Data Assimilation for the Transport Equation

Damon McDougall Wonjung Lee Andrew Stuart

The University of Warwick, Coventry, UK

THE UNIVERSITY OF WARWICK

Summary

Data assimilation is a term widely used to describe the incorporation of observed data into physical systems.

Our aim: To study the effect of model error with regard to data assimilation for the transport equation in the large data limit.

We show **consistency**: No model error present $\Rightarrow$ true signal is recovered.

This is **not robust**: An arbitrarily small model error can lead to inconsistent recovery of the signal in the large data limit.

Numerical results are presented which corroborate this theory.

Problem set-up

Given

$$\text{(PDE)} \quad \frac{\partial v}{\partial t} = c' \cdot \nabla v, \quad (x,t) \in \mathbb{T}^2 \times [0,T],$$
$$\text{(IC)} \quad v(x,0) = u(x),$$

and observations

$$\begin{aligned} y_{j,k} &= v(x_j, t_k) + \eta_{j,k}, \quad \eta_{j,k} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,\sigma^2), \\ \rightsquigarrow \quad y &= \mathcal{G}'(u) + \eta, \quad \eta \sim \mathcal{N}(0,B), \end{aligned} \tag{1}$$

find $u : \mathbb{T}^2 \to \mathbb{R}$ such that $u(x) = v(x,0)$.

This is called an inverse problem.

Data assimilation is act of incorporating y into (PDE) to get u.

Compare this with the forward problem: Find $v(x,t)$ for $t > 0$ given $u(x) = v(x,0)$.

Note: the solution to the forward problem is $v(x,t) = u(x + c't)$.

Solving the inverse problem

Given y, find u. That is, we want to know what $\mathbb{P}(u|y)$ looks like. This distribution is called the posterior ('after' the data) distribution. By Bayes' Rule, we have

$$\begin{aligned} \mathbb{P}(u|y) &= \frac{\mathbb{P}(y|u)\mathbb{P}(u)}{\mathbb{P}(y)} \\ &\propto \mathbb{P}(y|u)\mathbb{P}(u). \end{aligned}$$

What is $\mathbb{P}(y|u)$? It is called the likelihood. We already know it since by Equation (1), given u we have that $y \sim \mathcal{N}(\mathcal{G}'(u), B)$. That is, it is Gaussian.

What is $\mathbb{P}(u)$? It is called the prior ('before' the data) distribution. 'Prior knowledge' means we have an idea of some property of u. For our purposes, we will prescribe u to have $\alpha - 1$ derivatives and Gaussian. I.e., $\mathbb{P}(u) = \mathcal{N}(0, (-\Delta^{-\alpha}))$.

We have a conjugate prior. That is, Gaussian prior $\Rightarrow$ Gaussian posterior.

Data/model mismatch

There are many methods of sampling $\mathbb{P}(u|y)$, each needing an implementation of $\mathcal{G}'$ and each with their own pros and cons.

Question: What happens when the $\mathcal{G}$ you use to explore $\mathbb{P}(u|y)$ is not the same as the 'true' $\mathcal{G}'$ that generated y?

This happens all the time in climate science and weather prediction.

Our case: $\mathcal{G}$ also corresponds to an advection equation but possibly with a different wave velocity, c.

Define: $\delta c := c - c'$.

The results section that follows looks at $\mathbb{E}(u|y)$ for various cases of δc with initial condition (shown in Figure 1)

$$u(x,y) = \sum_{j=1}^{3} \sin(2\pi j x) + \sum_{k=1}^{3} \cos(2\pi k y).$$

Here $\mathbb{E}(u|y)$ has been computed from 10^6 samples from a random-walk Metropolis-Hasting Markov chain that samples $\mathbb{P}(u|y)$.

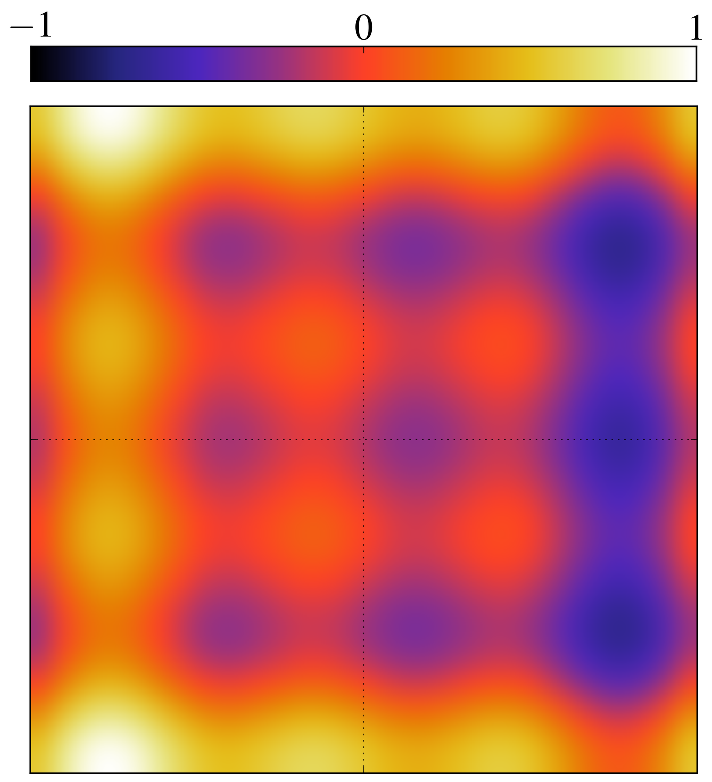


Figure: $u(x,y)$

Results: $\delta c = (0,0)$

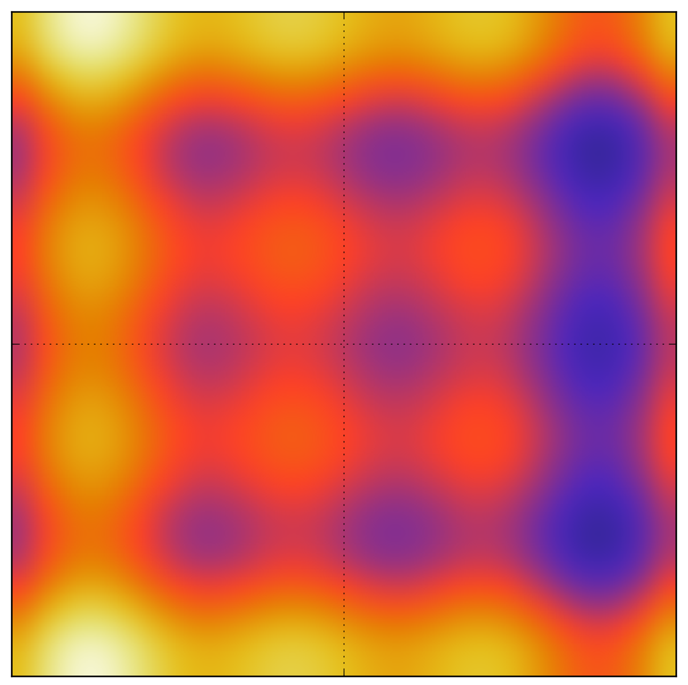
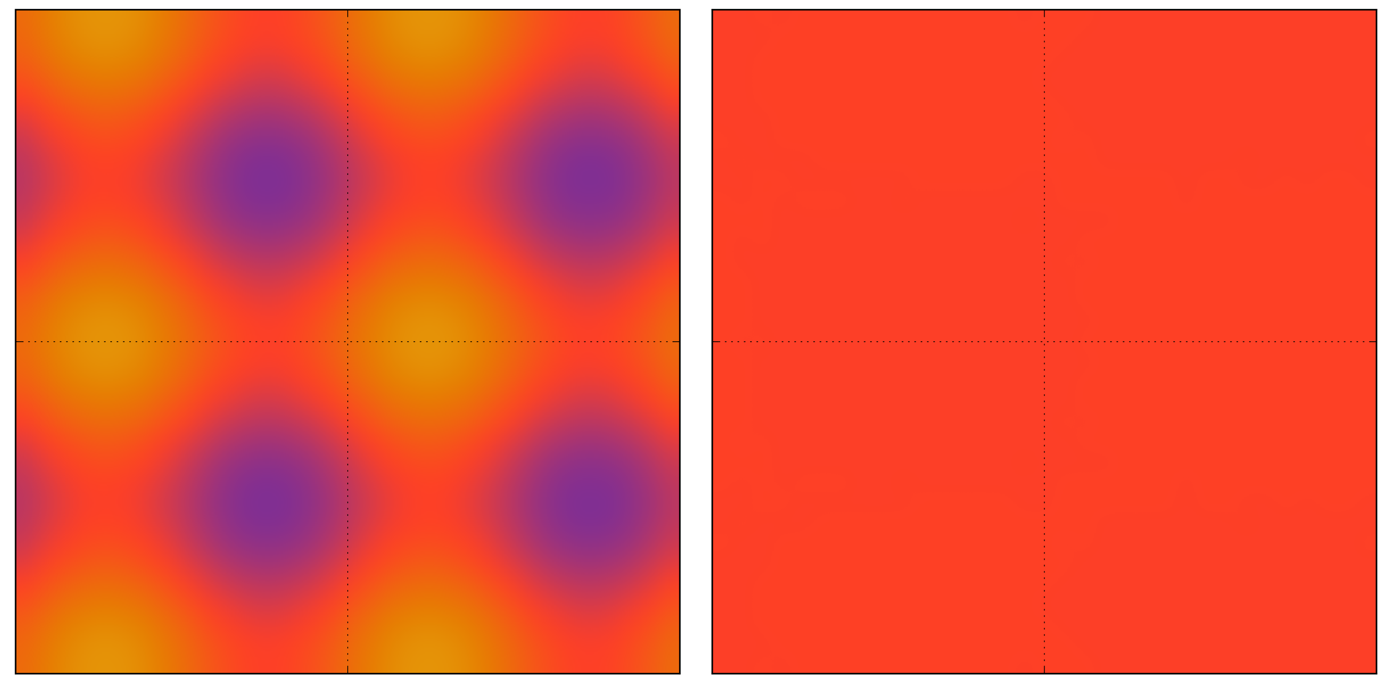


Figure: $\delta c = (0,0)$

Observation: With no model error, we recover the truth.

Results: δc time-independent

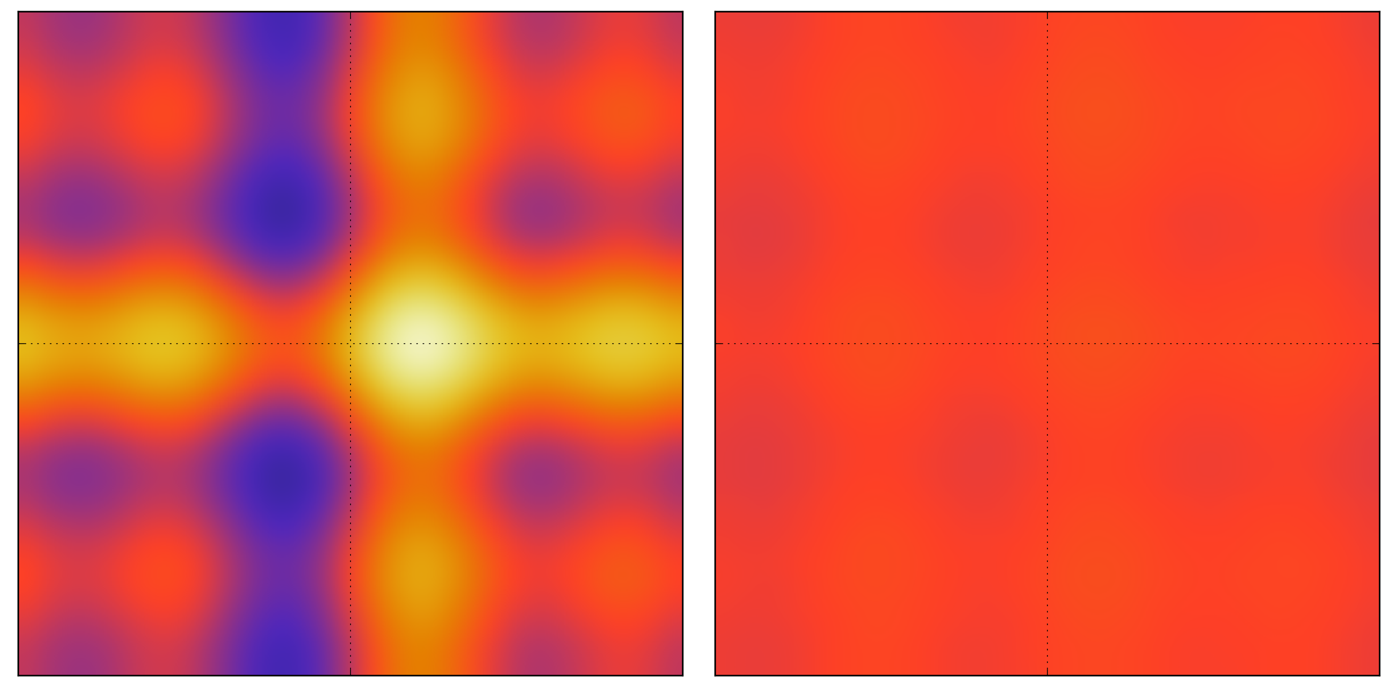


(a) $\delta c = (0.5, 0.5)$ (b) $\delta c \in \mathbb{R} \setminus \mathbb{Q} \times \mathbb{R} \setminus \mathbb{Q}$

Figure: Time-independent δc

Observation: We recover only part of the Fourier series of u.

Results: δc time-dependent



(a) $\int_0^\infty \delta c \, dt = (0.5, 0.5)$ (b) $\delta c = \dot{W} - c$

Figure: Time-dependent δc

Observation: If $\lim_{t\to\infty} \delta c = 0$ sufficiently fast, we recover a shift of u.

Conclusions and further work

- Even in the linear deterministic case, getting the model slightly wrong can have striking results.
- Each sample average used 10^6 samples however the posterior is Gaussian, so the Kalman filter is exact (but high dimensional)!
- Random-walk Metropolis-Hastings sampler is easily extended to $\mathbb{P}(u, c|y)$. This is very non-Gaussian.
- Advection is one aspect of weather and is very 'stiff' with respect to c.

The theory for $y_k = v(t_k) + \eta_k$ (complete observation of the field) exists in:

D. McDougall, W. Lee, A. M. Stuart.
Kalman filtering for linear wave equations with model error.
in preparation

 D.McDougall@warwick.ac.uk