BRISTAU: A State-Space Model for Non-Gaussian Hierarchical Emissions Inference

Stephen Pearson[a], Anita Ganesan[a], Matt Rigby[b], Luke Western[c], Andrew Zammit-Mangion[d]

[a] UoB, School of Geographical Sciences | [b] UoB, School of Chemistry | [c] MIT, Center for Sustainability Science and Strategy | [d] UNSW, School of Mathematics and Statistics

Scientific context

- Atmospheric measurements are increasingly used to constrain national emissions inventories
- The goal is to estimate regional surface fluxes, $\boldsymbol{x}$, from atmospheric observations, $\boldsymbol{y}$
- A chemical transport model (CTM) calculates a transport operator, linking surface sources to observations
- NAME[1] is a widely used example (a Lagrangian CTM)
- A statistical inverse framework is needed to estimate $\boldsymbol{x}$ while accounting for uncertainties and related parameters
- Markov chain Monte Carlo (MCMC) approaches have become the gold standard in this field

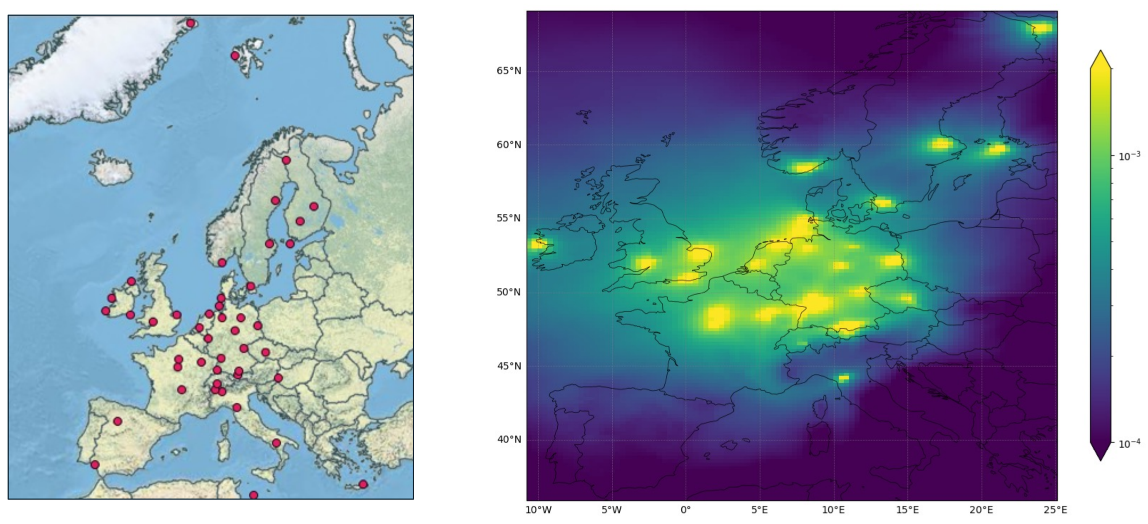


Fig.1. *Left: ICOS Atmosphere network[2], providing high frequency in-situ measurements of greenhouse gas concentrations. Right: map demonstrating the mean 2022 network measurement sensitivity to changes in surface emissions, as modelled by NAME[1].*

The problem

- Non-Gaussian methods needed to estimate non-CO_2 emissions
- Model hierarchy allows for estimation of "uncertainties in uncertainties", leading to more complete estimates[3]
- Traditional MCMCs are flexible, but computationally expensive
- MCMC emissions assumed **temporally invariant**
- Dimension reduction leads to **spatially coarse estimates**
- They rely on "batch" data usage – **difficult to scale**
- Large remotely sensed datasets under utilised

Model overview

$$\text{(O)}\;\; \boldsymbol{y}_t = \boldsymbol{H}_t\boldsymbol{z}_t + \boldsymbol{\epsilon}_t \qquad \text{(S)}\;\; \tilde{\boldsymbol{z}}_t = \boldsymbol{F}\tilde{\boldsymbol{z}}_{t-1} + \tilde{\boldsymbol{\omega}}_t$$

$$\boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}) \qquad \tilde{\boldsymbol{\omega}}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$$

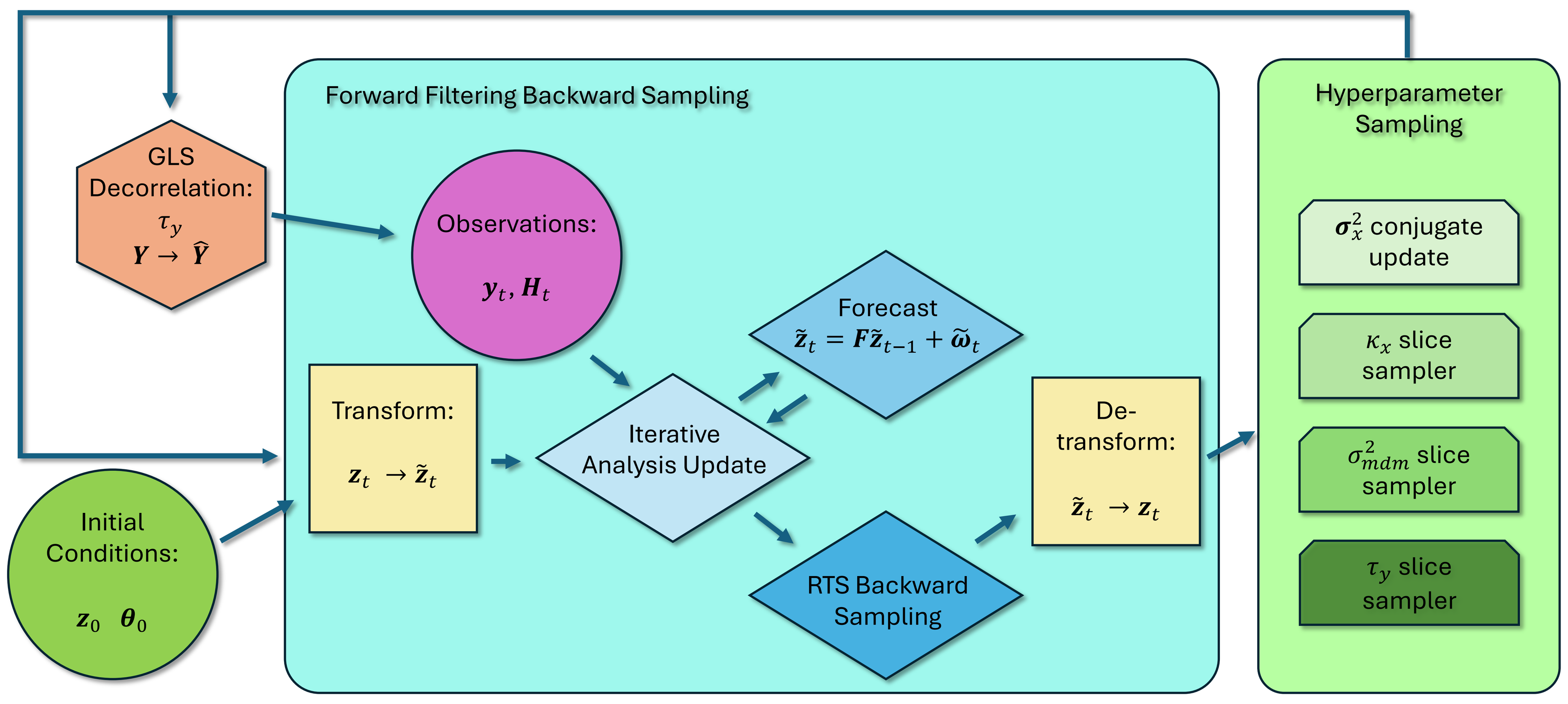
- This framework follows the mixed Gaussian-lognormal Kalman filter (MXKF)[4]
- $\boldsymbol{z}_t = [\boldsymbol{x}_t, \boldsymbol{x}_{bc}, \boldsymbol{r}_t]^T$ and $\tilde{\boldsymbol{z}}_t$ is the fully Gaussian transformed state
- $\boldsymbol{x}_t$ is the spatially aggregated surface flux basis functions
- $\boldsymbol{\omega}_t$ assumed lognormal for all basis functions
- $\boldsymbol{F}$ enforces VAR(1) relaxation of $\boldsymbol{x}_t$ towards $\boldsymbol{r}_t$ with

$$\tilde{x}_{t,k} = \kappa_x \tilde{x}_{t-1,k} + (1-\kappa_x)\tilde{r}_{t-1}, \qquad \kappa_x \in [0,1]$$

- $\boldsymbol{R}$ contains diagonal $(\sigma^2_{obs,t} + \sigma^2_{mdm})$ and off diagonal terms governed by decorrelation timescale τ_y
- $\boldsymbol{\epsilon}_t$ assumed to follow AR(1) (OU) process
- $\boldsymbol{Q} = diag([\boldsymbol{\sigma}_x^2, \boldsymbol{\sigma}_{bc}^2, \boldsymbol{\sigma}_r^2])$

Hyperparameter	Description	Prior
σ^2_{mdm}	MDM error variance	*Scaled Beta*
τ_y	MDM decorrelation timescale	*Scaled Beta*
σ^2_x	Forecast error variance	*Inverse Gamma*
κ_x	State persistence	*Uniform*

The algorithm



- At a high-level, this is just a conditional Gibbs sampler
- Latent state forward pass uses the MXKF, with RTS style backward sampling
- The analysis step utilises an **accelerated Newton-Gauss procedure** to optimise linearisation of the observation operator
- GLS decorrelation allows likelihood covariance to be treated as diagonal

OSSE

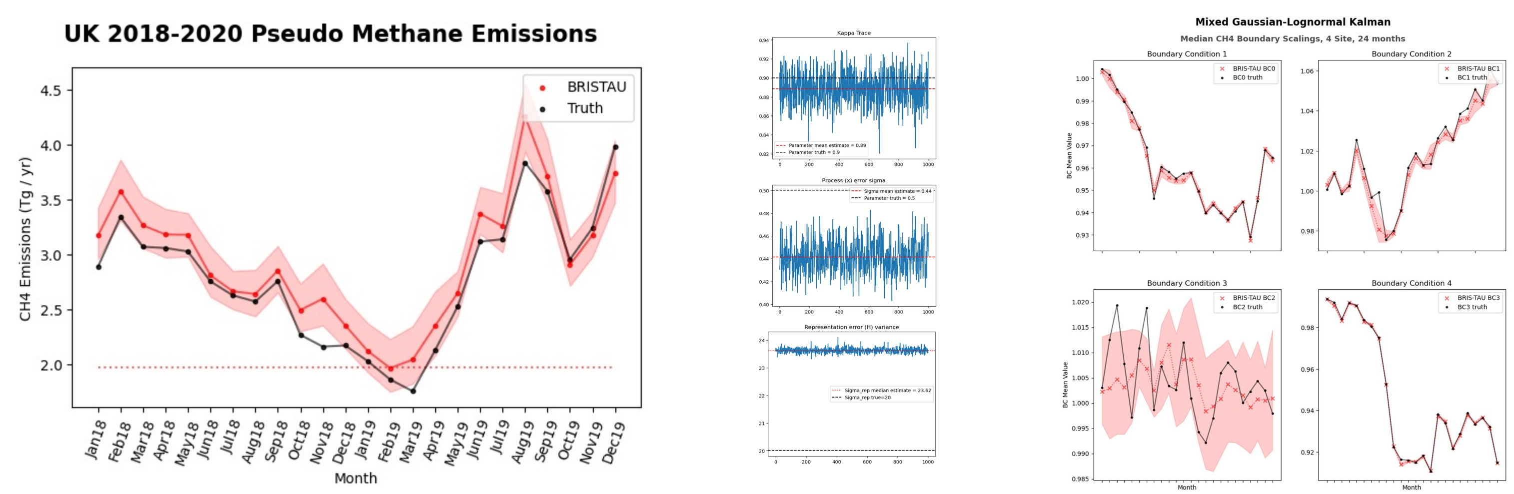


Fig.2. *Left: the fit to truth of UK aggregated pseudo-emissions over a 2-year period. Middle: the fit to truth of the hyperparameters – notice the identifiability issue between the 2 error variance terms (inverse correlation). Right: the fit to truth of the boundary conditions.*

- With a state-space formulation, BRISTAU can run over longer time windows than standard MCMCs

Real data

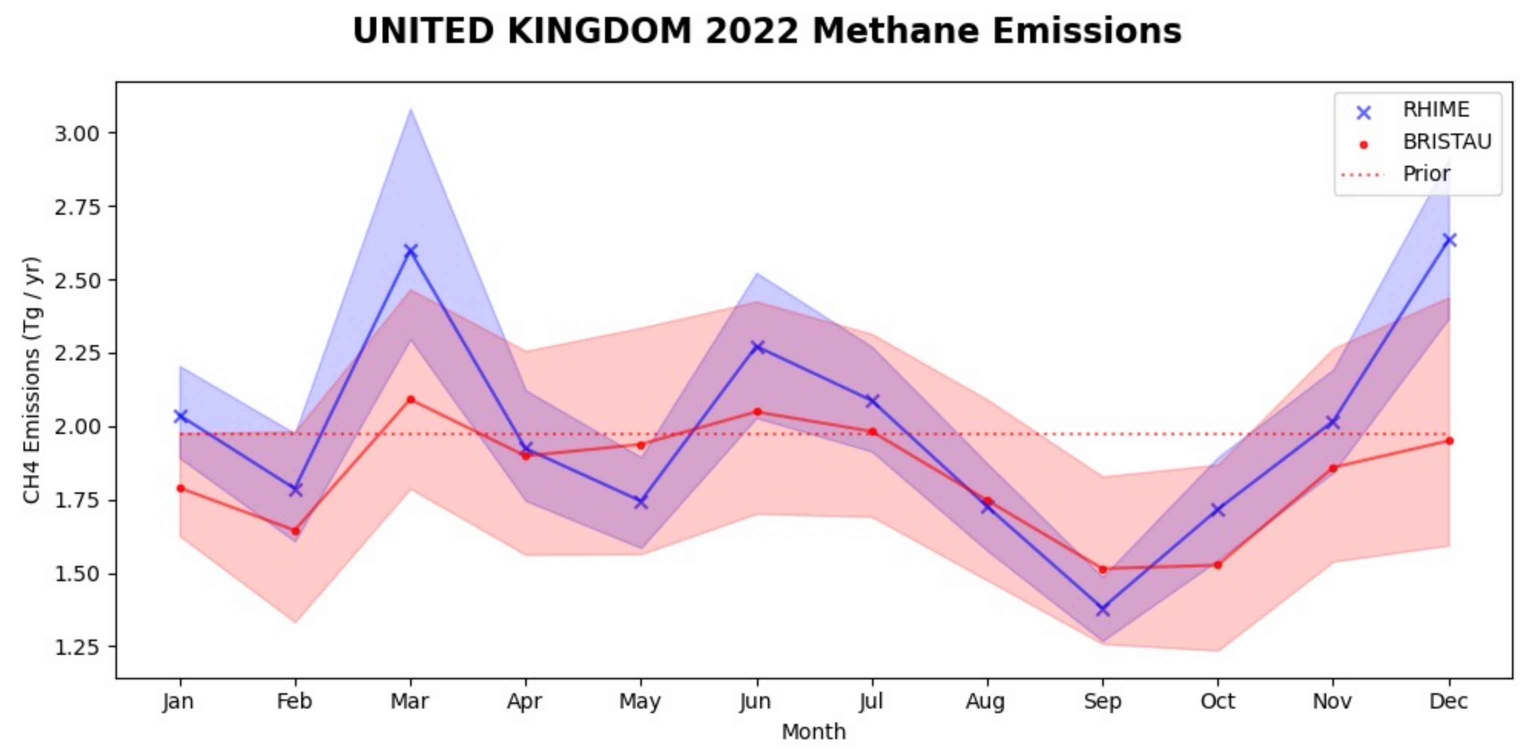


Fig.3. *Comparison of UK emission estimates from RHIME – a NUTS hierarchical MCMC – and BRISTAU. RHIME assumes complete temporal independence of both monthly emission estimates and observation errors.*

- This inversion used hourly data from 28 sites, ~220,000 observations ($\boldsymbol{Y}$), with 250 basis functions ($\boldsymbol{x}$)
- BRISTAU sampling time ~26 mins
- RHIME (MCMC) total sampling time ~5.5 hrs (~340 mins)

Get in touch!

If you have any suggestions or would like a chat, please drop me a line - ***s.pearson@bristol.ac.uk***

University of Bristol

Research England

[1] Jones et al. (2007) Met Office's Next-Generation Atmospheric Dispersion Model, NAME III. Doi: 10.1007/978-0-387-68854-1_62
[2] ICOS. (n.d.). *Atmosphere stations*. Integrated Carbon Observation System. Available from: https://www.icos-cp.eu/observations/atmosphere/stations
[3] Ganesan et al. (2014) Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods. Doi: 10.5194/acp-14-3855-2014
[4] Fletcher et al. (2023) Lognormal and Mixed Gaussian–Lognormal Kalman Filters. *Monthly Weather Review.* Doi: 10.1175/MWR-D-22-0072.1