Reliable assessment of marine biogeochemical models using diffusion based kernel density estimation

Maria-Theresia Verwega

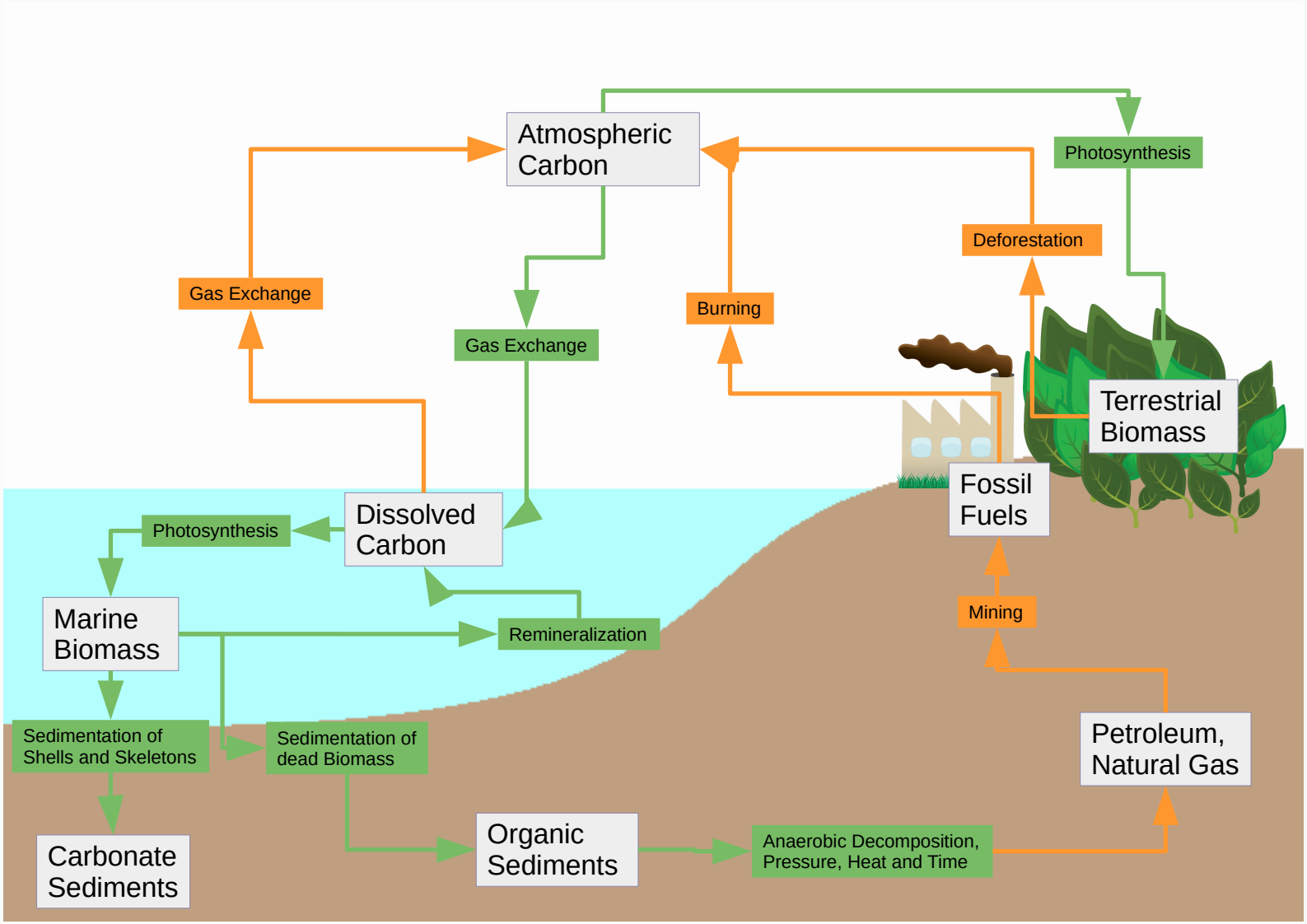
MarDATA HELMHOLTZ SCHOOL FOR MARINE DATA SCIENCE

GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel

C A U Christian-Albrechts-Universität zu Kiel Technische Fakultät

Background

The carbon cycle



Assessing a model

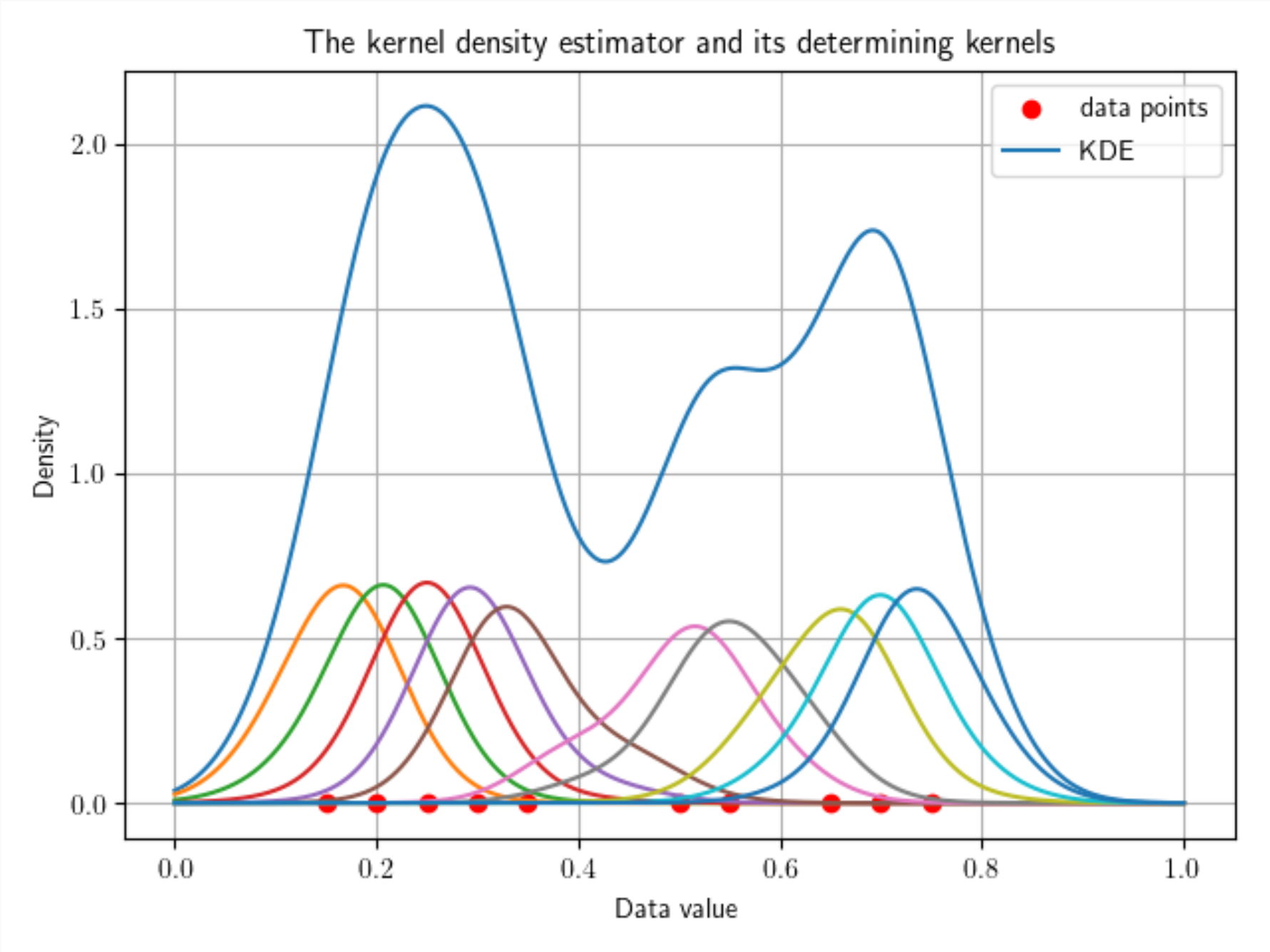
model-data-comparison:

- point by point
- probability density functions

Materials & Methods

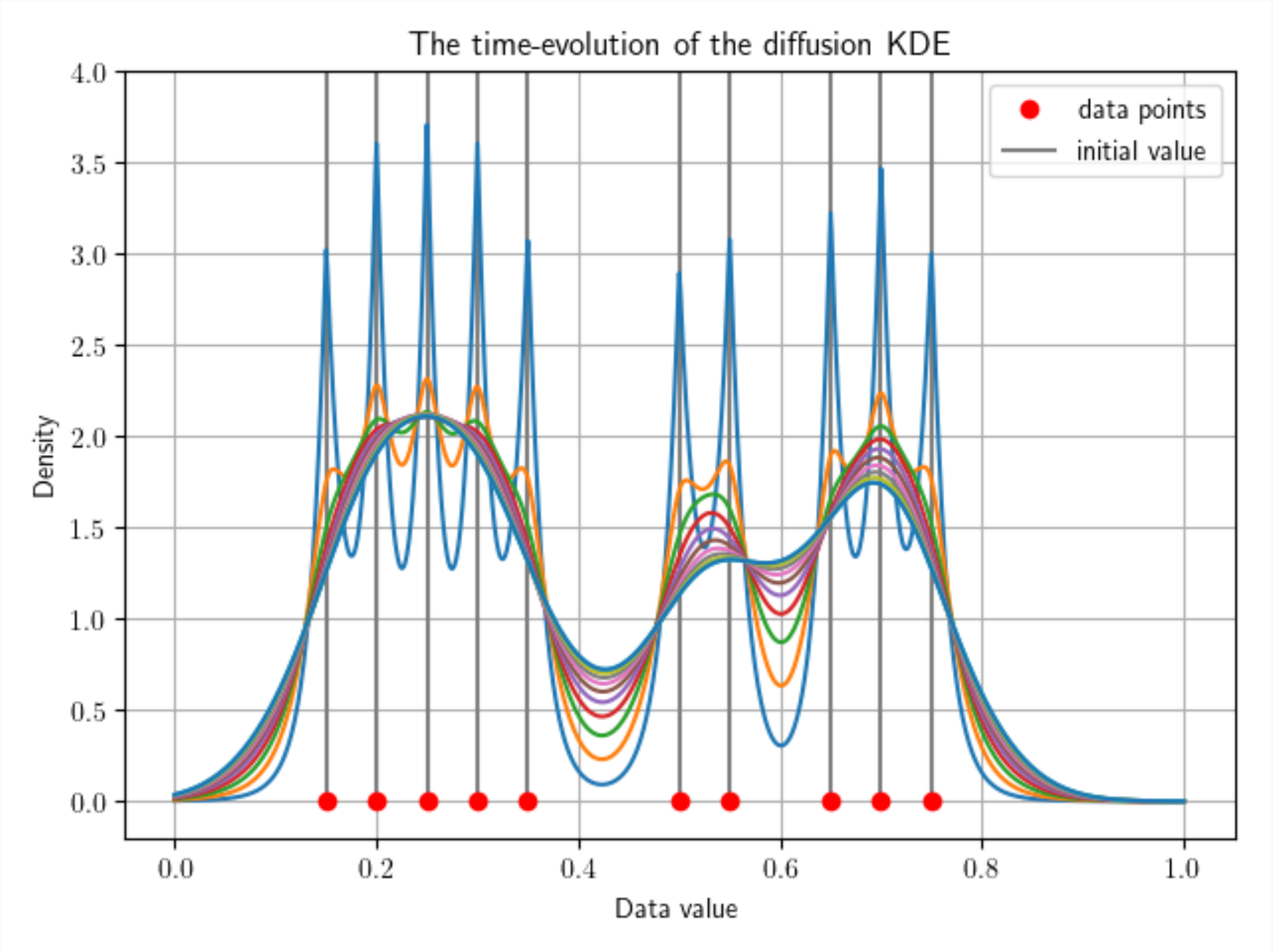
Kernel density estimator (KDE)

$$\hat{f}(x;t) = \frac{1}{n\sqrt{t}} \sum_{i=1}^{n} K\left(\frac{x - X_i}{\sqrt{t}}\right)$$


The kernel density estimator and its determining kernels
data points
KDE
Density
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Data value

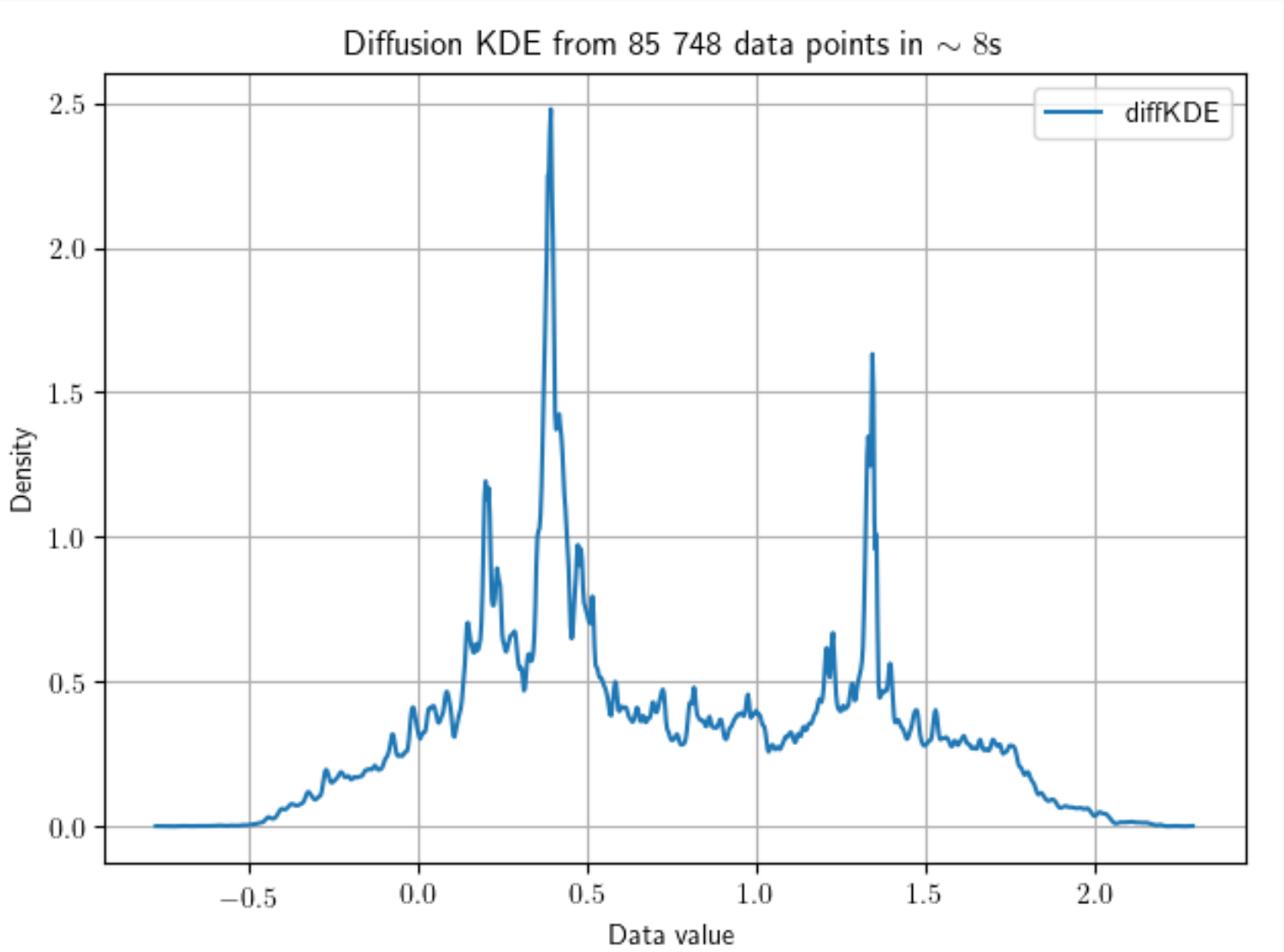
Diffusion KDE

$$\begin{cases} \frac{\partial}{\partial t} g(x;t) = Lg(x;t) \\ g(x;0) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - X_i) \end{cases}$$

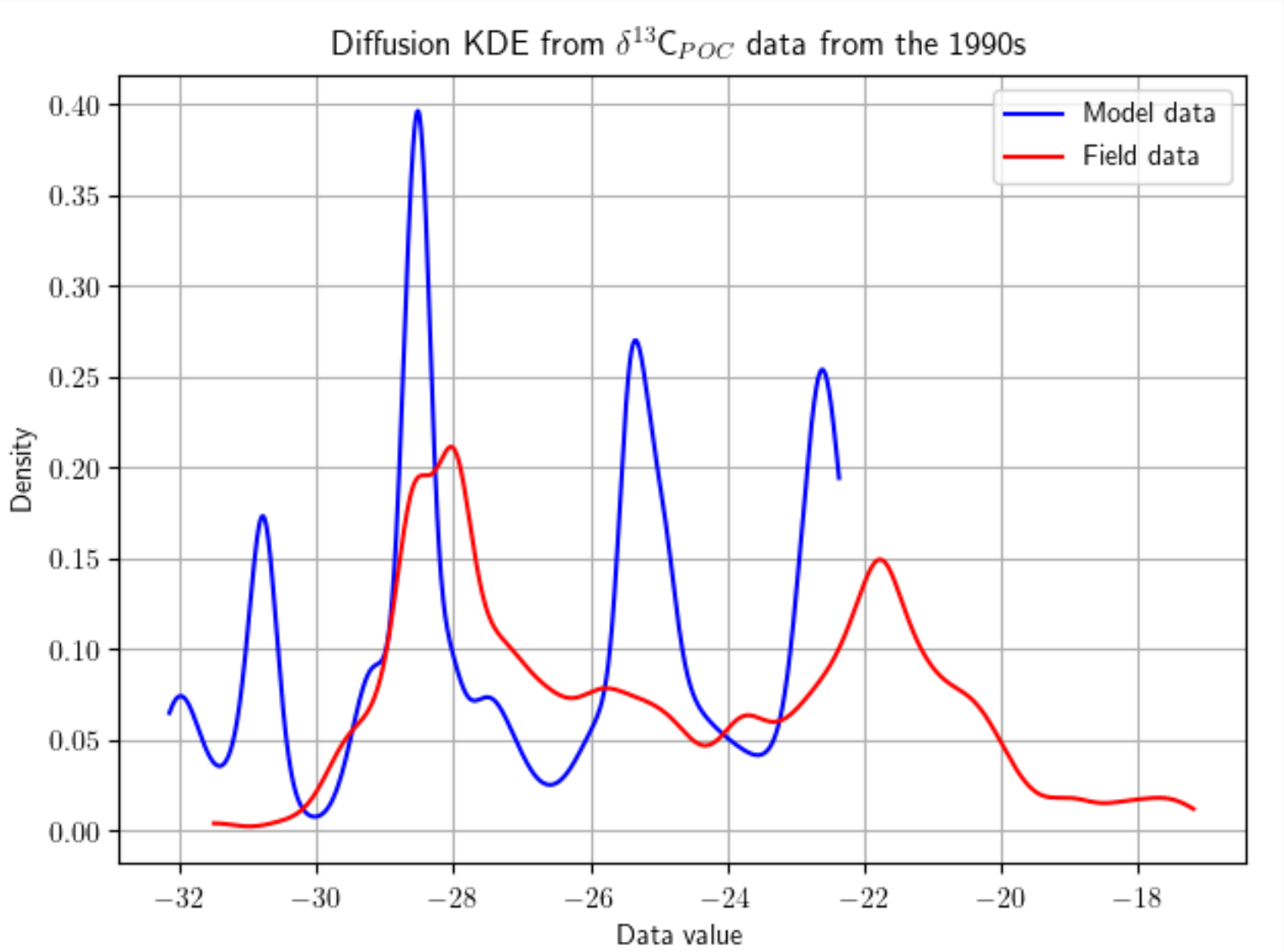

The time-evolution of the diffusion KDE
data points
initial value
Density
Data value
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.0
0.2
0.4
0.6
0.8
1.0

Results

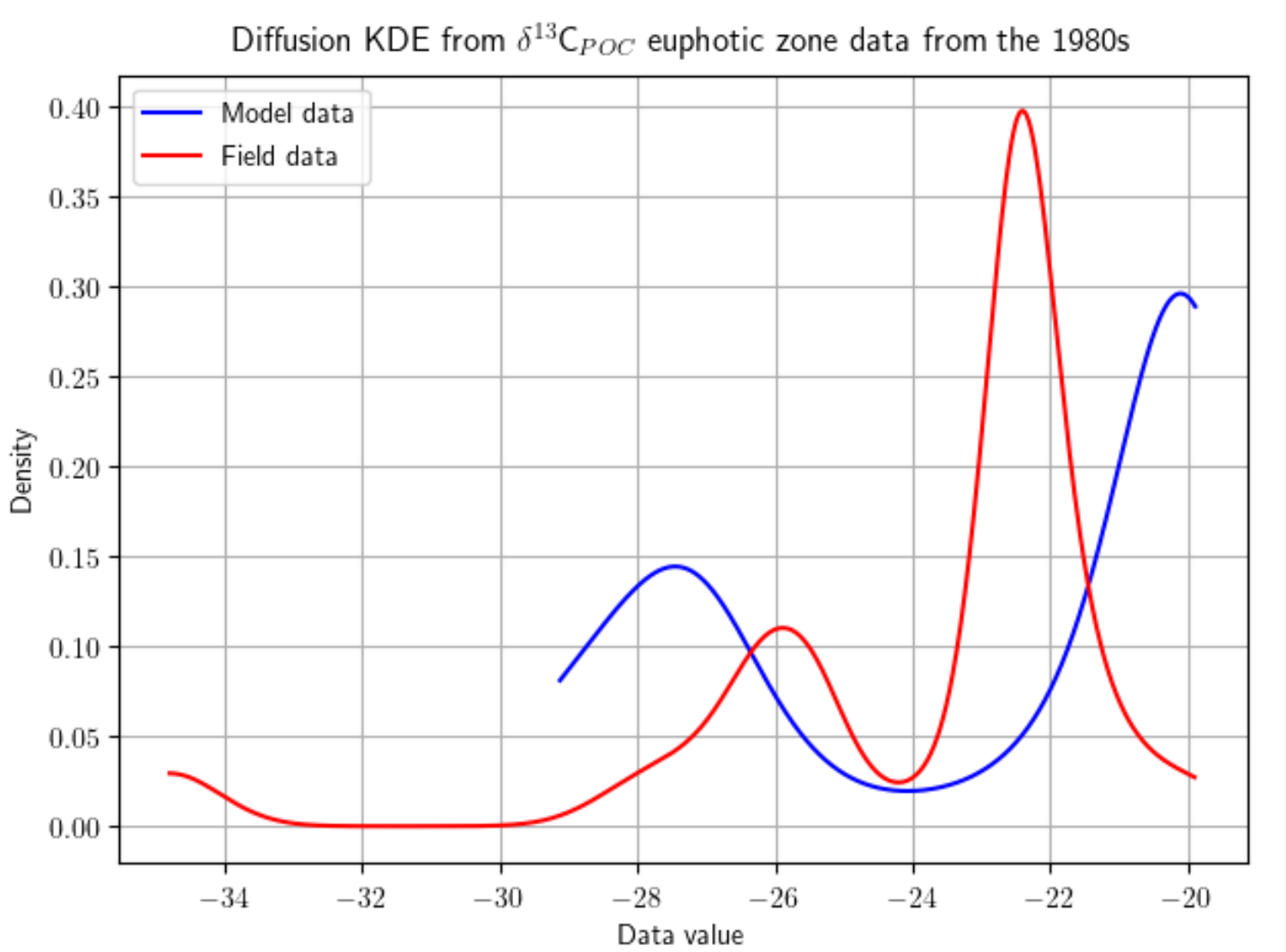
Fast computing



Revelation of bad model performance



Revelation of good model performance



Discussion

Shortcomings of existing KDEs

- boundary bias
- oversmoothing
- long computing time
- noise sensitivity

Outlook

- optimize marine biogeochemical models by the diffKDE
- provide the diffKDE free and open source

Thank You!