Chemometrics and Intelligent Laboratory Systems 149 (2015) 227–241

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Spatio-temporal bivariate statistical models for atmospheric trace-gas inversion

Andrew Zammit-Mangion [a,*], Noel Cressie [a], Anita L. Ganesan [b], Simon O'Doherty [c], Alistair J. Manning [d]

[a] National Institute for Applied Statistics Research Australia (NIASRA), School of Mathematics and Applied Statistics (SMAS), University of Wollongong, Northfields Avenue, Wollongong NSW 2522, Australia
[b] School of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, UK
[c] School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK
[d] Hadley Centre, Met Office, Fitzroy Road, Exeter, Devon EX1 3PB, UK

ARTICLE INFO

Article history:
Received 17 April 2015
Received in revised form 3 September 2015
Accepted 7 September 2015
Available online 12 September 2015

Keywords:
Conditional multivariate models
Methane emissions
Multivariate geostatistics
Hamiltonian Monte Carlo
Spatial statistics

ABSTRACT

Atmospheric trace-gas inversion refers to any technique used to predict spatial and temporal fluxes using mole-fraction measurements and atmospheric simulations obtained from computer models. Studies to date are most often of a data-assimilation flavour, which implicitly consider univariate statistical models with the flux as the variate of interest. This univariate approach typically assumes that the flux field is either a spatially correlated Gaussian process or a spatially uncorrelated non-Gaussian process with prior expectation fixed using flux inventories (e.g., NAEI or EDGAR). Here, we extend this approach in three ways. First, we develop a bivariate model for the mole-fraction field and the flux field. The bivariate approach allows optimal prediction of both the flux field and the mole-fraction field, and it leads to significant computational savings over the univariate approach. Second, we employ a lognormal spatial process for the flux field that captures both the lognormal characteristics of the flux field (when appropriate) and its spatial dependence. Third, we propose a new, geostatistical approach to incorporate the flux inventories in our updates, such that the posterior spatial distribution of the flux field is predominantly data-driven. The approach is illustrated on a case study of methane (CH_4) emissions in the United Kingdom and Ireland.

1. Introduction

Atmospheric trace-gas inversion refers to any technique used to predict spatial and temporal fluxes of a gas from observations of mole fractions. Since mole-fraction measurements are affected by weather patterns that are time varying, there is no straightforward relationship between the observations and the fluxes. The relationship, termed a 'source-receptor relationship' (SRR), is typically obtained by simulating from a computer model, such as a Lagrangian Particle Dispersion Model (LPDM), which maps the fluxes to the mole fractions [1]. The SRR can be characterised through a bivariate function $b_t(\mathbf{s}, \mathbf{u})$ where the spatial locations $\mathbf{s}, \mathbf{u} \in \mathbb{R}^2$ are in a given domain of interest, $D \subset \mathbb{R}^2$, and $t \in \{1, 2, \ldots\}$ is a discrete-time index.

Fig. 1, top-left panel, shows the UK and Ireland methane (CH_4) total fluxes by grid cell, in units g s^{-1}, obtained from a combination of flux inventories. Here, the component inventories are principally the National Atmospheric Emissions Inventory (NAEI) [2] and the Emissions Database for Global Atmospheric Research version 4.2 (EDGAR) [3]; see [4] for details. Fig. 1, top-right panel, shows an evaluation of $b_t(\mathbf{s}, \mathbf{u})$, in units s ng^{-1}, of methane fluxes, for $\mathbf{s}$ located at a mole-fraction monitoring station at Angus, Scotland (TTA), where $\mathbf{u}$ takes values on a discrete lattice, and where $t = 1$ (01 January 2014 at midnight). Fig. 1, bottom panel, shows the two-hourly averaged observations in parts per billion (ppb) at TTA, following background-removal (see Section 4.2 for details), for all of January 2014. In atmospheric trace-gas inversion, the aim is to recover a flux map, such as that in Fig. 1, top-left panel, from time-series observations taken at various monitoring stations and from the collection of atmospheric simulations (one such is shown in Fig. 1, top-right panel) that establish the SRR.

Atmospheric trace-gas inversion is an ill-posed problem. Consequently, "small uncertainties in the observational data correspond to much higher uncertainty in the emission[s]" [5]. However, in addition to observation and flux-field uncertainties, there is a third source of uncertainty arising from the use of an atmospheric model that does not perfectly match the true SRR due to physical parameterisations, solver discretisations, etc. Sometimes this third term is called the *discrepancy term*, and failure to acknowledge it can lead to over-confident predictions [6,7]. Until very recently, this discrepancy was not considered separately from the observation error; see [4]. However, it is crucial to distinguish between the errors due to observations and those due to model misspecification, as these are likely to have different statistical properties.

* Corresponding author.
E-mail address: azm@uow.edu.au (A. Zammit-Mangion).

http://dx.doi.org/10.1016/j.chemolab.2015.09.006
0169-7439/

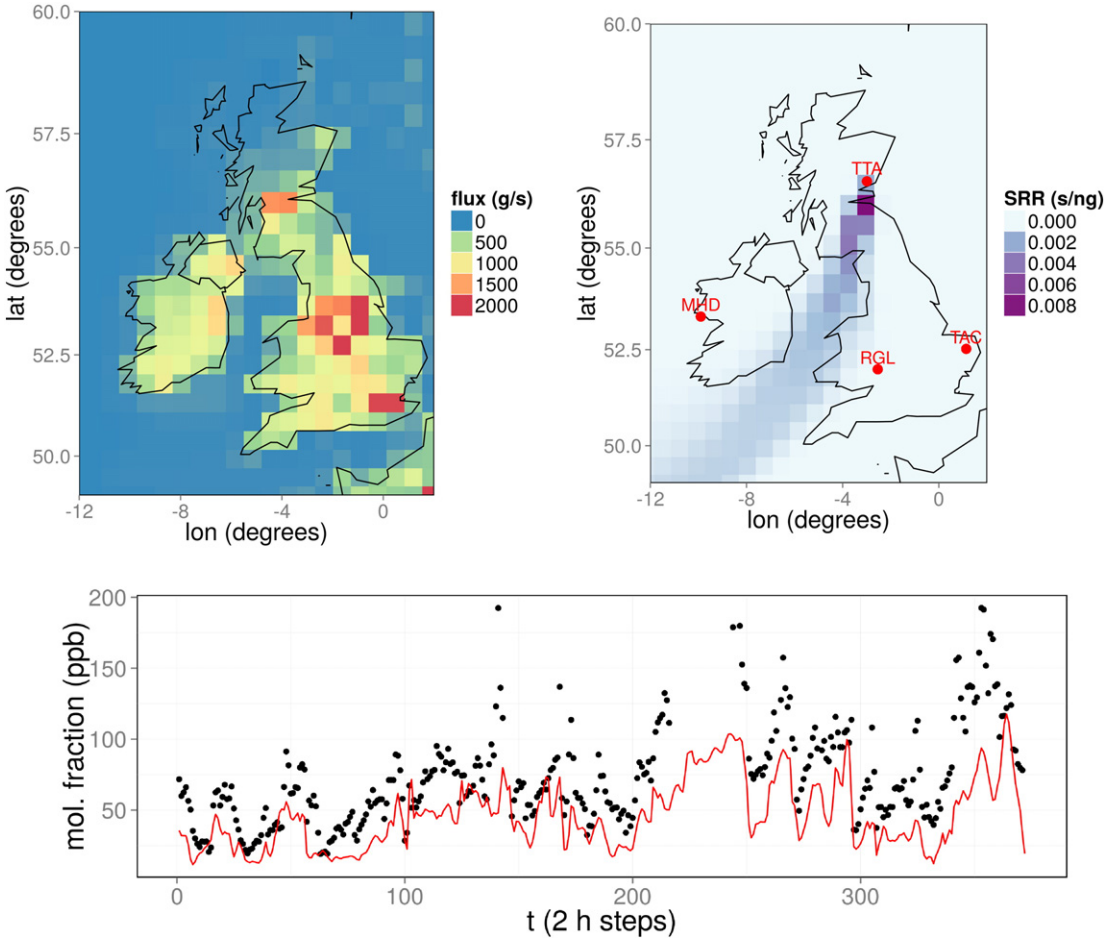

Fig. 1. Top-left panel: Total flux by grid cell in g s^{-1} obtained by combining methane inventories (see [4] for details) for January 2014. Top-right panel: The source-receptor relationship on 01 January 2014 at 00:00, $b_1(\mathbf{s}_{TTA}, \cdot)$, obtained from the UK Met Office's Lagrangian Particle Dispersion Model (LPDM), Numerical Atmospheric-dispersion Modelling Environment (NAME), where $\mathbf{s}_{TTA}$ is the coordinate vector of the Angus measurement station (TTA), in Scotland. Also shown are the three other stations used in this study, Mace Head (MHD), Ridge Hill (RGL), and Tacolneston (TAC). Bottom panel: Measurements of methane mole fraction in parts per billion (ppb) at TAC for January 2014, following background-removal (black dots), together with a straightforward prediction (red line) using NAME and the methane flux inventory. Each time step corresponds to an interval of 2 h.

A critical contribution of our work is to formalise the insight in [4] and treat the mole-fraction field as a second variable of interest. The consideration of modelling a mole-fraction field in addition to the flux field through a discrepancy term results in a *bivariate* model (see [8] for a recent review on such models). The bivariate model brings two advantages to this problem of trace-gas inversion: The locations at which the mole-fraction field is modelled need not coincide with the data points. This in turn allows the predictive distribution of mole fraction at any unobserved locations to be found, and then averaged over any subset of the spatio-temporal domain, with relative ease (provided the SRR is available for these locations/domain). The other advantage is that the decoupling leads to computationally efficient methods in spatial statistics that can be used to scale up the inversion to large, remote-sensing datasets.

Another contribution of our work is to introduce the lognormal spatial process as a prior distribution for the flux field. This model acts as a bridge between the two sides of the dichotomous literature that either assumes spatial (possibly truncated) Gaussian-process priors (e.g., [9]) or spatially uncorrelated lognormal priors (e.g., [10]). A lognormal spatial process is attractive, as it is able to capture both (i) the nonnegativity and heavy tails in the distribution of the flux (valid for some trace gases such as methane) and (ii) the spatial correlation of the flux field. We show that expectations and covariances for both the flux field and the mole-fraction field can be obtained analytically if the flux field is defined as a lognormal process, by directly applying results from the univariate case ([11], [12], p. 135).

The third contribution of our work is to propose a new way to carry out assimilation in atmospheric trace-gas inversion. Frequently, the prior expectation of the flux process is set from one or more inventories that are many times unreliable and that may have unquantifiable effects on a posterior assessment. Here, we only use *characteristics* of the inventories, namely the spatial length scales and the marginal variance, whilst setting the prior expectation to be spatially constant. In this way, the inventory fluxes are not used directly in the assimilation, and the spatial distribution of fluxes obtained from the posterior expectation will be predominantly data-driven. Our contribution addresses a concern in [13], Section 5.2, that suggests that such an approach is difficult when prior distributions are non Gaussian.

The article is structured as follows. In Section 2, we discuss the three contributions outlined above. In Section 3, we detail our approach to inferring the fields of interest using a combination of approximate inferential methods. The proposed framework is then assessed in Section 4, first in a study in one-dimensional space with simulated datasets, and then for emissions prediction in the UK and Ireland using the four measurement stations illustrated in Fig. 1, top-right panel. Section 5 contains conclusions and an outline of future research directions.

2. Theory

The notation we use corresponds to that commonly found in the spatial-statistics literature (e.g., [14]). Here, stochastic processes are denoted using regular typeface, whilst bold typeface is used to denote