Causal Spatial and Spatio-Temporal Models

"An application to flux estimation"

Andrew Zammit-Mangion
and
Noel Cressie

National Institute for Applied Statistics Research Australia
University of Wollongong

Outline

Section 1

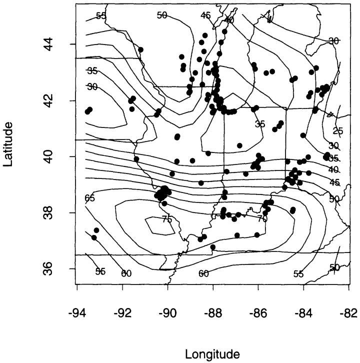
Introduction

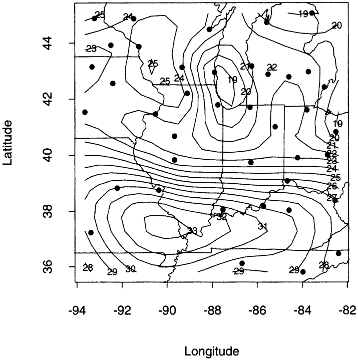
- **Univariate** spatial model.

- **Multivariate** spatial model.
  - Two or more interacting spatial variables.
  - Improve prediction on one of the variates by observing the others: **Cokriging**.
  - Determine which variate caused the observed phenomenon: **Source separation**.

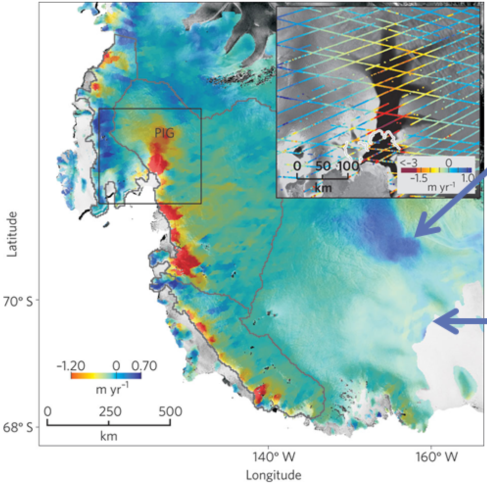
NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

Royle and Berliner (1999), Midwestern USA.

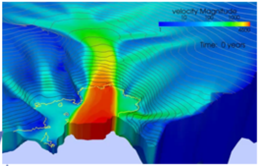


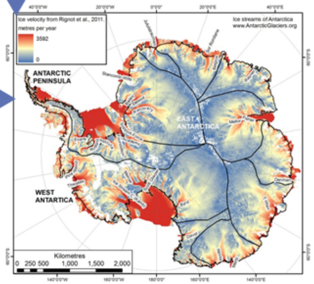


ICESAT data (elevation change per year)

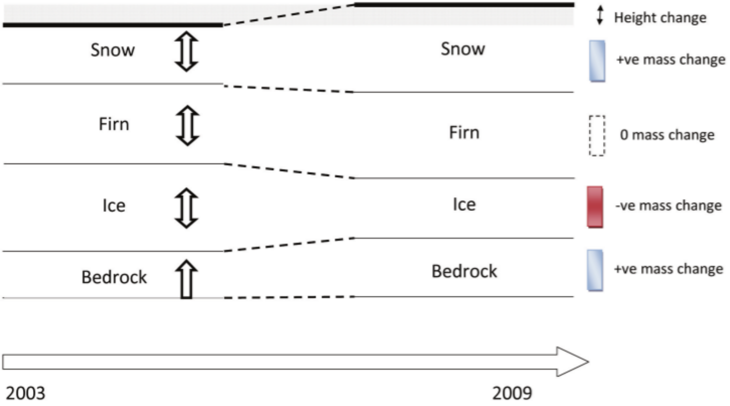


Elmer/Ice (LGGE) model





- Zammit-Mangion et al. (2014, 2015b,a), Antarctica.



NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

- **Modelling**: Given a bivariate process $(Y_1(\cdot), Y_2(\cdot))$, what is a valid *cross-covariance function matrix* (CCFM)

$$\begin{pmatrix} C_{11}(\cdot,\cdot) & C_{12}(\cdot,\cdot) \\ C_{21}(\cdot,\cdot) & C_{22}(\cdot,\cdot) \end{pmatrix}, \tag{1}$$

such that **any** covariance matrix derived from it is positive-definite?

- **Computational**: Sometimes we struggle with univariate models – how do our algorithms scale to multivariate models?

- **Linear model of co-regionalisation** (LMC, Wackernagel, 1995): Define

$$Y_1(\cdot) = a_{11}\widetilde{Y}_1(\cdot) + a_{12}\widetilde{Y}_2(\cdot), \quad (2)$$

$$Y_2(\cdot) = a_{21}\widetilde{Y}_1(\cdot) + a_{22}\widetilde{Y}_2(\cdot), \quad (3)$$

where, independently,

$$\widetilde{Y}_1(\cdot) \sim \mathcal{N}(\mu_1(\cdot), C_1(\cdot,\cdot)), \quad (4)$$

$$\widetilde{Y}_2(\cdot) \sim \mathcal{N}(\mu_2(\cdot), C_2(\cdot,\cdot)). \quad (5)$$

- $C_{ij}(\cdot,\cdot) = a_{i1}a_{j1}C_1(\cdot,\cdot) + a_{i2}a_{j2}C_2(\cdot,\cdot)$.
- CCFM is positive-definite for any $\{a_{ij} : i,j = 1,\ldots,2\}$.

- **Bivariate parsimonious Matérn model** (Gneiting et al., 2010): Let $C^o(\cdot)$ be a stationary, isotropic covariance function. Define

$$C^o_{ij}(\cdot) \equiv \beta_{ij} M(\cdot; \nu_{ij}, \kappa_{ij}), \tag{6}$$

  where $M(\cdot)$ is a Matérn covariance function. Let $\kappa_{ii} = \kappa_{ij} = \kappa$ and set $\nu_{ij} = (\nu_{ii} + \nu_{jj})/2$. Then if $(\beta_{ij} : i, j = 1, 2)$ is positive-definite, the CCFM is positive-definite.

- **Bivariate full Matérn model**: Relaxes assumptions on smoothness and scales, but finding valid parameters is much more involved.

Current approaches (limitations)

- Stuck with homogeneous models (e.g., convolution methods).
- Stuck with fixed scales (parsimonious Matérn).
- Stuck with Matérn models (e.g., full Matérn models).
- **Stuck with symmetry (e.g., LMC)**.

- $Y_1(\cdot)$: precipitation at present.
- $Y_2(\cdot)$: precipitation in 5 minutes time.

Section 2

Causal spatial models

Specification:

$$E\left(Y_2(\mathbf{s}) \mid Y_1(\cdot)\right) = \int_D b(\mathbf{s},\mathbf{v})\, Y_1(\mathbf{v})\,\mathrm{d}\mathbf{v}; \quad \mathbf{s} \in D, \tag{7}$$

$$\operatorname{cov}\left(Y_2(\mathbf{s}), Y_2(\mathbf{u}) \mid Y_1(\cdot)\right) = C_{2|1}(\mathbf{s},\mathbf{u}); \quad \mathbf{s},\mathbf{u} \in \mathbb{R}^d. \tag{8}$$

Building blocks:

- $C_{11}(\cdot,\cdot)$,
- $C_{2|1}(\cdot,\cdot)$,
- $b(\cdot,\cdot)$ (interaction function).

- CCFM is easy to find:

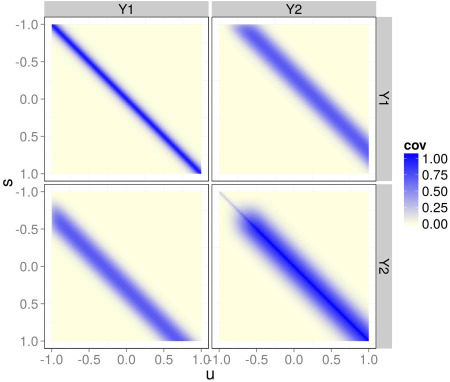
$$\begin{bmatrix} C_{11}(\mathbf{s},\mathbf{u}) & \int_D C_{11}(\mathbf{s},\mathbf{v})b(\mathbf{u},\mathbf{v})\mathrm{d}\mathbf{v} \\ \int_D b(\mathbf{s},\mathbf{v})C_{11}(\mathbf{v},\mathbf{u})\mathrm{d}\mathbf{v} & C_{22}(\mathbf{s},\mathbf{u}) \end{bmatrix}; \tag{9}$$

$$C_{22}(\mathbf{s},\mathbf{u}) = C_{2|1}(\mathbf{s},\mathbf{u}) + \int_D \int_D b(\mathbf{s},\mathbf{v})C_{11}(\mathbf{v},\mathbf{w})b(\mathbf{w},\mathbf{u})\mathrm{d}\mathbf{v}\mathrm{d}\mathbf{w}, \tag{10}$$

  and is always valid (we will outline the proof soon).

- Asymmetry (i.e., $C_{12}(\mathbf{s},\mathbf{u}) \neq C_{21}(\mathbf{s},\mathbf{u})$) is guaranteed if $b(\cdot,\cdot)$ is not symmetric.

- Assume $b^o(\cdot) = b(\cdot, \cdot)$ and that it is off-centre.
- $\mathbf{s}, \mathbf{u} \in \{-1, -0.9, \ldots, 1\}$.



Properties of causal spatial models

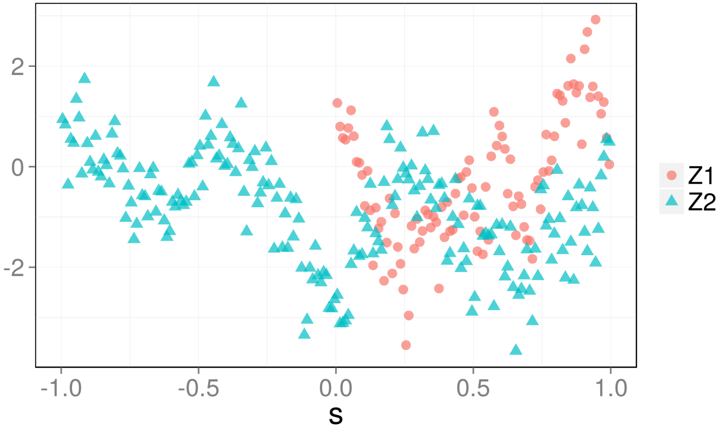
- Heterogeneity, since $C_{11}(\cdot,\cdot)$, $C_{2|1}(\cdot,\cdot)$ need not be homogeneous and $b(\mathbf{s},\mathbf{u})$ need not be symmetric.
- We are not restricted to Matérn fields. The bivariate parsimonious Matérn field is a **special case**.
- $Y_2(\cdot)$ can be arbitrarily smoother than $Y_1(\cdot)$ **and** have a different scale.

- Assume all parameters are known and $Y_1(\cdot)$ is only partially observed.

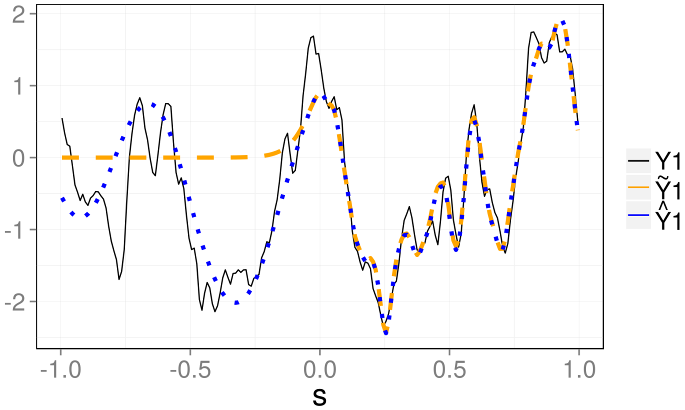
- Use simple cokriging **or** simple kriging to estimate $Y_1(\cdot)$:

$$\hat{Y}_1(\mathbf{s}_0) \equiv E(Y_1(\mathbf{s}_0) \mid \mathbf{Z}_1, \mathbf{Z}_2) \quad \text{simple cokriging predictor,}$$
$$\widetilde{Y}_1(\mathbf{s}_0) \equiv E(Y_1(\mathbf{s}_0) \mid \mathbf{Z}_1) \quad \text{simple kriging predictor.}$$

Example



Example



Is the bivariate model always valid?

- If $C_{11}(\mathbf{s},\mathbf{u})$ and $C_{2|1}(\mathbf{s},\mathbf{u})$ are positive-definite, then $C_{22}(\cdot,\cdot)$ is positive-definite (recall quadratic form).
- $C_{12}(\mathbf{s},\mathbf{u}) = C_{21}(\mathbf{u},\mathbf{s})$.
- CCFM is positive-definite if, for any n_1, n_2 such that $n_1 + n_2 > 0$, any locations $\{\mathbf{s}_{1k}\}, \{\mathbf{s}_{2l}\}$ and any real numbers $\{a_{1k}\}, \{a_{2l}\}$,

$$
\begin{aligned}
&\mathrm{var}\left(\sum_{k=1}^{n_1} a_{1k} Y_1^0(\mathbf{s}_{1k}) + \sum_{l=1}^{n_2} a_{2l} Y_2^0(\mathbf{s}_{2l})\right) \\
&= \sum_{k=1}^{n_1}\sum_{k'=1}^{n_1} a_{1k}a_{1k'} C_{11}^0(\mathbf{s}_{1k},\mathbf{s}_{1k'}) + \sum_{l=1}^{n_2}\sum_{l'=1}^{n_2} a_{2l}a_{2l'} C_{22}^0(\mathbf{s}_{2l},\mathbf{s}_{2l'}) \\
&+ \sum_{k=1}^{n_1}\sum_{l'=1}^{n_2} a_{1k}a_{2l'} C_{12}^0(\mathbf{s}_{1k},\mathbf{s}_{2l'}) + \sum_{l=1}^{n_2}\sum_{k'=1}^{n_1} a_{2l}a_{1k'} C_{21}^0(\mathbf{s}_{2l},\mathbf{s}_{1k'}) \;\geq 0.
\end{aligned}
$$

- It can be shown that

$$\mathrm{var}\left(\sum_{k=1}^{n_1} a_{1k} Y_1^0(\mathbf{s}_{1k}) + \sum_{l=1}^{n_2} a_{2l} Y_2^0(\mathbf{s}_{2l})\right)$$
$$= \sum_{l=1}^{n_2}\sum_{l'=1}^{n_2} a_{2l}a_{2l'} C_{2|1}(\mathbf{s}_{2l}, \mathbf{s}_{2l'}) + \int_D \int_D a(\mathbf{s})a(\mathbf{u})C_{11}(\mathbf{s}, \mathbf{u})\, \mathrm{d}\mathbf{s}\mathrm{d}\mathbf{u},$$

where

$$a(\mathbf{s}) \equiv \sum_{k=1}^{n_1} a_{1k}\delta(\mathbf{s} - \mathbf{s}_{1k}) + \sum_{l=1}^{n_2} a_{2l} b(\mathbf{s}_{2l}, \mathbf{s}); \quad \mathbf{s} \in \mathbb{R}^d.$$

- $[Y_1(\cdot), \ldots, Y_p(\cdot)]$ can be decomposed as,

$$[Y_p(\cdot) \mid Y_{p-1}(\cdot), Y_{p-2}(\cdot), \ldots, Y_1(\cdot)] \ldots [Y_1(\cdot)].$$

NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

- $[Y_1(\cdot), \ldots, Y_p(\cdot)]$ can be decomposed as,

$$[Y_p(\cdot) \mid Y_{p-1}(\cdot), Y_{p-2}(\cdot), \ldots, Y_1(\cdot)] \ldots [Y_1(\cdot)].$$

- The conditional expectation is

$$E(Y_q(\mathbf{s}) \mid \{Y_r(\cdot) : r = 1, \ldots, (q-1)\}) \equiv \sum_{r=1}^{q-1} \int_D b_{qr}(\mathbf{s}, \mathbf{v}) Y_r(\mathbf{v}) \mathrm{d}\mathbf{v}; \quad \mathbf{s} \in D.$$

- $[Y_1(\cdot), \ldots, Y_p(\cdot)]$ can be decomposed as,

$$[Y_p(\cdot) \mid Y_{p-1}(\cdot), Y_{p-2}(\cdot), \ldots, Y_1(\cdot)] \ldots [Y_1(\cdot)].$$

- The conditional expectation is

$$E(Y_q(\mathbf{s}) \mid \{Y_r(\cdot) : r = 1, \ldots, (q-1)\}) \equiv \sum_{r=1}^{q-1} \int_D b_{qr}(\mathbf{s}, \mathbf{v}) Y_r(\mathbf{v}) \mathrm{d}\mathbf{v}; \quad \mathbf{s} \in D.$$

- The conditional covariance is

$$\operatorname{cov}(Y_q(\mathbf{s}), Y_q(\mathbf{u}) \mid \{Y_r(\cdot) : r = 1, \ldots, (q-1)\}) \equiv C_{q|(r<q)}(\mathbf{s}, \mathbf{u}); \quad \mathbf{s}, \mathbf{u} \in \mathbb{R}^d,$$

where $\{b_{qr}(\cdot, \cdot) : r = 1, \ldots, (q-1);\ q = 2, \ldots, p\}$ are integrable.

We need to show that the p-variate process is well defined. The proof is by induction:

- We know that the bivariate process is well defined.
- Assume that the $(p-1)$-variate process is well defined.
- Show that the p-variate process is well defined.

$$
\begin{aligned}
\mathrm{var}\left(\sum_{q=1}^{p}\sum_{m=1}^{n_q} a_{qm} Y_q(\mathbf{s}_{qm})\right) &= \sum_{m=1}^{n_p}\sum_{m'=1}^{n_p} a_{pm}a_{pm'}C_{p|(q<p)}(\mathbf{s}_{pm},\mathbf{s}_{pm'}) \\
&+ \sum_{q=1}^{p-1}\sum_{r=1}^{p-1}\int_D\int_D a_q(\mathbf{s})a_r(\mathbf{u})C_{qr}(\mathbf{s},\mathbf{u})\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{u},
\end{aligned}
$$

where

$$
a_q(\mathbf{s}) \equiv \left(\sum_{k=1}^{n_q} a_{qk}\delta(\mathbf{s}-\mathbf{s}_{qk}) + \sum_{m=1}^{n_p} a_{pm}b_{pq}(\mathbf{s}_{pm},\mathbf{s})\right).
$$

The following can all be shown to be special cases of causal spatial models:

- The parsimonious Matérn model of Gneiting et al. (2010),
- The full Matérn model of Gneiting et al. (2010),
- The linear model of coregionalisation, used for example by Wackernagel (1995),
- The moving average model of Ver Hoef and Barry (1998).

Graphical structure

- No restriction on graphical structure. Starting with a well defined joint distribution, the structure could be undirected, directed, or a chain graph (Lauritzen, 1996).
- Computationally-efficient algorithms available for some structures.

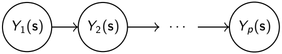
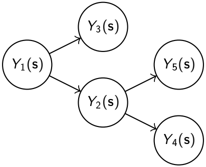
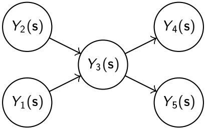


Figure : Ordered nodes





Model flexibility

NIASRA
NATIONAL INSTITUTE FOR APPLIED
STATISTICS RESEARCH AUSTRALIA

$C_{11}(\cdot,\cdot)$	
$C_{21}(\cdot,\cdot)$	$C_{22}(\cdot,\cdot)$

Figure : Bivariate system: Need to specify three marginal/cross-covariance functions.

Available building blocks: Three functions, $C_{11}(\cdot,\cdot)$, $C_{2|1}(\cdot,\cdot)$, $b(\cdot,\cdot)$.

Model flexibility

$C_{11}(\cdot,\cdot)$		
$C_{21}(\cdot,\cdot)$	$C_{22}(\cdot,\cdot)$	
$C_{31}(\cdot,\cdot)$	$C_{32}(\cdot,\cdot)$	$C_{33}(\cdot,\cdot)$

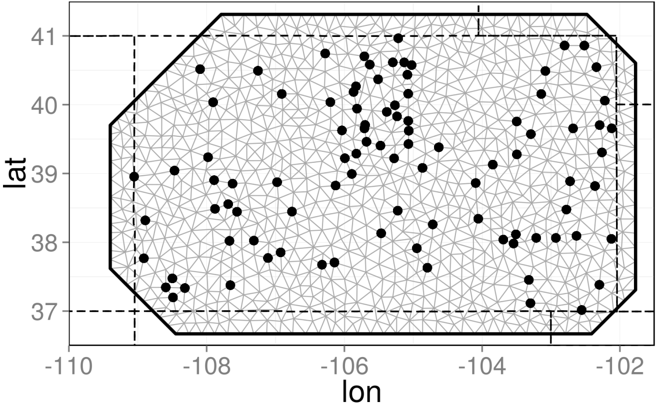
Figure : Trivariate system: Need to specify six marginal/cross-covariance functions.

Available building blocks: Six functions, $C_{11}(\cdot,\cdot)$, $C_{2|1}(\cdot,\cdot)$, $C_{3|1,2}(\cdot,\cdot)$, $b_{21}(\cdot,\cdot)$, $b_{31}(\cdot,\cdot)$, $b_{32}(\cdot,\cdot)$.

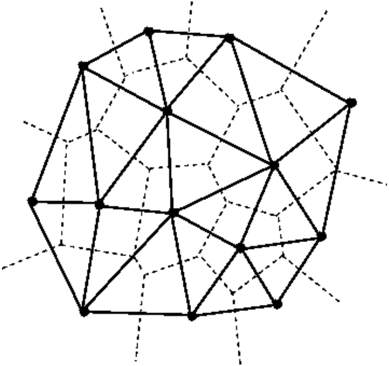
- Minimum and maximum temperatures taken on September 19, 2004 in the state of Colorado, USA.
- 94 measurement stations (collocated measurements); residuals are obtained by subtraction of statewide mean.
- Maximum-temperature residual later in the afternoon ($Y_2(\cdot)$) highly dependent on minimum-temperature residual in the early morning hours ($Y_1(\cdot)$).
- Fit three models and compare using DIC:

$$
\begin{array}{ll}
\text{Model 1:} & b_o(\mathbf{h}) \equiv 0, \\
\text{Model 2:} & b_o(\mathbf{h}) \equiv A\delta(\mathbf{h}), \\
\text{Model 3:} & b_o(\mathbf{h}) \equiv \begin{cases} A\{1 - (\|\mathbf{h} - \boldsymbol{\Delta}\|/r)^2\}^2, & \|\mathbf{h} - \boldsymbol{\Delta}\| \leq r \\ 0, & \text{otherwise.} \end{cases}
\end{array}
$$

- Consider a discretisation of $Y_1(\cdot)$ and $Y_2(\cdot)$, $\mathbf{Y}_1$ and $\mathbf{Y}_2$ respectively, and let $\mathbf{Y} \equiv (\mathbf{Y}_1, \mathbf{Y}_2)'$, $\mathbf{Z} \equiv (\mathbf{Z}_1, \mathbf{Z}_2)'$.



Numerical integrations



$$E\left(Y_2(\mathbf{s}) \mid Y_1(\cdot)\right) = \int_D b(\mathbf{s}, \mathbf{v}) Y_1(\mathbf{v})\,\mathrm{d}\mathbf{v}; \quad \mathbf{s} \in D.$$

$$E(Y_2(\mathbf{s}_l) \mid Y_1(\cdot)) \simeq \sum_{k=1}^{n} A_k b(\mathbf{s}_l, \mathbf{v}_k) Y_1(\mathbf{v}_k),$$

where $\{A_k\}$ are the tessellation areas.

- Observation model:

$$\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \Bigg| \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \boldsymbol{\theta} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{D}\mathbf{Y}_1 \\ \mathbf{D}\mathbf{Y}_2 \end{pmatrix}, \sigma_\varepsilon^2 \begin{pmatrix} \mathbf{I} & \rho_\varepsilon \mathbf{I} \\ \rho_\varepsilon \mathbf{I} & \mathbf{I} \end{pmatrix} \right),$$

where $\mathbf{D}$ is an incidence matrix and $\boldsymbol{\theta}$ includes σ_ε^2 and ρ_ε.

- Process model:

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \Bigg| \boldsymbol{\theta} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{11}\mathbf{B}' \\ \mathbf{B}\boldsymbol{\Sigma}_{11} & \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}' + \boldsymbol{\Sigma}_{2|1} \end{pmatrix} \right),$$

where $\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{2|1}$ and $\mathbf{B}$ depend on parameters in $\boldsymbol{\theta}$.

We have many unknown parameters!

- Assume that $C_{11}(\cdot)$ and $C_{2|1}(\cdot)$ are Matérn covariance functions with smoothness parameter $\nu = 3/2$.

Parameter	Model 1	Model 2	Model 3
σ^2_ε	x	x	x
ρ_ε	x	x	x
σ^2_{11}	x	x	x
$\sigma^2_{2\|1}$	x	x	x
κ_{11}	0.98 [0.76, 1.22]	1 [0.8, 1.26]	1.03 [0.83, 1.25]
$\kappa_{2\|1}$	0.76 [0.56, 1]	0.62 [0.46, 0.81]	3.65 [1.16, 6.72]
A		x	x
r			x
Δ_1			x
Δ_2			x
DIC	992.45	985.17	982.45

Median fields

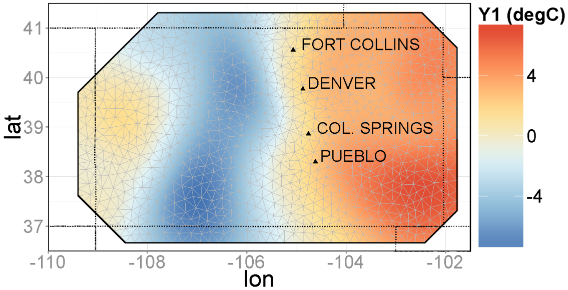


Figure : Interpolated map in degrees Celsius (degC) of $E(\mathbf{Y}_1 \mid \mathbf{Z}_1, \mathbf{Z}_2)$.

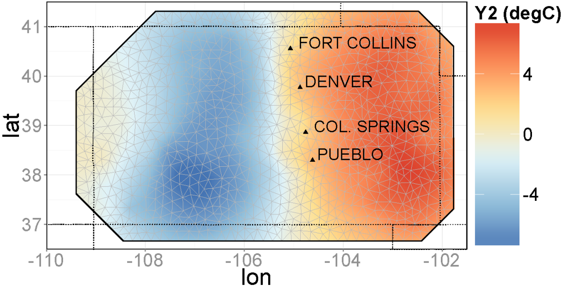


Figure : Interpolated map in degrees Celsius (degC) of $E(\mathbf{Y}_2 \mid \mathbf{Z}_1, \mathbf{Z}_2)$.

Interaction function

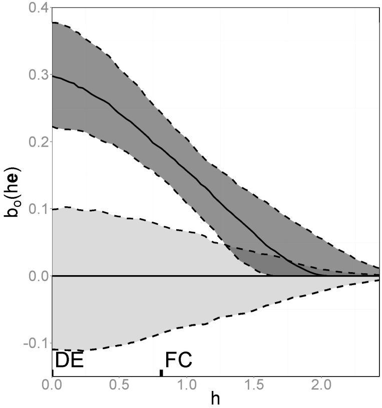


Figure : Prior (light grey) and posterior (dark grey) median (solid line) and inter-quartile ranges (enclosed by dashed lines) of the interaction function $b_o(\cdot)$ of Model 3, along a unit vector **e** originating at Denver (DE) in the direction of Fort Collins (FC)

Section 3

Atmospheric trace-gas inversion in the UK and Ireland

Atmospheric trace-gas inversion

- $Y_1(\mathbf{s})$ is methane emissions per unit area – we assume this is temporally invariant over a period of one month.
- $Y_{2,t}(\mathbf{s})$ is methane mole fraction and is spatio-temporally varying.
- **Aim:** Infer the (spatial) emissions from observation of (spatio-temporal) mole fraction.

- Prior information on methane emissions is available from flux inventories, principally the National Atmospheric Emissions Inventory (NAEI).
- A Lagrangian Particle Dispersion Model (LPDM) is used to trace particles backwards in time and thus establish the flux-mole-fraction ‘source-receptor relationship’ (SRR) that we characterise through the interaction function $b_t(\mathbf{s}, \mathbf{u})$.
- The LPDM we use is the Met Office’s National Atmospheric Modelling Environment (NAME).
- Data available at the measuring stations were averaged over 2 h time periods. The NAME output was provided at this 2 h resolution.

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

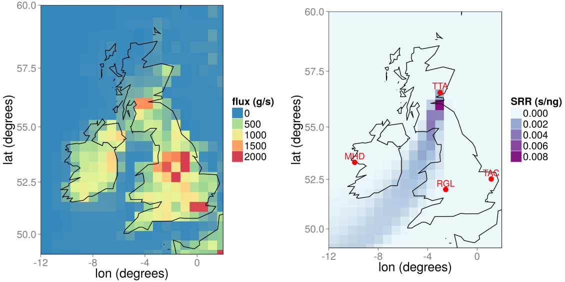


Figure : Emissions map obtained from the NAEI for January 2014 (left panel) and the interaction function $b_t(\mathbf{s}, \cdot)$ obtained from the Met Office's NAME with $\mathbf{s}$ set to the coordinates of the Angus measurement station (TTA), Scotland (right panel).

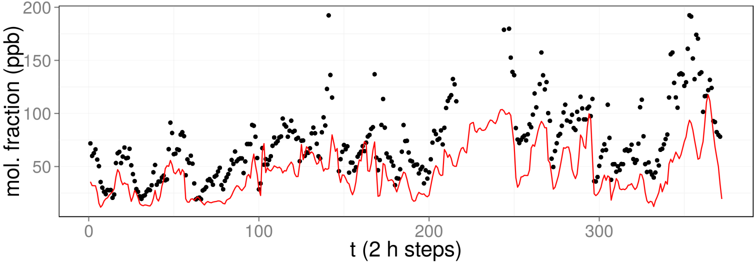


Figure : Measurements of methane mole fraction in parts per billion (ppb) at Tacolneston, England (TAC) for January 2014, following background-removal (black dots) together with a forward-model prediction (red line) using NAME and the methane flux inventory. Each time step corresponds to an interval of 2 h.

- The inversion is an ill-posed problem: We need to estimate a (spatial) emissions field from measurements that aggregate spatio-temporally.

- Let $Y_1(\mathbf{s})$ be a lognormal spatial process, that is, $\widetilde{Y}_1(\cdot) \equiv \log Y_1(\cdot)$ is a Gaussian process.
- Let $E(\widetilde{Y}_1(\mathbf{s})) \equiv \widetilde{\mu}_1(\mathbf{s};\boldsymbol{\vartheta})$ and $\mathrm{cov}(\widetilde{Y}_1(\mathbf{s}), \widetilde{Y}_1(\mathbf{u})) \equiv \widetilde{C}_{11}(\mathbf{s},\mathbf{u};\boldsymbol{\vartheta})$.

$$\begin{aligned}\mu_1(\mathbf{s};\boldsymbol{\vartheta}) &\equiv E(Y_1(\mathbf{s}))\\ &\equiv \exp(\widetilde{\mu}_1(\mathbf{s};\boldsymbol{\vartheta}) + (1/2)\widetilde{C}_{11}(\mathbf{s},\mathbf{s};\boldsymbol{\vartheta})); \quad \mathbf{s} \in D,\end{aligned}$$

$$\begin{aligned}C_{11}(\mathbf{s},\mathbf{u};\boldsymbol{\vartheta}) &\equiv \mathrm{cov}(Y_1(\mathbf{s}), Y_1(\mathbf{u}))\\ &\equiv \mu_1(\mathbf{s};\boldsymbol{\vartheta})\mu_1(\mathbf{u};\boldsymbol{\vartheta})[\exp(\widetilde{C}_{11}(\mathbf{s},\mathbf{u};\boldsymbol{\vartheta})) - 1]; \quad \mathbf{s},\mathbf{u} \in D.\end{aligned}$$

Lognormal bivariate model

- We have a causal **spatio-temporal** bivariate model:

$$E\left(Y_{2,t}(\mathbf{s}) \mid Y_1(\cdot)\right) = \int_D b_t(\mathbf{s},\mathbf{v})\, Y_1(\mathbf{v})\,\mathrm{d}\mathbf{v}; \quad \mathbf{s} \in D,$$
$$\operatorname{cov}\left(Y_{2,t}(\mathbf{s}), Y_{2,t'}(\mathbf{u}) \mid Y_1(\cdot)\right) = C_{2|1,t,t'}(\mathbf{s},\mathbf{u}); \quad \mathbf{s},\mathbf{u} \in \mathbb{R}^d,$$

  with the mole-fraction covariance:

$$C_{22,t,t'}(\mathbf{s},\mathbf{u}) = C_{2|1,t,t'}(\mathbf{s},\mathbf{u}) + \int_D \int_D b_t(\mathbf{s},\mathbf{v})\, C_{11}(\mathbf{v},\mathbf{w})\, b_{t'}(\mathbf{w},\mathbf{u})\,\mathrm{d}\mathbf{v}\,\mathrm{d}\mathbf{w}.$$

- The conditional covariance $C_{2|1,t,t'}$ is used to account for simulator discrepancy (boundary conditions, model discretisation, linearisation, etc.).

NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

$$\mathbf{Y}_t(\cdot) \sim \mathrm{Dist}\left(\begin{pmatrix}\mu_1(\cdot)\\ \mu_{2,t}(\cdot)\end{pmatrix}, \begin{pmatrix} C_{11}(\cdot,\cdot) & C_{12,t'}(\cdot,\cdot) \\ C_{21,t}(\cdot,\cdot) & C_{22,t,t'}(\cdot,\cdot)\end{pmatrix}\right), \quad t, t' \in \mathbb{R}^+.$$

- What to choose for $C_{2|1,t,t'}(\mathbf{s}, \mathbf{u})$?
- *Strategy 1*: If $\dim(\mathbf{Z}_{2,t}) < 10$, say, then use standard spatio-temporal covariance functions, which yield (dense) covariance matrices, and evaluate $Y_{2,t}(\cdot)$ only where we take observations.
- *Strategy 2*: If $\dim(\mathbf{Z}_{2,t}) \gg 10$, then we need to use sequential estimation methods, dimensionality reduction and/or **matrix sparsity**.

- The discrepancy is a separable spatio-temporal Gaussian process with

$$\begin{aligned} C_{2|1,t,t'}(\mathbf{s},\mathbf{u}) &= \sigma^2_{2|1}\rho_s(\mathbf{s},\mathbf{u};d)\rho_t(t,t';a), \\ \rho_s(\mathbf{s},\mathbf{u};d) &\equiv \exp(\|\mathbf{s}-\mathbf{u}\|/d); \quad d > 0, \\ \rho_t(t,t';a) &\equiv a^{|t-t'|}; \quad |a| < 1, \end{aligned}$$

- Then $\boldsymbol{\Sigma}_{2|1} = \sigma^2_{2|1}\widetilde{\boldsymbol{\Sigma}}_{2|1,t} \otimes \widetilde{\boldsymbol{\Sigma}}_{2|1,s}$.
- $\mathbf{B}$ is obtained by concatenating $\{\mathbf{B}_t : t = 1,2,\ldots\}$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{11}\mathbf{B}' \\ \mathbf{B}\boldsymbol{\Sigma}_{11} & \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}' + \boldsymbol{\Sigma}_{2|1} \end{pmatrix}.$$

.

- $\mathbf{B}$ is dense: Sparse covariance matrices are not computationally helpful.

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \mathbf{B}'\mathbf{Q}_{2|1}\mathbf{B} + \mathbf{Q}_{11} & -\mathbf{B}'\mathbf{Q}_{2|1} \\ -\mathbf{Q}_{2|1}\mathbf{B} & \mathbf{Q}_{2|1} \end{pmatrix}.$$

- Large benefit by making sure the (very large) matrix $\mathbf{Q}_{2|1}$ is sparse.

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \mathbf{B}'\mathbf{Q}_{2|1}\mathbf{B} + \mathbf{Q}_{11} & -\mathbf{B}'\mathbf{Q}_{2|1} \\ -\mathbf{Q}_{2|1}\mathbf{B} & \mathbf{Q}_{2|1} \end{pmatrix}.$$

- Large benefit by making sure the (very large) matrix $\mathbf{Q}_{2|1}$ is sparse.
- We define

$$\mathbf{Q}_{2|1} \equiv \sigma_{2|1}^{-2}\widetilde{\mathbf{Q}}_{2|1,t} \otimes \widetilde{\mathbf{Q}}_{2|1,s} \quad ,$$

where

$$\widetilde{\mathbf{Q}}_{2|1,t} \equiv \begin{pmatrix} 1 & -a & 0 & & & & 0 \\ -a & (1+a^2) & -a & & & & 0 \\ & & & \ddots & & & \\ 0 & & & & -a & (1+a^2) & -a \\ 0 & & & & 0 & -a & 1 \end{pmatrix}.$$

and we get $\mathbf{Q}_{2|1,s}$ from an intrinsic Gaussian Markov random field specification.

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \mathbf{B}'\mathbf{Q}_{2|1}\mathbf{B} + \mathbf{Q}_{11} & -\mathbf{B}'\mathbf{Q}_{2|1} \\ -\mathbf{Q}_{2|1}\mathbf{B} & \mathbf{Q}_{2|1} \end{pmatrix}.$$

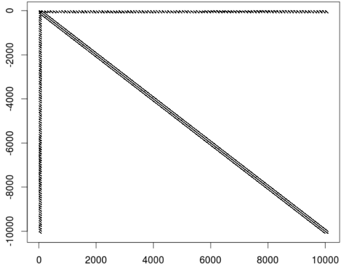


Figure : Sparsity pattern of $\boldsymbol{\Sigma}^{-1}$. Note how the top-left, top-right and bottom-left blocks are dense. The largest block, $\mathbf{Q}_{2|1}$ is very sparse due to its (sparse) Kronecker structure.

- Interaction function (transport model) $b_t(\mathbf{s}, \mathbf{u})$ is assumed known from NAME.

- We can obtain reasonable estimates of the parameters appearing in $C_{11}(\mathbf{s}, \mathbf{u})$ from inventories (range, marginal variance, and nugget).

- We have no idea what the parameters appearing in $C_{2|1}(\mathbf{s}, \mathbf{u})$ are. These need to be estimated.

- This is an ill-posed problem: MCMC chains where both the parameters and the fields are unknown will not mix well.

- This is an ill-posed problem: MCMC chains where both the parameters and the fields are unknown will not mix well.
- We know the first two moments – use a **Laplace-approximated-EM** to estimate the parameters, then fix the parameters and use MCMC for inferring the fields.

- This is an ill-posed problem: MCMC chains where both the parameters and the fields are unknown will not mix well.
- We know the first two moments – use a **Laplace-approximated-EM** to estimate the parameters, then fix the parameters and use MCMC for inferring the fields.
- This is known as an empirical hierarchical model (EHM; Cressie and Wikle, 2011).

Computational strategy

- This is an ill-posed problem: MCMC chains where both the parameters and the fields are unknown will not mix well.
- We know the first two moments – use a **Laplace-approximated-EM** to estimate the parameters, then fix the parameters and use MCMC for inferring the fields.
- This is known as an empirical hierarchical model (EHM; Cressie and Wikle, 2011).
- We can compute all the (horrible) gradients analytically, so we use an MCMC method that takes advantage of these. This is known as Hamiltonian Monte Carlo (HMC).

HMC

- Use Hamiltonian dynamics to propose the next state in an MCMC chain (Duane et al., 1987).
- Need knowledge of gradient to simulate dynamics.
- Suitable when variables are highly correlated *a posteriori* (ill-posed problem).
- Dynamics are simulated using standard methods (Euler or leapfrog method).
- One-step updates = Langevin method.
- HMC chains are ergodic and reversible (Neal, 2011).

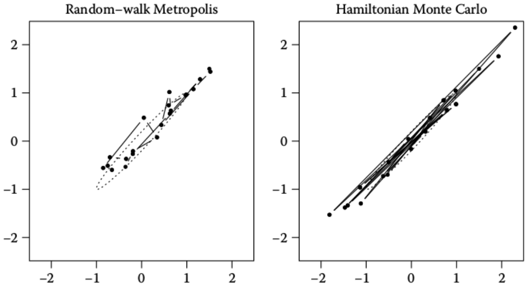


FIGURE 5.4
Twenty iterations of the random-walk Metropolis method (with 20 updates per iteration) and of the Hamiltonian Monte Carlo method (with 20 leapfrog steps per trajectory) for a two-dimensional Gaussian distribution with marginal standard deviations of one and correlation 0.98. Only the two position coordinates are plotted, with ellipses drawn one standard deviation away from the mean.

- Figure and caption taken from Neal (2011).

1. Assume the properties of the lognormal flux spatial process (i.e., $\widetilde{C}_{11}$ and $\widetilde{\mu}_1$ are known), and simulate a realisation.

2. Simulate a spatio-temporal interaction function (assumed known).

3. Simulate mole fraction observations at a few locations (Model 1) and at many (1000) locations (Model 2).

4. Infer the flux $Y_1(\mathbf{s})$ from the data in both cases.

- Simulated interaction function.

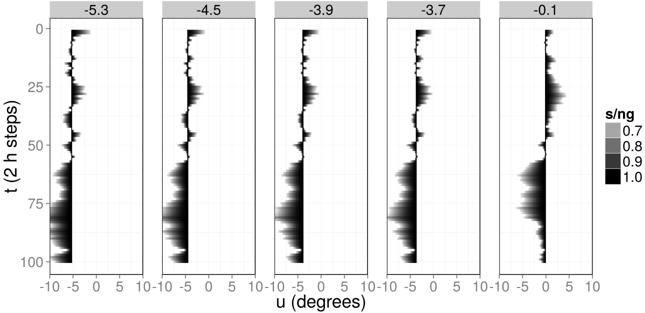


Figure : The source-receptor relationship $b_t(s, u)$ synthesised at five observation locations $s \in D_m^O = \{-5.3^\circ, -4.5^\circ, -3.9^\circ, -3.7^\circ, -0.1^\circ\}$.

- Flux and time-averaged mole-fraction field.

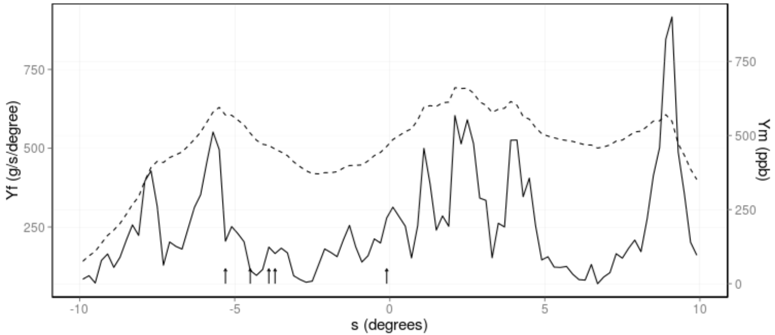


Figure : A sample realisation of the flux field (solid line), the resulting mole-fraction field averaged over the entire temporal domain (dashed line) and the five observation locations (arrows).

- Laplace-EM proved relatively straightforward to implement.
- Convergence with Model 1 $\simeq$ 80 iterations.
- Convergence with Model 2 $\simeq$ 5 iterations (much more data).
- Convergence may be hard to achieve when mode is close to zero and tails are heavy.
- "Bouncing method" needs to be implemented for the HMC chain to respect positivity constraint on flux (Neal, 2011).

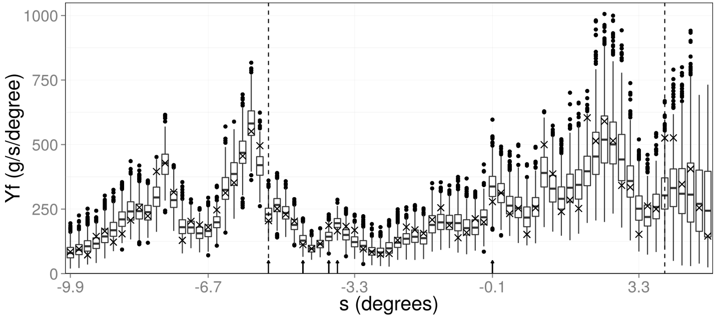


Figure : Samples from the posterior distribution of the lognormal flux field obtained using Hamiltonian Monte Carlo (HMC). The crosses denote the true (simulated) fluxes, and the arrows denote the locations of the measurements. The vertical dashed lines show the spatial locations displayed in the next slide.

- Why include the HMC if we have a Laplace-EM algorithm?

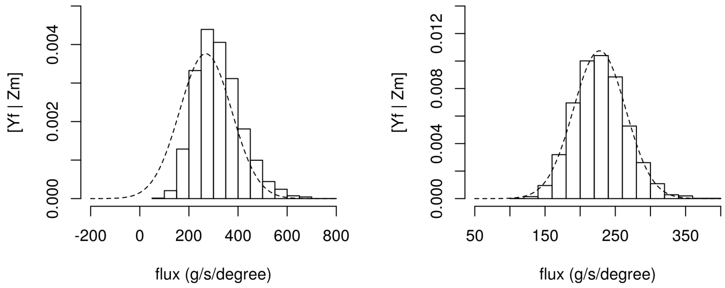


Figure : Laplace approximation (dashed line) and a histogram of the empirical posterior distribution from the MCMC samples for the flux at $s = 3.9°$ (left panel) and $s = -5.3°$ (right panel).

- Extract spatial properties from the emissions inventory.

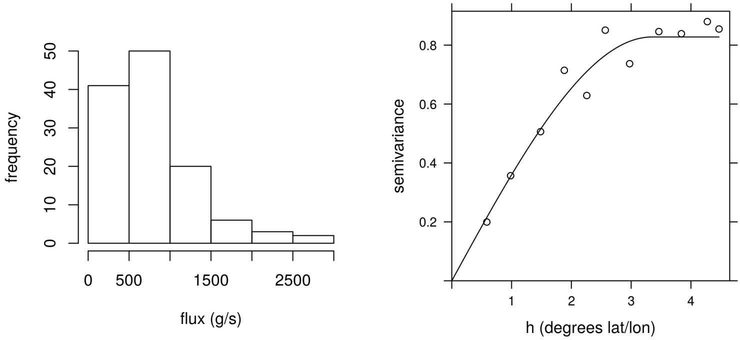


Figure : Histogram of inventory fluxes in the UK and Ireland following regridding (left panel) and the empirical (open circles) and fitted (solid line) semi-variogram as a function of lag distance in degrees lat/lon.

- We set the prior-flux expectation to be spatially constant.
- We used Model 1 since we only have 4 stations.
- Laplace-EM converged in $\simeq$ 30 iterations.
- Simulator discrepancy is not negligible:
  1. $\hat{\sigma}_{2|1} \simeq 26$ ppb,
  2. $\hat{d} \simeq 200$ km,
  3. $\hat{a} \simeq 0.97$ ($1/e$ rate of 66 h).

NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

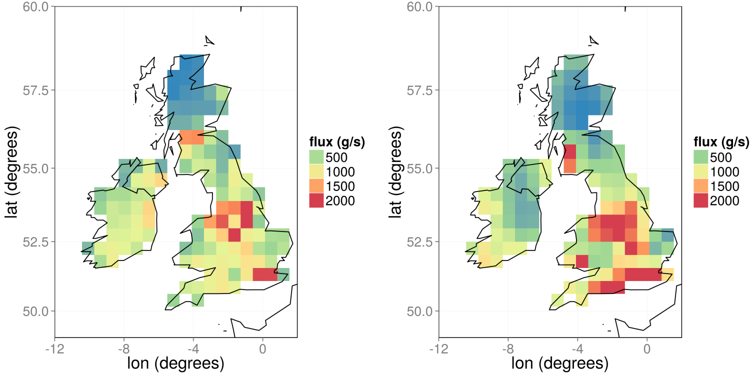


Figure : Emissions inventory (left panel) and 95 percentile (right panel) methane emissions in the UK and Ireland, obtained using the Laplace-EM/HMC approach. Emissions in the white grid cells were assumed known and used to correct the observations.

Section 4

Conclusions

- Bivariate and multivariate models often appear in environmental studies. Usually, one or more of these are 'explained away' prior to commencing the analysis.
- Causal models allow for a (very) flexible model class through interaction functions that can be arbitrarily complex.
- Computation is key: For large, non-Gaussian systems, approximate message passing + variational techniques are probably needed (Cseke et al., 2014).
- Slides and reproducible code available at `https://github.com/andrewzm/bicon`.
- Thanks to Anita Ganesan and Matthew Rigby (University of Bristol) for help with the case study of CH_4 fluxes.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. John Wiley and Sons, Hoboken, NJ.

Cseke, B., Mangion, A. Z., Sanguinetti, G., and Heskes, T. (2014). Sparse approximations in spatio-temporal point-process models. *arXiv preprint arXiv:1305.4152*.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195:216–222.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105:1167–1177.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford, UK.

Neal, R. (2011). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Press, Boca Raton, FL.

Royle, J. A. and Berliner, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 4:29–56.

Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69:275–294.

Wackernagel, H. (1995). *Multivariate Geostatistics*. Springer, Berlin, DE.

Zammit-Mangion, A., Bamber, J. L., Schoen, N. W., and Rougier, J. C. (2015a). A data-driven approach for assessing ice-sheet mass balance in space and time. *Annals of Glaciology, in press*.

Zammit-Mangion, A., Rougier, J. C., Bamber, J. L., and Schoen, N. W. (2014). Resolving the Antarctic contribution to sea-level rise: a hierarchical modelling framework. *Environmetrics*, 25:245–264.

Zammit-Mangion, A., Rougier, J. C., Schoen, N., Lindgren, F., and Bamber, J. (2015b). Multivariate spatio-temporal modelling for assessing Antarctica's present-day contribution to sea-level rise. *Environmetrics*, :doi: 10.1002/env.2323, in press.