + \sqrt{2k_B T\gamma}\,\mathrm{d}\boldsymbol{W}(t)$$

Nondimensionalize:

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{V}\,\mathrm{d}t,$$
$$\mathrm{d}\boldsymbol{V} = -a\boldsymbol{V}\,\mathrm{d}t - a\boldsymbol{X}\,\mathrm{d}t + a\,\mathrm{d}\boldsymbol{W}(t)$$

where $a = \gamma^2/(m\alpha)$ is ratio of position to momentum time scale.

Explore simple Ornstein-Uhlenbeck (OU) model

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{V}\,\mathrm{d}t,$$
$$m\mathrm{d}\boldsymbol{V} = -\gamma\boldsymbol{V}\,\mathrm{d}t - \alpha\boldsymbol{X}\,\mathrm{d}t + \sqrt{2k_BT\gamma}\,\mathrm{d}\boldsymbol{W}(t)$$

Nondimensionalize:

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{V}\,\mathrm{d}t,$$
$$\mathrm{d}\boldsymbol{V} = -a\boldsymbol{V}\,\mathrm{d}t - a\boldsymbol{X}\,\mathrm{d}t + a\,\mathrm{d}\boldsymbol{W}(t)$$

Exact drift-diffusion coarse-graining when $a \gg 1$:

$$\mathrm{d}\boldsymbol{X} = -\boldsymbol{X}\,\mathrm{d}t + \mathrm{d}\boldsymbol{W}(t)$$

Explore simple Ornstein-Uhlenbeck (OU) model

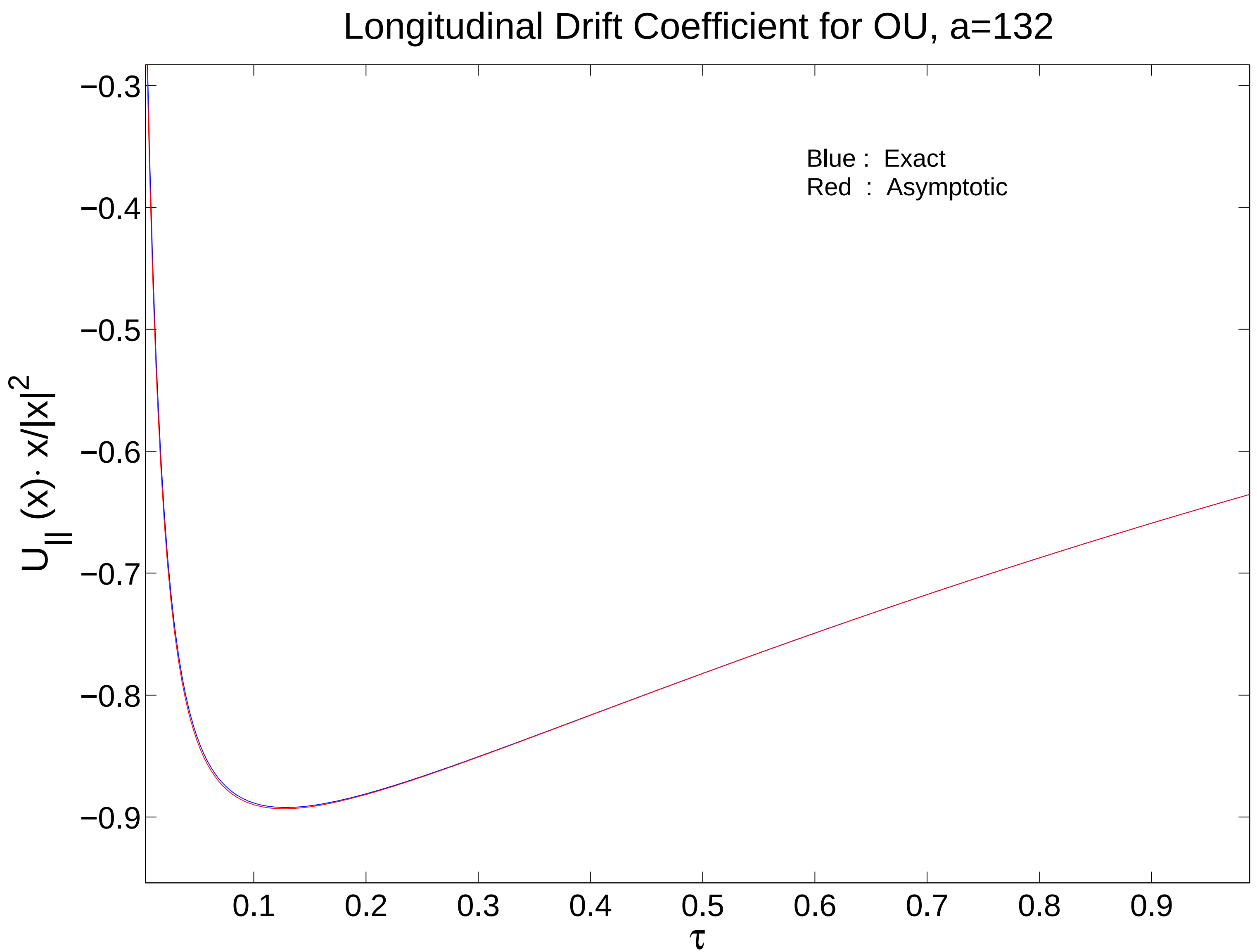
$$
\begin{aligned}
\mathrm{d}\boldsymbol{X} &= \boldsymbol{V}\,\mathrm{d}t, \\
m\mathrm{d}\boldsymbol{V} &= -\gamma\boldsymbol{V}\,\mathrm{d}t - \alpha\boldsymbol{X}\,\mathrm{d}t + \sqrt{2k_BT\gamma}\,\mathrm{d}\boldsymbol{W}(t)
\end{aligned}
$$

Nondimensionalize:

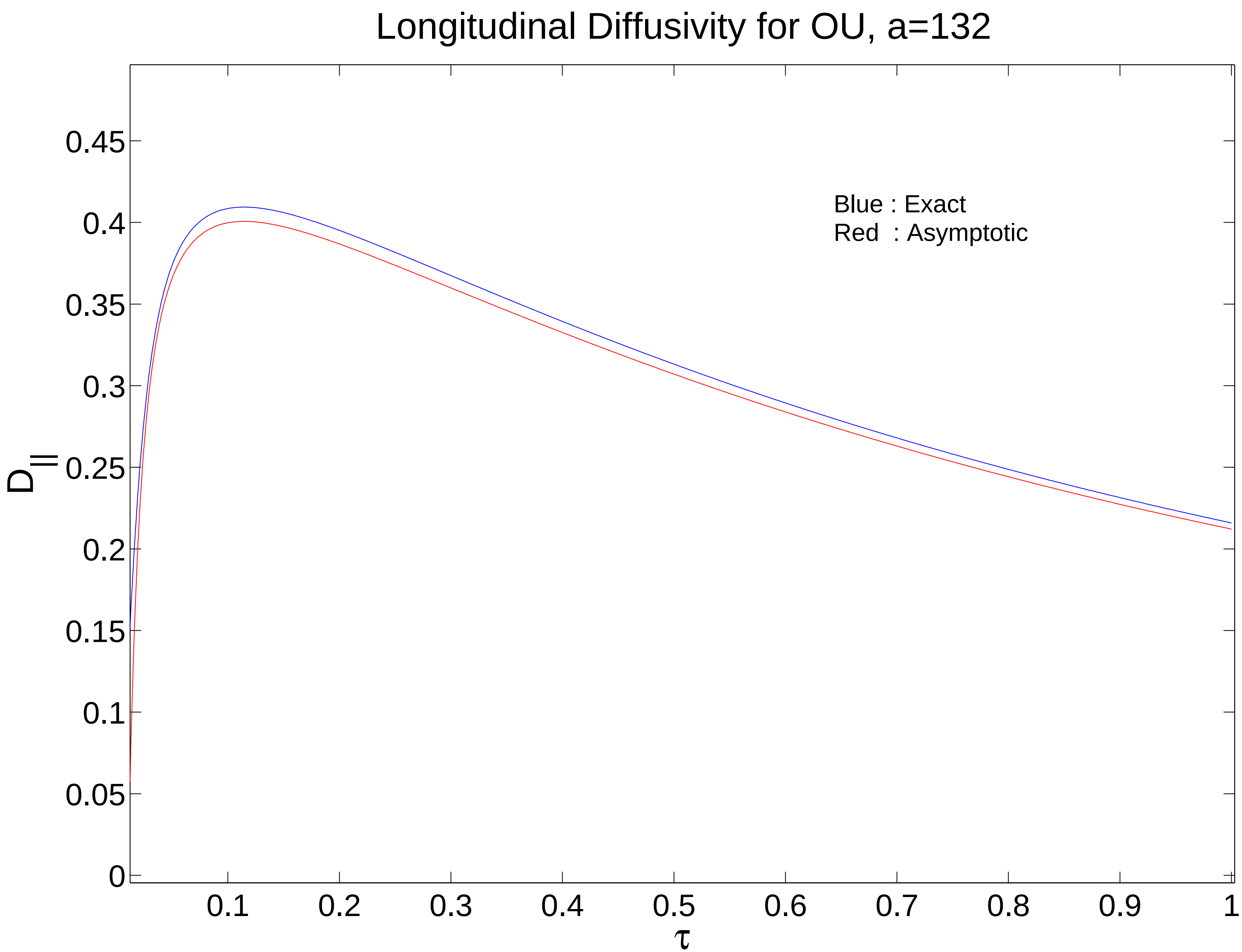
$$
\begin{aligned}
\mathrm{d}\boldsymbol{X} &= \boldsymbol{V}\,\mathrm{d}t, \\
\mathrm{d}\boldsymbol{V} &= -a\boldsymbol{V}\,\mathrm{d}t - a\boldsymbol{X}\,\mathrm{d}t + a\,\mathrm{d}\boldsymbol{W}(t)
\end{aligned}
$$

What if we try to obtain this from analysis of trajectories with finite but large a?

Drift and diffusion coefficients of exact OU solution sampled with finite time difference τ.



Drift and diffusion coefficients of exact OU solution sampled with finite time difference τ.

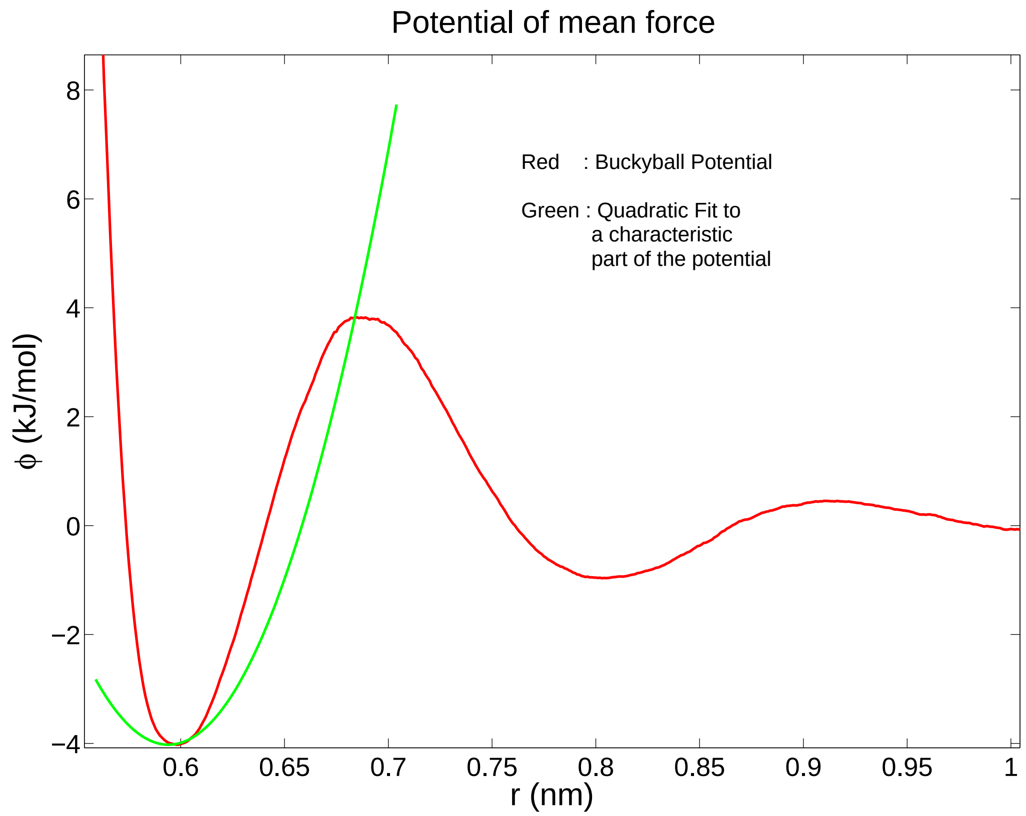


Inferences from OU model

- Good choice of τ may be the one which maximizes drift magnitude and diffusivity.
- Beginning estimate obtained from OU model with same a value.

OU Model Insights → MD Data Parameterization in DD-II Model

To obtain time scales, approximate main well in potential of mean force by quadratic.



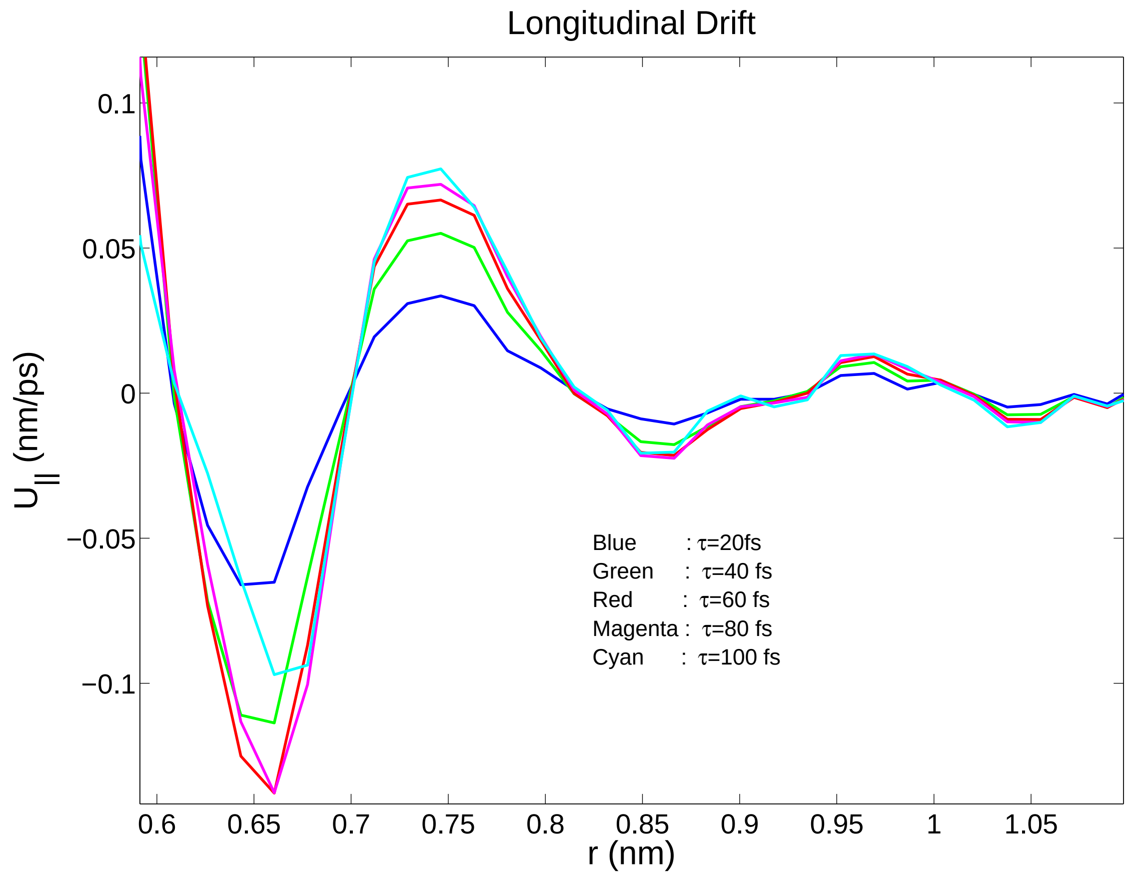
This gives $a = 132$.

OU Model Insights $\longrightarrow$ MD Data Parameterization in DD-II Model

Examine drift and diffusivity computed from various choices of τ.

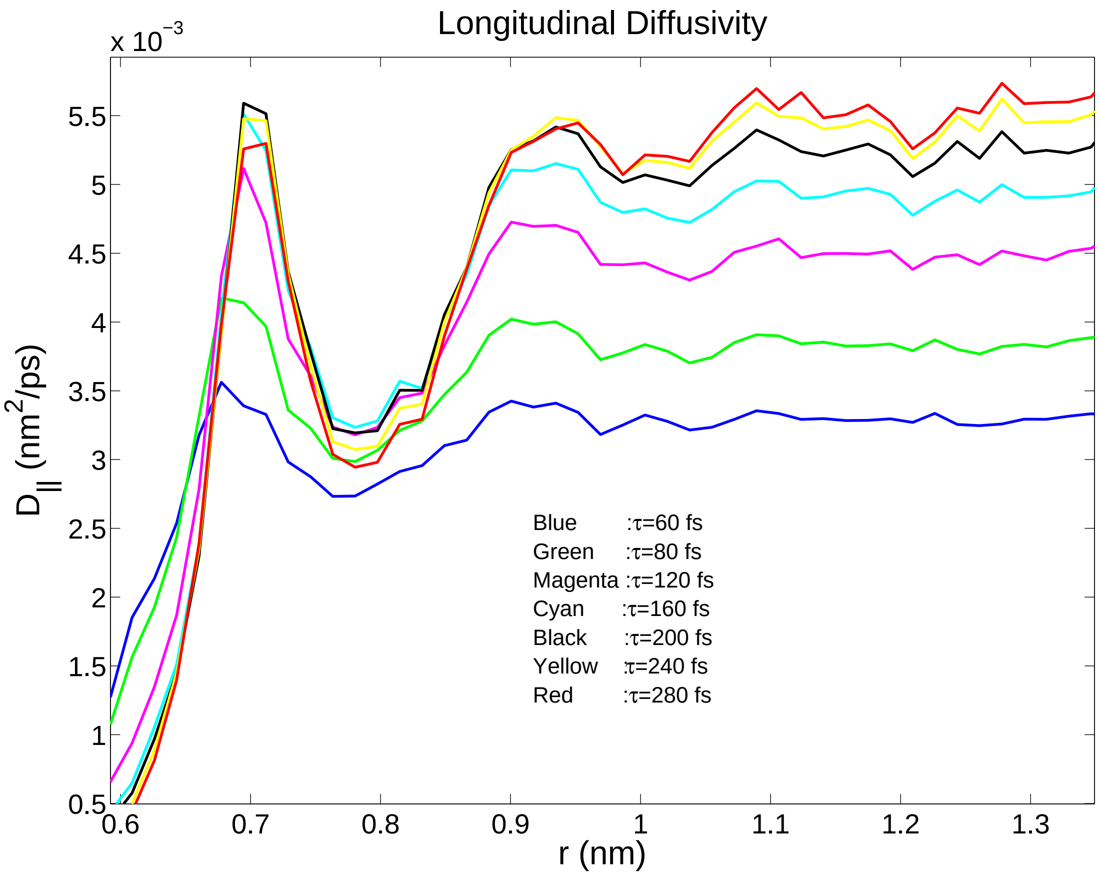
OU Model Insights → MD Data Parameterization in DD-II Model

Examine drift and diffusivity computed from various choices of τ.



OU Model Insights → MD Data Parameterization in DD-II Model

Examine drift and diffusivity computed from various choices of τ.



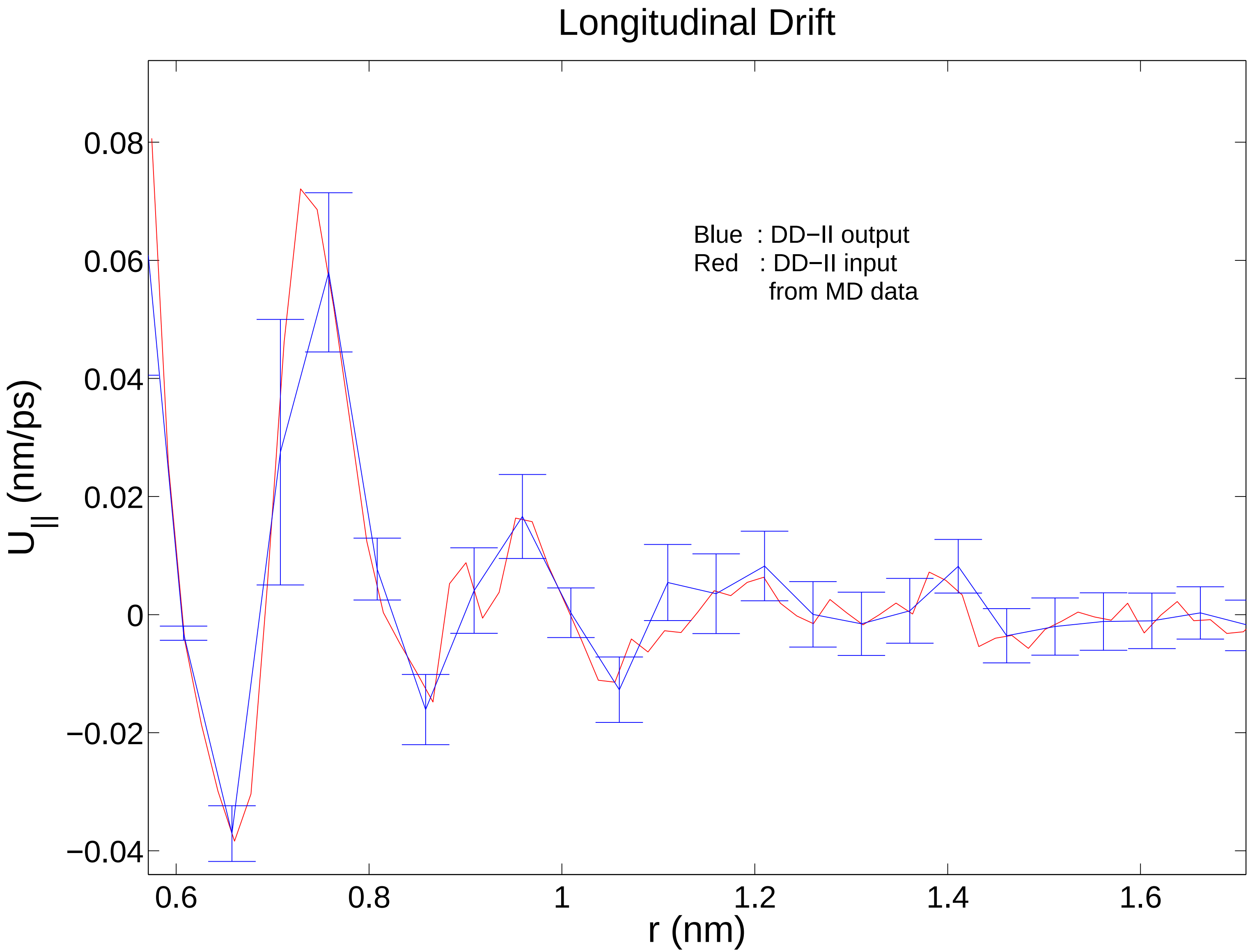
OU Model Insights $\rightarrow$ MD Data Parameterization in DD-II Model

Examine drift and diffusivity computed from various choices of τ. Both desiderata about drift and diffusivity behavior not simultaneously satisfiable.

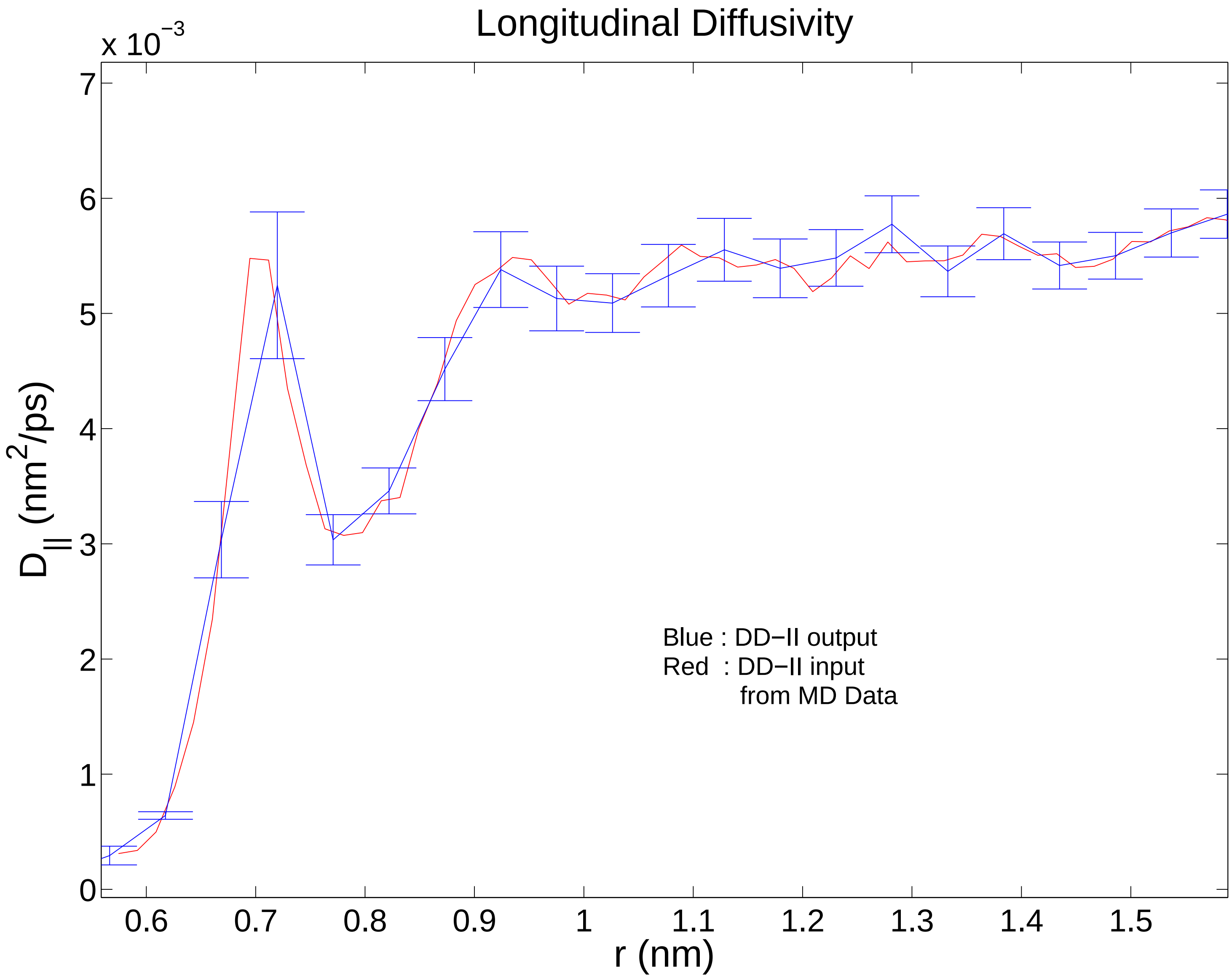
- Correct bulk diffusivity behavior more important

We choose $\tau = 0.2$ ps $= 200$ fs.

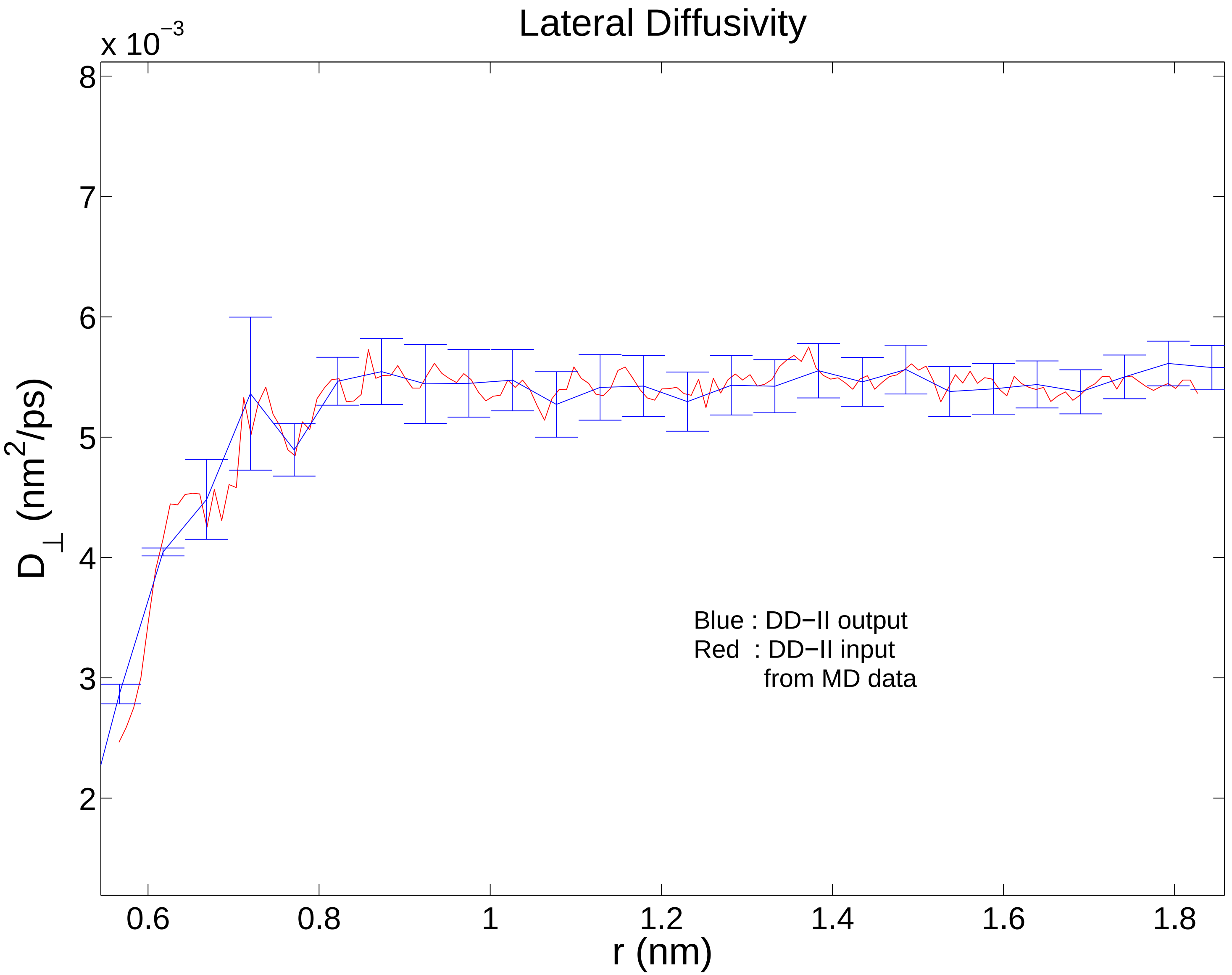
Parameterization Used in DD-II Model



Parameterization Used in DD-II Model



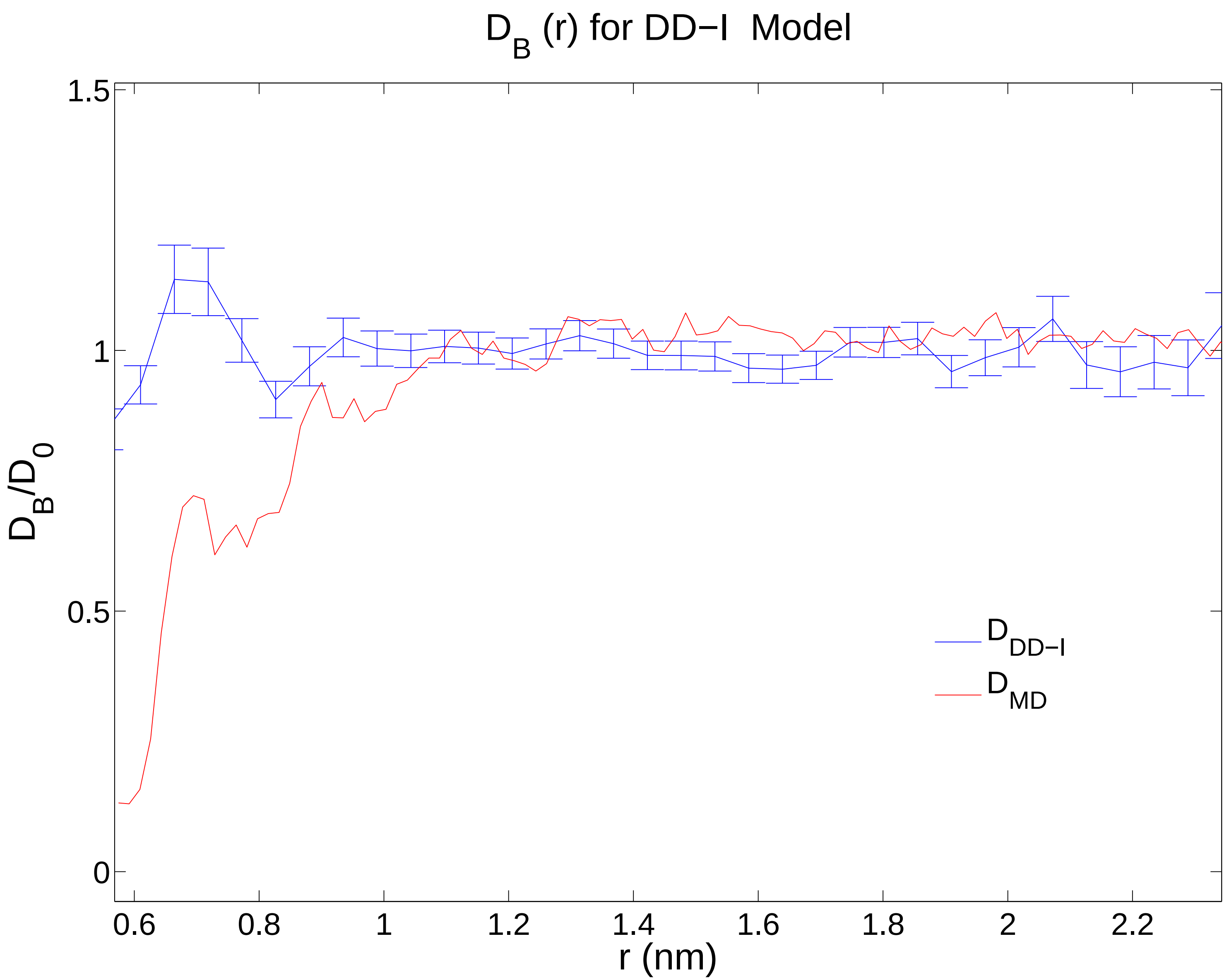
Parameterization Used in DD-II Model



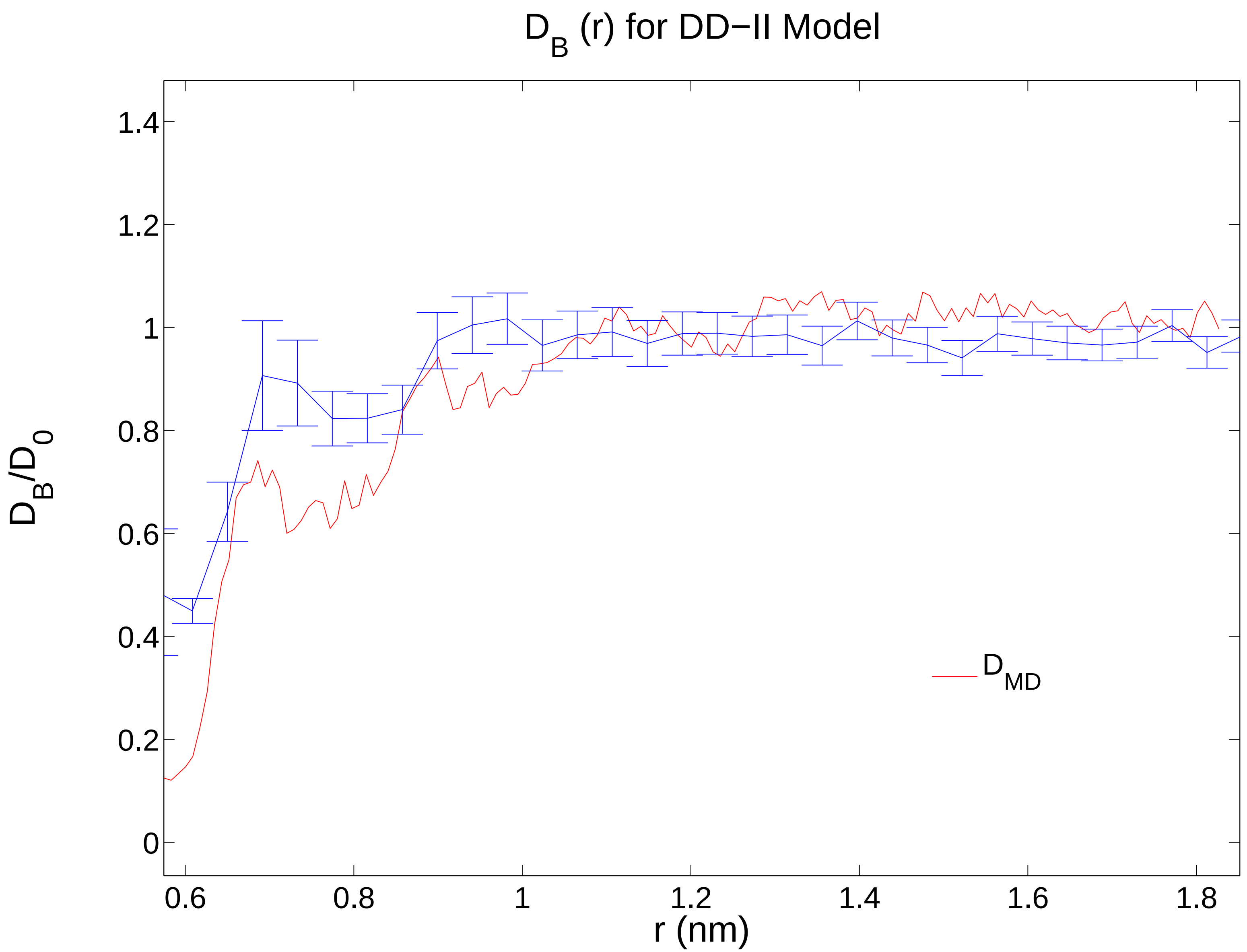
Compare Predictions of Biophysical Diffusivity Formula

$$D_B(\boldsymbol{r}) \equiv \left\langle \left. \frac{|\boldsymbol{X}(t+2\tau) - \boldsymbol{X}(t)|^2}{6\tau} \right| \boldsymbol{X}(t) = \boldsymbol{r} \right\rangle - \left\langle \left. \frac{|\boldsymbol{X}(t+\tau) - \boldsymbol{X}(t)|^2}{6\tau} \right| \boldsymbol{X}(t) = \boldsymbol{r} \right\rangle$$

Compare Predictions of Biophysical Diffusivity Formula



Compare Predictions of Biophysical Diffusivity Formula



Future Work

Next steps

- anisotropies
- chemical heterogeneity

Unresolved Mesoscale Turbulence in Ocean Circulation Models

Computational ocean models for climate prediction have resolution $\sim$ 100 km:

- does not adequately resolve mesoscale turbulent structures on length scales $\lesssim$ 100 km
- even smaller scales $\lesssim$ 100 m : Kolmogorov turbulence

We focus on representing turbulent transport by unresolved mesoscale turbulence.

Wind-driven double-gyre 2000 km basin-scale computational simulation, 5 km resolution, by Shafer Smith.

Mathematical Framework for Transport

$$\frac{\partial T(\boldsymbol{x},t)}{\partial t} + \boldsymbol{u}(\boldsymbol{x},t)\cdot\boldsymbol{\nabla}T(\boldsymbol{x},t) = \kappa\Delta T(\boldsymbol{x},t),$$
$$T(\boldsymbol{x},t=0) = T_{\mathrm{in}}(\boldsymbol{x})$$

- Passive scalar field $T(\mathbf{x},t)$
- Velocity field $\boldsymbol{u} = \boldsymbol{V} + \boldsymbol{v}$: large-scale mean flow + small-scale fluctuations
- "Molecular" diffusion coefficient κ

Parameterization Problem

- Obtain an equation for coarse-grained $\langle T \rangle$:

$$\frac{\partial \langle T \rangle}{\partial t} + \boldsymbol{V} \cdot \boldsymbol{\nabla} \langle T \rangle = \kappa \Delta \langle T \rangle - \boldsymbol{\nabla} \cdot \boldsymbol{F}$$

$$\langle T \rangle(\boldsymbol{x}, t = 0) = T_{\text{in}}(\boldsymbol{x}),$$

- Turbulent Flux $\boldsymbol{F} = \langle \boldsymbol{v}\, T \rangle$
- Problem of Parametrization:

$$\boldsymbol{F} = \mathcal{F}(\langle T \rangle, \boldsymbol{V})$$

where $\mathcal{F}$ only involves a few parameters.

Parameterization of Small Scales

One parameterization used in AOS and engineering (Gent and McWilliams 1990, Griffies 1998, Middleton and Loder 1989):

$$\boldsymbol{F} = -\mathsf{K}^*(\boldsymbol{x}, t) \cdot \boldsymbol{\nabla} \langle T \rangle,$$

with generally nonsymmetric $\mathsf{K}^* = \mathsf{S}^* + \mathsf{A}^*$.

- Symmetric part S^*: variably enhanced diffusion
- Antisymmetric part A^*: effective drift $\boldsymbol{U}^* = -\boldsymbol{\nabla} \cdot \mathsf{A}^*$ relative to mean flow

Parameterization Approaches

Within the class of "effective diffusivity" or "eddy diffusivity" schemes, the way in which K^* is modeled differs. Some examples:

- Mixing-length theory: $\mathsf{K}^* \sim \ell v$.
- $K - \varepsilon$ theory: parameterize based on local energy and energy dissipation rate
- Gent-McWilliams: related to slope of surfaces of constant potential density

These are empirical, with varying degrees of success.
Ocean circulation models often employ even simpler practice of choosing K^* as constant multiple of identity, tuned *a posteriori*.

Homogenization-Based Parameterization

Under assumption of scale separation (not so crazy for ocean), homogenization theory provides rigorous support and formula for effective diffusivity (Avellaneda, Majda, McLaughlin, Papanicolaou, Varadhan, Fannjiang, Pavliotis & K):

$$\mathsf{K}^*(\boldsymbol{x}, t) = \kappa \left(\mathsf{I} - \langle \boldsymbol{v} \otimes \boldsymbol{\chi} \rangle \right)$$

where

$$\frac{\partial \boldsymbol{\chi}}{\partial \tau} + (\boldsymbol{V} + \boldsymbol{v}) \cdot \boldsymbol{\nabla}_{\boldsymbol{y}} \boldsymbol{\chi} - \kappa \Delta_{\boldsymbol{y}} \boldsymbol{\chi} = -\boldsymbol{v},$$

is cell problem, solved on sub-grid scale with frozen mean flow $\boldsymbol{V}$. $\langle \cdot \rangle$ denotes subgrid-scale average.

Computational Homogenization

- Heterogenous multiscale methods (HMM) natural, but perhaps too complex a modification to existing codes.
- More negotiable: parameterization formula for K^* based on statistical subgrid-scale model with small number of nondimensional parameters
- Our approach: build up from simple models, combine numerical solutions with new and existing asymptotic results (Avellaneda, Majda, Vergassola, Bonn, McLaughlin, Kesten, Papanicolaou, etc.)

Example: Dynamical Poisson Vortex Model (Cinlar, Caglar,...)

Stream function:

$$\psi(\boldsymbol{y}, t) = \sum_n \psi_{\text{vor}}(\boldsymbol{y} - \mathbf{Y}_c^{(n)}(t - \xi^{(n)}), t - \xi^{(n)})$$

Example: Dynamical Poisson Vortex Model (**Cinlar**, **Caglar**,...)

Stream function:

$$\psi(\boldsymbol{y},t) = \sum_n \psi_{\text{vor}}(\boldsymbol{y} - \mathbf{Y}_c^{(n)}(t-\xi^{(n)}), t-\xi^{(n)})$$

composed of simple vortex blobs, e.g.,

$$\boldsymbol{v}(\boldsymbol{y},t) = \begin{bmatrix} \partial\psi_{\text{vor}}/\partial y_2 \\ -\partial\psi_{\text{vor}}/\partial y_1 \end{bmatrix},$$

$$\psi_{\text{vor}}(\boldsymbol{y}, t=0) = 12\bar{v}_{\text{vor}}\ell_{\text{vor}} \left[\frac{1}{9}\left|\frac{\boldsymbol{y}}{\ell_{\text{vor}}}\right|^3 - \frac{1}{6}\left|\frac{\boldsymbol{y}}{\ell_{\text{vor}}}]\right|^2 \right]$$

for $|\boldsymbol{y}| \leq \ell_{\text{vor}}$

Example: Dynamical Poisson Vortex Model (**Cinlar**, **Caglar**,...)

Stream function:

$$\psi(\boldsymbol{y}, t) = \sum_{n} \psi_{\mathrm{vor}}(\boldsymbol{y} - \mathbf{Y}_c^{(n)}(t - \xi^{(n)}), t - \xi^{(n)})$$

with centers $\mathbf{Y}_c^{(n)}(t)$ obeying certain simple dynamical law

- Brownian motion
- intervortical advection

and $\psi_{\mathrm{vor}}(\boldsymbol{y}, t) \to 0$ for $|t| \to \infty$,

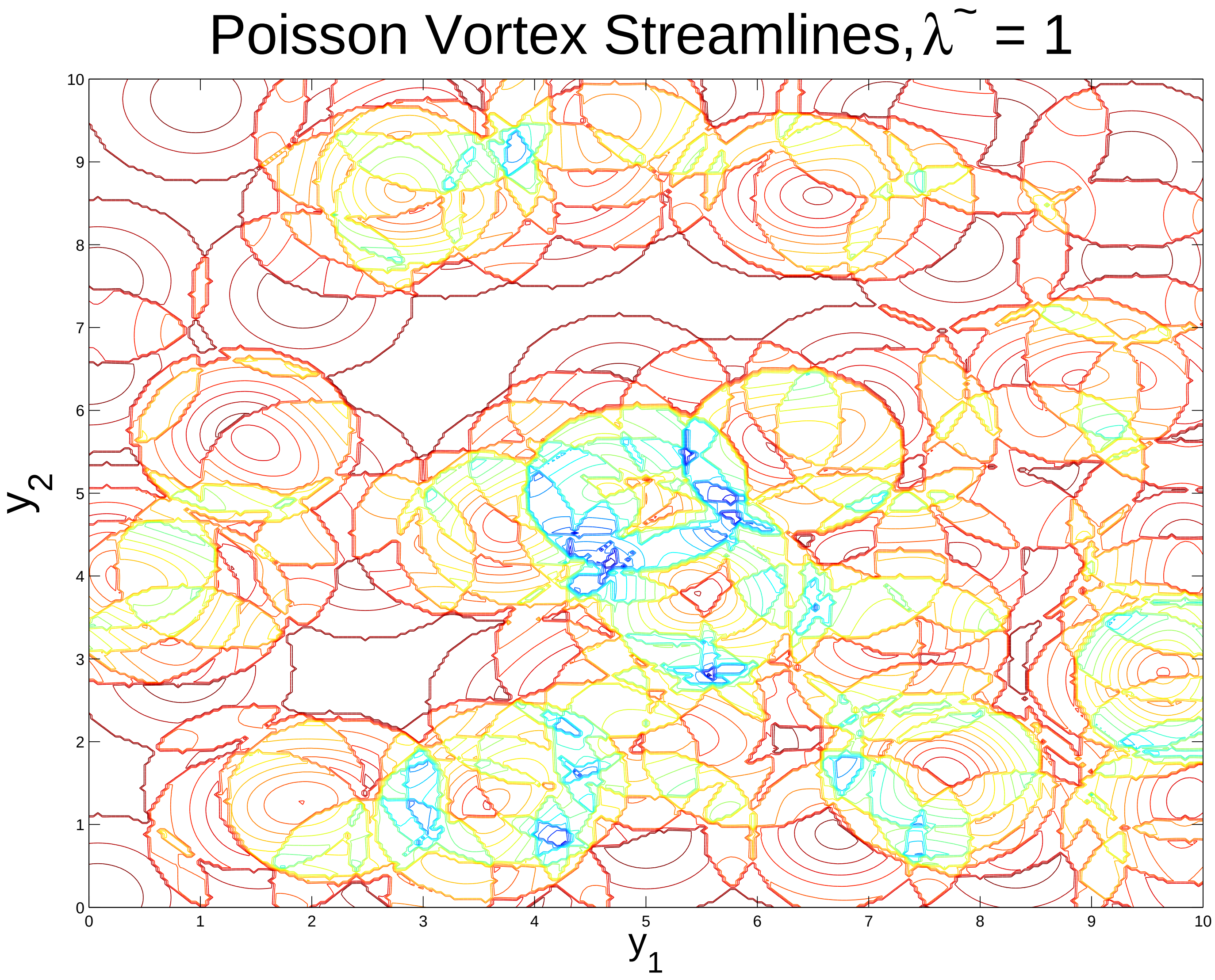
Example: Dynamical Poisson Vortex Model (Cinlar, Caglar,...)

Stream function:

$$\psi(\boldsymbol{y},t) = \sum_n \psi_{\text{vor}}(\boldsymbol{y} - \mathbf{Y}_c^{(n)}(t-\xi^{(n)}), t-\xi^{(n)})$$

vortices nucleated according to space-time Poisson process $(\boldsymbol{\eta}^{(n)}, \xi^{(n)})$ of prescribed intensity λ, with $\mathbf{Y}_c^{(n)}(0) = \boldsymbol{\eta}^{(n)}$

Sample Snapshot of Poisson Vortex Field



Uncertainty Modeling Issues

How choose the statistical model parameters for dynamical parameterization?

- for observed subgrid-scale data, can try parametric statistical fitting
- but want statistical subgrid-scale model to represent unobserved small scales, with only coarse-scale data available
- so how predict small-scale statistical parameters from large-scale observations?

References regarding Statistical Modeling in Molecular Dynamics

- T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, 2002.
- V. A. Makarov et al, *Biophys J.*, **75** (1998) 150–158.
  - biophysical modeling framework for water near surface
- G. Pavliotis and A. M. Stuart, *J. Stat. Phys.* **124** (2007) 741–781.
  - mathematical framework for computing coarse-grained coefficients statistically from microscale data

References regarding Statistical Modeling in Ocean Turbulence

- *Mathematics of Subgrid-Scale Phenomena in Atmospheric and Oceanic Flows*, Institute for Pure and Applied Mathematics, `www.ipam.ucla.edu/programs/atm2002`
- A. J. Majda and P. R. Kramer, *Phys. Rep.* **314** (1999) 237–574.
  - review of some statistical flow models
- M. Çaglar, *et al*, *J. Atmos. Ocean. Tech.* **23** (2006) 1745–1758.
  - parameterically fitting observational data to statistical vortex model