Locating and quantifying gas emission sources using remotely obtained concentration data

Philip Jonathan

Lancaster University, Department of Mathematics & Statistics, UK.
Shell Research Ltd., London, UK.

Seminar: Data Science of the Natural Environment
(slides at *www.lancs.ac.uk/∼jonathan*)

Over 2 million wells in North America

GWP values and lifetimes from 2013 IPCC AR5 p714 (with climate-carbon feedbacks)[8]	Lifetime (years)	GWP	
		20 years	100 years
Methane	12.4	86	34
HFC-134a (hydrofluorocarbon)	13.4	3790	1550
CFC-11 (chlorofluorocarbon)	45.0	7020	5350
Nitrous oxide (N_2O)	121.0	268	298
Carbon tetrafluoride (CF_4)	50000	4950	7350

Global warming potentials (wiki: time-integrated energy released from instantaneous release of 1 kg of trace substance relative to that of 1 kg of CO_2)

Thanks

All of what follows is joint work led by

- Bill Hirst (Shell)

with

- David Randell (Shell)
- Oliver Kosut (MIT, now Arizona State)
- Fernando Gonzalez del Cueto (Shell, now Lumos Imaging, Denver)

Thanks also to

- Rutger Ijzermans, Matthew Jones (Shell)

Outline

Motivation

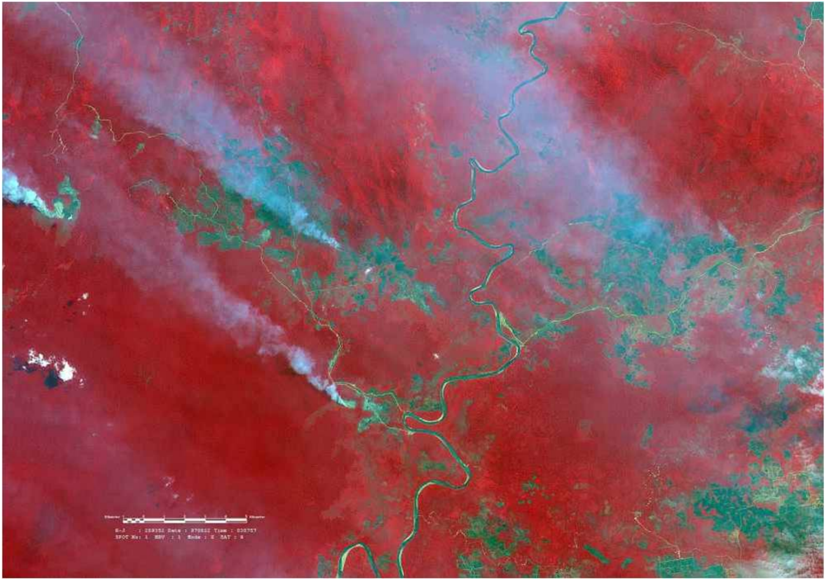
- A method for detecting, locating and quantifying sources of gas emissions to the atmosphere
- From remotely obtained atmospheric gas concentration measurements

Issues

- Potentially **large** background gas concentrations ($\approx 1800ppb$ for CH_4)
- Need to detect small signals ($\approx 5 - 35ppb$ for CH_4)
- Gas dispersion determined by prevailing wind conditions

Approach

- Plume model represents gas dispersion between source and measurement location
- Measured concentration is sum of contributions from sources and **relatively smooth** background
- Infer source locations, source emission rates, background level, plume biases and uncertainties



Smoke plumes (Gaussian plume in far field)

Survey aircraft ($\approx 50ms^{-1}$, $\approx 200m$ above ground)

Motivating test applications

Synthetic problem

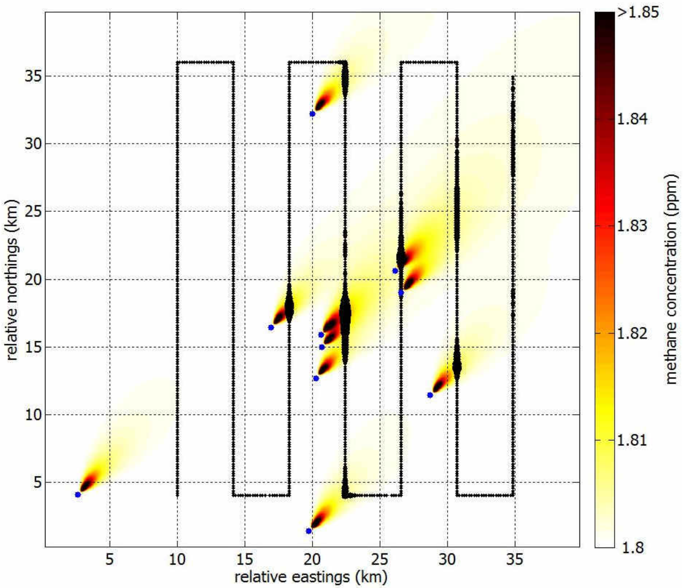
- **Known** wind field, sources and background, 10 sources

Landfill

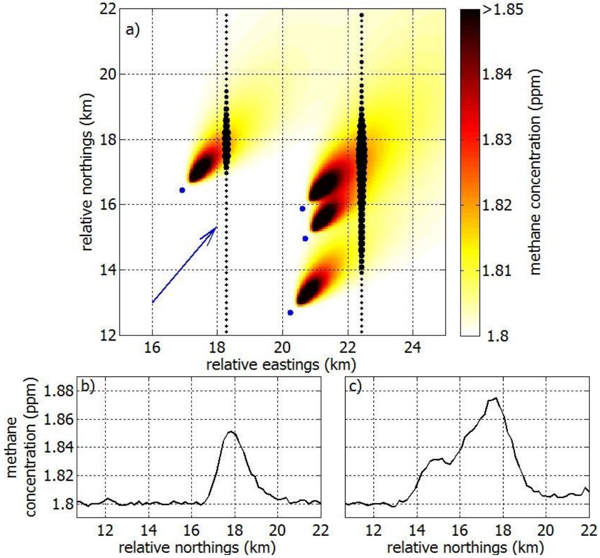
- 2 landfill regions, probable diffuse sources
- Wind field from UK met–office **global circulation model**

Flare stack

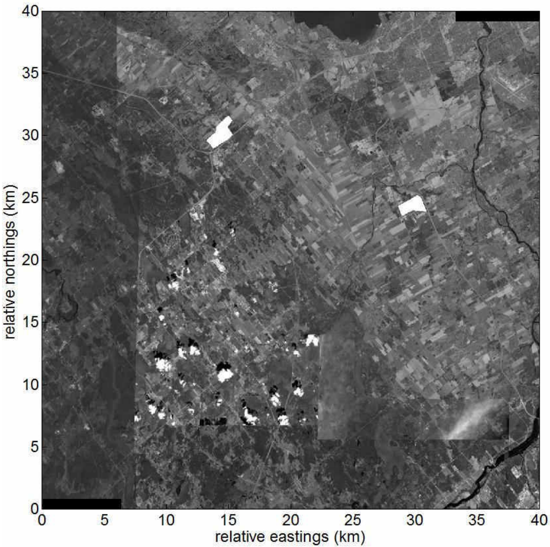
- Single elevated near–point source
- Wind field from UK met–office global circulation model
- **Coastal** location



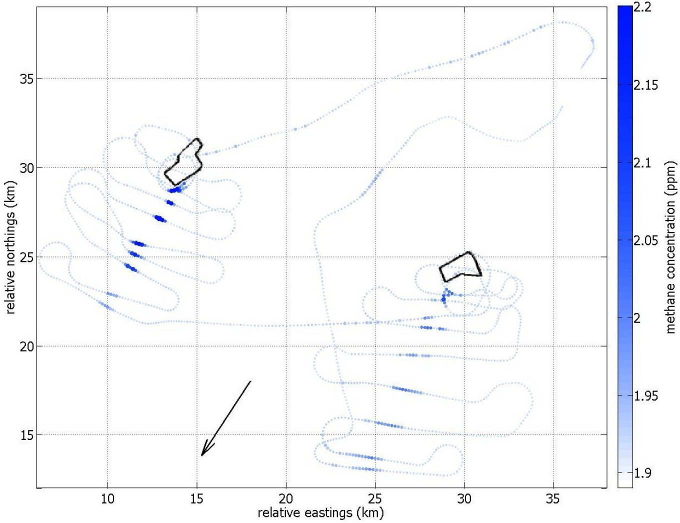
Synthetic problem revealed



(a) two passes x–y (b) first pass in time (c) second pass in time

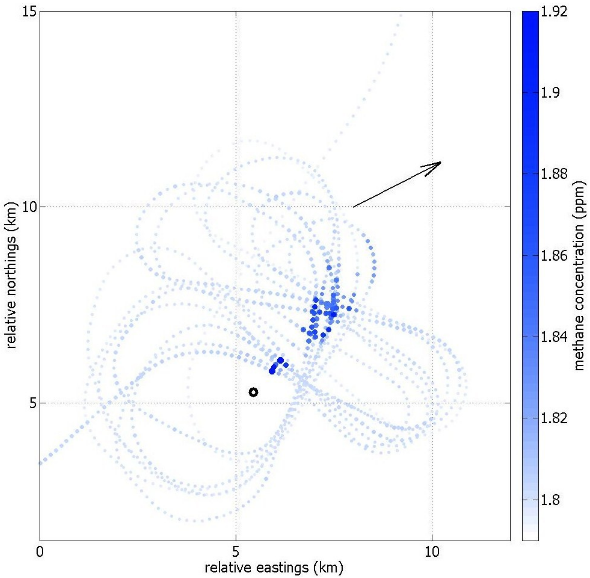


Landfill from above



Landfill measurements

Flare stack



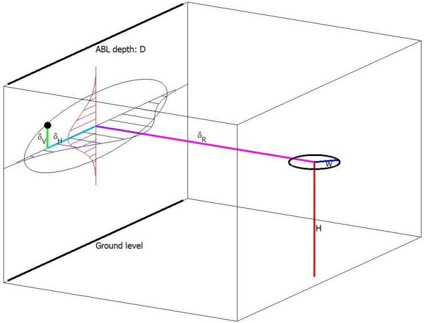
Flare stack measurements (wind direction bias)

Model formulation

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{b} + \boldsymbol{\epsilon}$$

- $\mathbf{y}$: measured concentrations
- $\mathbf{A}$: assumed known from plume model
- $\mathbf{s}$: sources to be estimated
- $\mathbf{b}$: background to be estimated
- $\boldsymbol{\epsilon}$: measurement error (assumed Gaussian), variance to be estimated

Plume model



- Red: Source height H
- Blue: Source half–width w
- Magenta: Downwind offset δ_R
- Cyan: Horizontal offset δ_H
- Green: Vertical offset δ_V
- **ABL** height: D
- Horizontal extent: $\sigma_H = \delta_R \tan(\gamma_H) + w$
- Vertical extent: $\sigma_V = \delta_R \tan(\gamma_V)$
- *Opening angles*: γ_H, γ_V

$$a = \frac{1}{2\pi|\mathbf{U}|\sigma_H\sigma_V} \exp\left\{-\frac{\delta_H^2}{2\sigma_H^2}\right\} \times \left\{ \exp\left\{-\frac{(\delta_V - H)^2}{2\sigma_V^2}\right\} + \exp\left\{-\frac{(\delta_V + H)^2}{2\sigma_V^2}\right\} \right.$$
$$\left. + \exp\left\{-\frac{(\mathbf{2D} - \delta_V - H)^2}{2\sigma_V^2}\right\} + \exp\left\{-\frac{(2D - \delta_V + H)^2}{2\sigma_V^2}\right\} \right\}$$

Background model

Requirements

- Positive and smoothly–varying, spatially and temporally
- Basis function representation: $\mathbf{b} = \mathbf{P}\boldsymbol{\beta}$
- We use Gaussian Markov random field
- Explicit spatial dependence due to wind transport incorporated

Random field prior

$$f(\boldsymbol{\beta}) \propto \exp\{-\tfrac{\mu}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{J}_{\boldsymbol{\beta}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\}$$

- $\mathbf{J}_{\boldsymbol{\beta}}$ is sparse, $\mathbf{P} = \mathbf{I}$
- Fast estimation

Inference strategy

Initial point estimation

- Sources and background
- Source locations assumed on **fixed grid**
- **Fast** estimation of starting solution for Bayesian inference

Subsequent Bayesian inference

- Sources, background, measurement error, wind–field parameters, ...
- **Grid-free** sources modelled using Gaussian **mixture model**
- Reversible jump MCMC inference
- Quantified parameter **uncertainties** and **dependencies**

Initial point estimation

Background prior

$$f(\boldsymbol{\beta}) \propto \exp\{-\tfrac{\mu}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T \mathbf{J}_{\boldsymbol{\beta}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\}$$

Source prior (**Laplace**)

$$f(\mathbf{s}) \propto \exp\{-\lambda\|\mathbf{Q}\mathbf{s}\|_1\}$$

Likelihood

$$f(\mathbf{y}|\mathbf{s},\boldsymbol{\beta}) \propto \exp\{-\tfrac{1}{2\sigma_\epsilon^2}\|A\mathbf{s}+P\boldsymbol{\beta}-\mathbf{y}\|^2\},$$

Posterior

$$f(\mathbf{s},\boldsymbol{\beta}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{s},\boldsymbol{\beta})f(\mathbf{s})f(\boldsymbol{\beta})$$

Maximum a-posteriori estimate

$$\mathrm{argmin}_{\mathbf{s},\boldsymbol{\beta}} \qquad \tfrac{1}{2\sigma_\epsilon^2}\|A\mathbf{s}+P\boldsymbol{\beta}-\mathbf{y}\|^2 + \tfrac{\mu}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T J(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \lambda\|Q\mathbf{s}\|_1$$

Bayesian inference

Parameters

- Source locations $\mathbf{z}$, "widths" $\mathbf{w}$ and emission rates $\mathbf{s}$ for mixture of m sources
- Random field background parameters $\boldsymbol{\beta}$
- Measurement error standard deviation σ_ϵ
- Wind–direction correction δ_ϕ
- Others (e.g. plume opening angles)
- Call these $\boldsymbol{\theta}$ which can be partitioned $\{\boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_{\overline{\kappa}}\}$

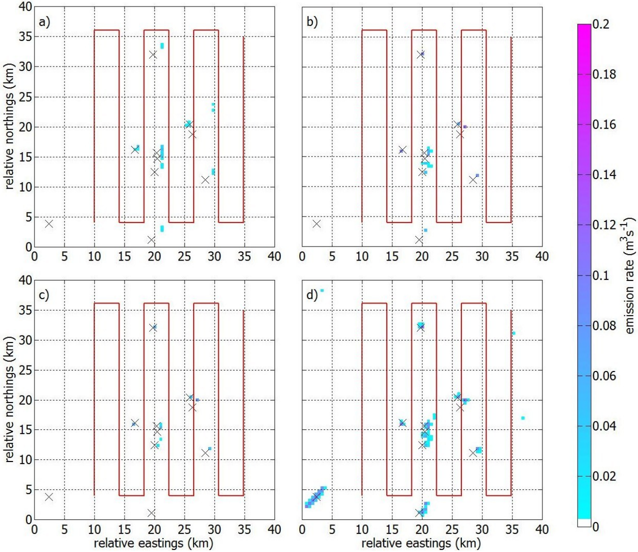
Full conditional

$$f(\boldsymbol{\theta}_\kappa | \mathbf{y}, \boldsymbol{\theta}_{\overline{\kappa}}) \propto f(\mathbf{y} | \boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_{\overline{\kappa}}) f(\boldsymbol{\theta}_\kappa | \boldsymbol{\theta}_{\overline{\kappa}})$$

Inference tools

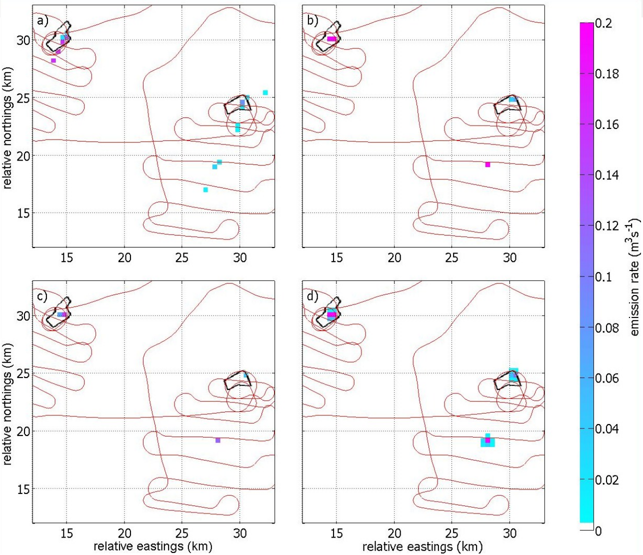
- Gibbs' sampling
- **Reversible jump**
- (Metropolis–Hastings)

Synthetic



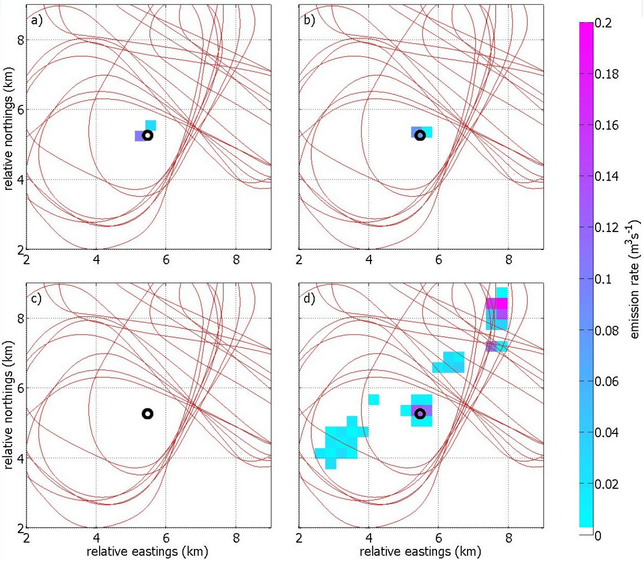
(a) initial (b) median (c) 2.5% (d) 97.5%

Landfill



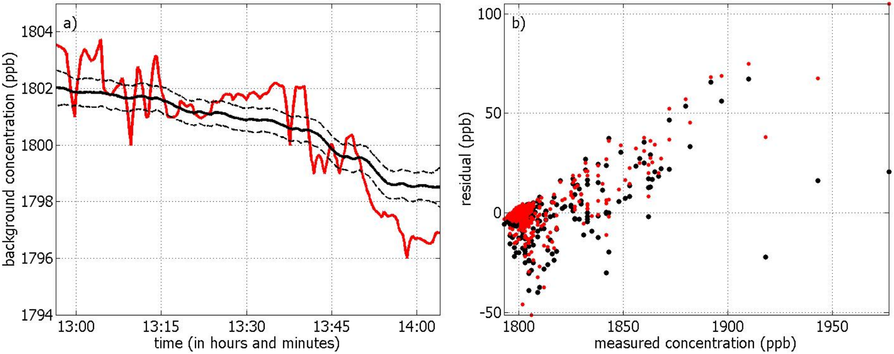
(a) initial (b) median (c) 2.5% (d) 97.5%

Flare stack



(a) initial (b) median (c) 2.5% (d) 97.5%

Flare stack



(a) background in time (b) residual vs measured concentration
initial (red); posterior median (black)

Wind direction correction of 18^o

Conclusions and on–going work

Conclusions

- **Data structure** and management
- Flexible inference using combination of standard methods
- Good performance on synthetic and field applications
- **Scalability** from iterative estimation

Extensions (on-going and potential)

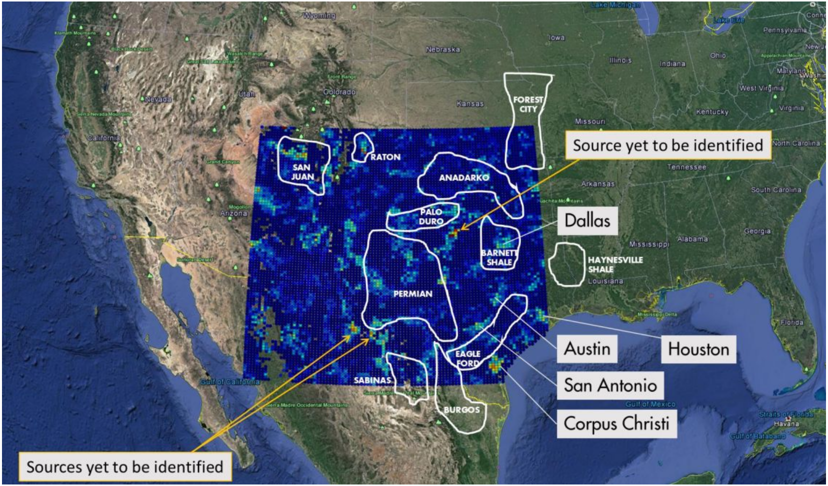
- Multiple flights, multiple wind data sources
- Enhanced plume model
- Internal **calibration**
- Improved prior characterisation of sources, intermittent sources
- Simultaneous inference using multiple measurement types
- Optimal **design**
- **Line-of-sight** and **satellite** applications

Slides and articles at *www.lancs.ac.uk/∼jonathan*

Line-of-sight sensing

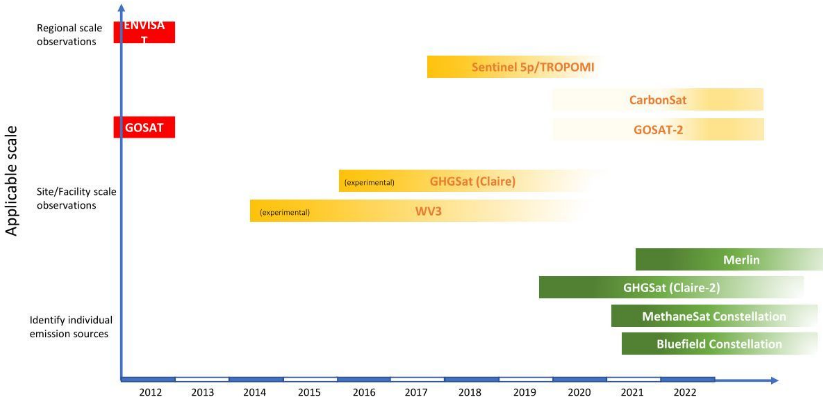
Line-of-sight laser

Satellite



Tropomi satellite and Google Earth

Satellite Service Dates



Potential to measure individual sources