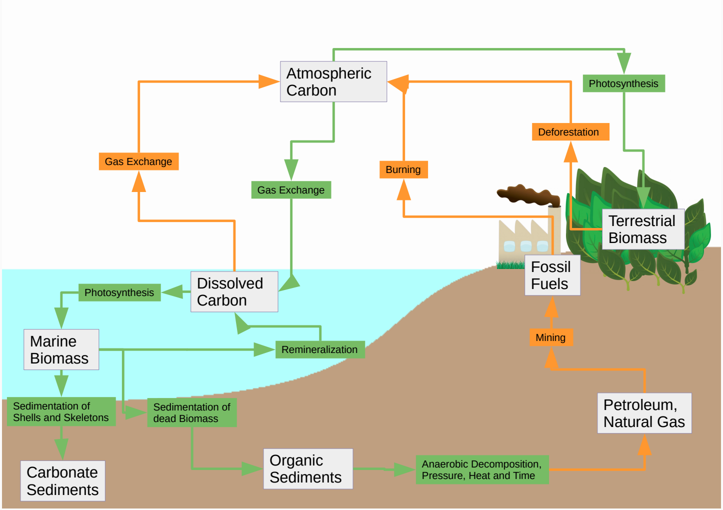
Reliable Assessment of marine biogeochemical Models using diffusion based Kernel Density Estimation

Background

The carbon cycle



Assessing a model

model-data-comparison:

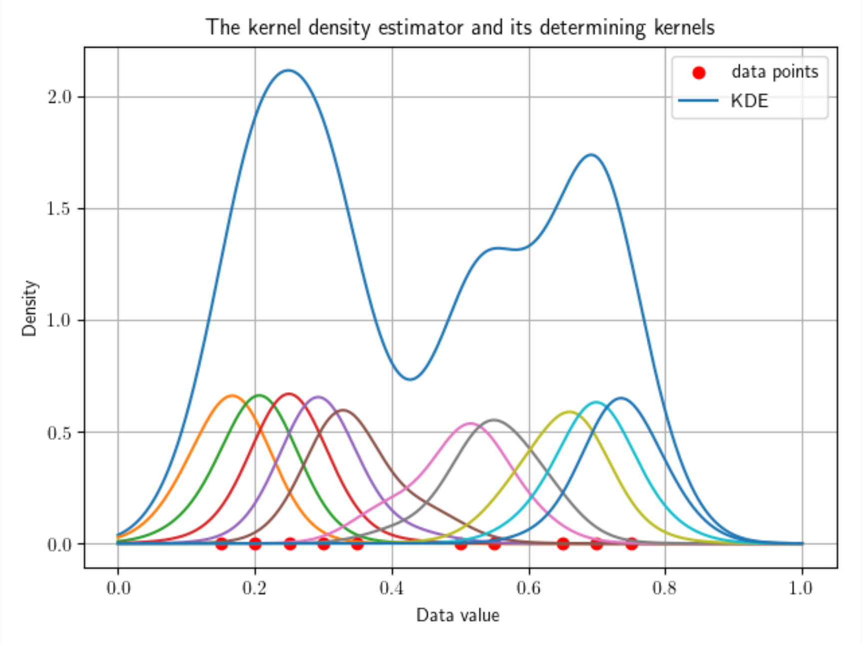
- point by point
- probability density functions

Materials & Methods

Kernel density estimator (KDE)

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

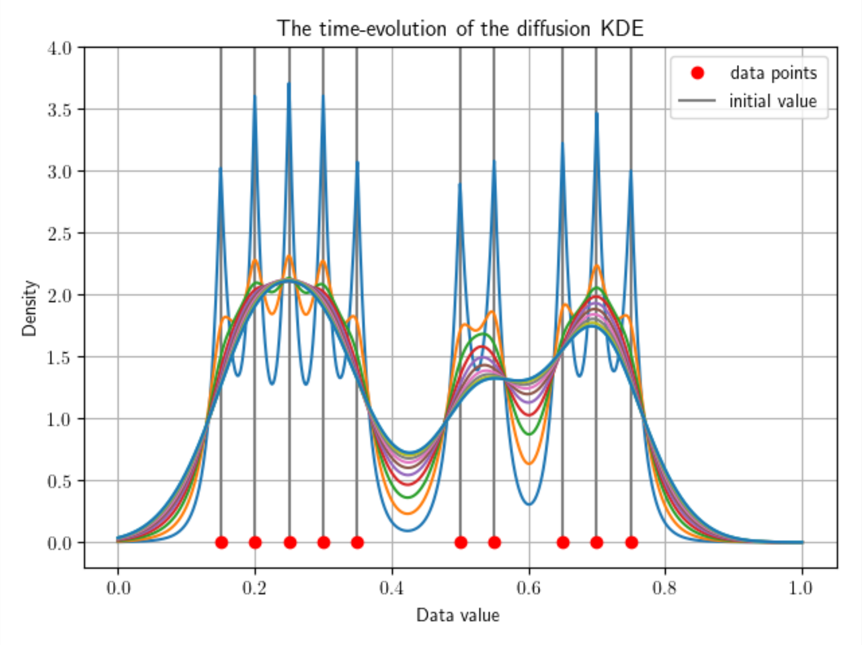
$$x \in \mathbb{R},\ t \in \mathbb{R}_{>0},\ X \in \mathbb{R}^n$$


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

Diffusion KDE

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

$$Lg := \frac{1}{2}\frac{d}{dx}\left(a\frac{d}{dx}\left(\frac{g}{p}\right)\right)$$

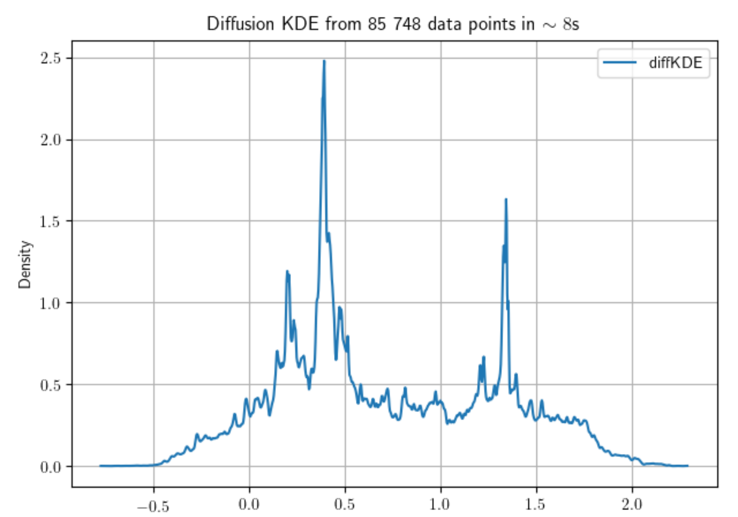

The time-evolution of the diffusion KDE
data points
initial value
Density
Data value
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Implementation based on:

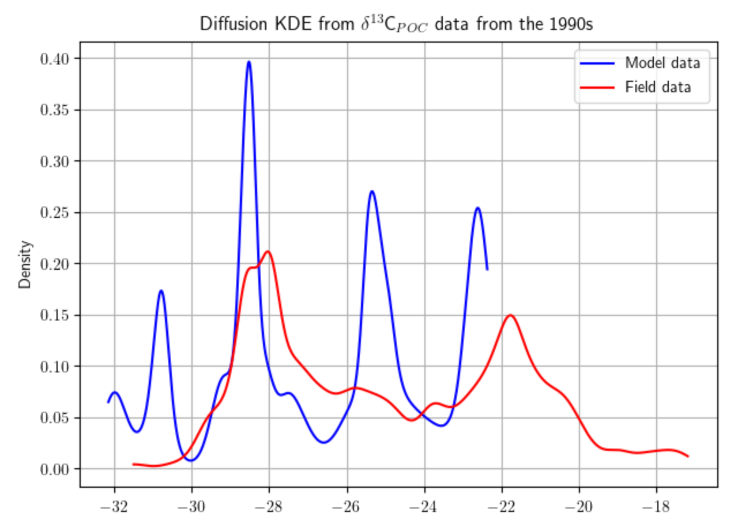
- finite elements (FEniCS)
- plug-in bandwidth selection

Results

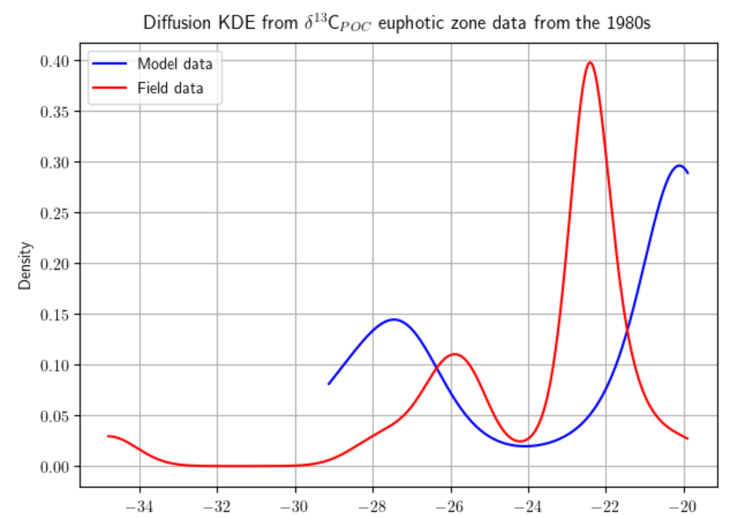
Fast computing



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