Chemometrics and Intelligent Laboratory Systems 146 (2015) 407–417

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Optimal designs for the methane flux in troposphere

Sándor Baran [a], Kinga Sikolya [a], Milan Stehlík [b,c,*]

[a] Faculty of Informatics, University of Debrecen, Kassai út 26, H-4028 Debrecen, Hungary
[b] Institut für Angewandte Statistik, Johannes Kepler University in Linz, Altenberger Straße 69, A-4040 Linz, Austria
[c] Departamento de Matemática, Universidad Técnica Federico Santa María, Casilla 110-V, Valparaíso, Chile

ARTICLE INFO

Article history:
Received 7 September 2014
Received in revised form 2 June 2015
Accepted 10 June 2015
Available online 18 June 2015

Classification codes:
Primary 62K05
Secondary 62M30

Keywords:
Arrhenius model
Bias reduction
Filling designs
Integrated mean square prediction error
Ornstein–Uhlenbeck sheet
Tropospheric methane

ABSTRACT

The understanding of methane emission and methane absorption plays a central role both in the atmosphere and on the surface of the Earth. Several important ecological processes, e.g., ebullition of methane and its natural microergodicity request better designs for observations in order to decrease variability in parameter estimation. Thus, a crucial fact, before the measurements are taken, is to give an optimal design of the sites where observations should be collected in order to stabilize the variability of estimators. In this paper we introduce a realistic parametric model of covariance and provide theoretical and numerical results on optimal designs. For parameter estimation D-optimality, while for prediction integrated mean square error and entropy criteria are used. We illustrate applicability of obtained benchmark designs for increasing/measuring the efficiency of the engineering designs for estimation of methane rate in various temperature ranges and under different correlation parameters. We show that in most situations these benchmark designs have higher efficiency.

1. Introduction

The understanding of methane emission and methane absorption plays a central role both in the atmosphere (for troposphere see, e.g., [30]) and on the surface of the Earth (see, e.g., [19] regarding the methane emissions from natural wetlands and references therein or [13] for efficient and robust model of the methane emission from sedge–grass marsh in South Bohemia). Several important ecological processes, e.g., ebullition of methane and its natural microergodicity request better designs for observations in order to decrease variability in parameter estimation [14]. In this context by a design we mean a set of locations where the investigated process is observed. Thus, a crucial fact, before the measurements are taken, is to give an optimal design of the sites where observations should be collected. Rodríguez-Díaz et al. [25] provided a comparison of filling and D-optimal designs for a one-dimensional design variable, e.g., temperature. However, such a model oversimplifies the important fact that variation of other variables, e.g., rates k_1 of the considered modified Arrhenius model, could disturb the efficiency of the learning process. The latter statement is also in agreement with common sense in physical chemistry. In this paper the difficulties of modeling and design are treated, mainly by allowing an Ornstein–Uhlenbeck (OU) sheet error model.

We concentrate on efficient estimation of the parameters of the modified Arrhenius model (model popular in chemical kinetics), which is used by Vaghjiani and Ravishankara [30] as a flux model of methane in troposphere. This generalized exponential (GE) model can be expressed as

$$Y = Ax^{\mu}e^{-Bx} + \epsilon = \eta(x, \mu, B) + \epsilon, \tag{1.1}$$

where $A, B, \mu \in \mathbb{R}$, $A, B \geq 0$, are constants and ε is a random error term. In the case of correlated errors such a model was studied by Rodríguez-Díaz et al. [25], however, in that work error structures were univariate stochastic processes. In [24,25] the authors concentrated on the Modified-Arrhenius (MA) model, which is equivalent to the GE model through the change of variable $x = 1/t$. This model is useful for chemical kinetic mainly because it is a generalization of Arrhenius model describing the influence of temperature t on the rates of chemical processes, see, e.g., [17] for general discussion and [23] for optimal designs. However, for specific setups, for instance, long temperature ranges, Arrhenius model is insufficient and the MA (or GE model) appears to be the good alternative (see for instance [10]). Other applications of model (1.1) in chemistry are related to the transition state theory (TST) of chemical reactions [12].

* Corresponding author at: Institut für Angewandte Statistik, Johannes Kepler University in Linz, Altenberger Straße 69, A-4040 Linz, Austria. Tel.: +43 732 2468 6806, fax: +43 732 24689846.
E-mail address: Milan.Stehlik@jku.at (M. Stehlík).

http://dx.doi.org/10.1016/j.chemolab.2015.06.002
0169-7439/

In practical chemical kinetics two steps are taken: first the rates k_1 are estimated (typically with symmetric estimated error) and then MA is fitted to the rates, i.e.,

$$k_1 = A(1/t)^{\mu} e^{-B/t} + \tilde{\varepsilon}(t). \tag{1.2}$$

Statistically correct would be to assess both steps by one optimal experimental planning. Rodríguez-Díaz et al. [25] concentrated on the second phase, i.e., what is the optimal distribution of temperature for obtaining statistically efficient estimators of trend parameters A, B, μ and correlation parameters of the error term $\tilde{\varepsilon}$. In this paper we provide designs both for rates and temperatures, and in this way substantially generalize the previously studied model.

Correlation is the natural dependence measure fitting for elliptically symmetric distributions (e.g., Gaussian). By taking s (this variable can play, for example, the role of atmospheric pressure, latitude or location of the measuring balloon in troposphere, either vertically or horizontally) and temperature t to be variables of covariance, our model (1.1) takes a form of a stationary process

$$Y(s,t) = k_1 + \varepsilon(s,t), \tag{1.3}$$

where the design points are taken from a compact design space $\mathcal{X} = [a_1, b_1] \times [a_2, b_2]$, with $b_1 > a_1$ and $b_2 > a_2$, and $\varepsilon(s,t)$, $s, t \in \mathbb{R}$, is a stationary OU sheet, that is a zero mean Gaussian process with covariance structure

$$\mathrm{E}\varepsilon(s_1,t_1)\varepsilon(s_2,t_2) = \frac{\tilde{\sigma}^2}{4\alpha\beta} exp(-\alpha|s_1 - s_2| - \beta|t_1 - t_2|), \tag{1.4}$$

where $\alpha > 0, \beta > 0, \ \tilde{\sigma} > 0$. We remark that $\varepsilon(s,t)$ can also be represented as

$$\varepsilon(s,t) = \frac{\tilde{\sigma}}{2\sqrt{\alpha\beta}} e^{-\alpha s - \beta t} \mathcal{W}\left(e^{2\alpha s}, e^{2\beta t}\right),$$

where $\mathcal{W}(s,t)$, $s, t \in \mathbb{R}$, is a standard one-dimensional Brownian sheet [4,5]. Covariance structure (1.4) implies that for $\mathbf{d} = (d, \delta)$, $d \geq 0$, $\delta \geq 0$, the variogram $2\gamma(\mathbf{d}) := \mathrm{Var}(\varepsilon(s + d, t + \delta) - \varepsilon(s,t))$ equals

$$2\gamma(\mathbf{d}) = \frac{\tilde{\sigma}^2}{2\alpha\beta}\left(1 - e^{-\alpha d - \beta\delta}\right)$$

and the correlation between two measurements depends on the distance through the semivariogram $\gamma(\mathbf{d})$.

As can be visible from relation (1.2) between rates and parameters A, μ and B of the MA model, the second variable s is missing from trend since it is not chemically understood as driving mechanism of chemical kinetics, however, in this context it is an environment variable.

In order to apply the usual notations of spatial modeling [16] we introduce $\sigma := \tilde{\sigma}/(2\sqrt{\alpha\beta})$ and instead of Eq. (1.4) we investigate

$$\mathrm{E}\varepsilon(s_1,t_1)\varepsilon(s_2,t_2) = \sigma^2 exp(-\alpha|s_1 - s_2| - \beta|t_1 - t_2|), \tag{1.5}$$

where σ is considered as a nuisance parameter.

We remark that in order to reduce the length of the paper proofs of all theorems presented here together with calculations corresponding to Examples 2.8 and 3.5 are given in a separate Supplementary section which is available on the website of the publisher. These details can also be found in [6].

2. Benchmarking grid designs for the OU sheet with constant trend

In this section we derive several optimal design results for the case of constant trend and regular grids resulting in a Kronecker product covariance structure. These theoretical contributions will serve as benchmarks for optimal designs in a methane flux model. Thus we consider the stationary process

$$Y(s,t) = \theta + \varepsilon(s,t) \tag{2.1}$$

with the design points taken from a compact design space $\mathcal{X} = [a_1, b_1] \times [a_2, b_2]$, where $b_1 > a_1$ and $b_2 > a_2$ and $\varepsilon(s,t)$, $s, t \in \mathbb{R}$, are a stationary Ornstein–Uhlenbeck sheet, i.e., a zero mean Gaussian process with covariance structure (1.5).

2.1. D-optimality

As a first step we derive D-optimal designs, that is arrangements of design points that maximize the objective function $\Phi(M) := det(M)$, where M is the Fisher information matrix of observations of the random field Y. This method, "plugged" from the widely developed uncorrelated setup, is offering considerable potential for automatic implementation, although further development is needed before it can be applied routinely in practice. Theoretical justifications of using the Fisher information for D-optimal designing under correlation can be found in [1,22,29].

We investigate grid designs of the form $\{(s_i, t_j) : i = 1, 2, \ldots, n, j = 1, 2, \ldots, m\} \subset \mathcal{X} = [a_1, b_1] \times [a_2, b_2]$, $n, m \geq 2$, and without loss of generality we may assume $a_1 \leq s_1 < s_2 < \ldots < s_n \leq b_1$ and $a_2 \leq t_1 < t_2 < \ldots < t_m \leq b_2$. Usually, the grid containing the design points can be arranged arbitrary in the design space $\mathcal{X}$, but we also consider restricted D-optimality, when $s_1 = a_1$, $s_n = b_1$ and $t_1 = a_2$, $t_m = b_2$, i.e., the vertices of $\mathcal{X}$ are included in all designs.

2.1.1. Estimation of trend parameter only

Let us assume first that parameters α, β and σ of the covariance structure (1.5) of the OU sheet ε are given and we are interested in estimation of the trend parameter θ. In this case the Fisher information on θ based on observations $\{Y(s_i, t_j),\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m\}$ equals $M_\theta(n, m) = \mathbf{1}_{nm}^{\top} C^{-1}(n, m, r)\mathbf{1}_{nm}$, where $\mathbf{1}_k$, $k \in \mathbb{N}$, denotes the column vector of ones of length k, $r = (\alpha, \beta)^T$, and $C(n, m, r)$ is the covariance matrix of the observations [22,31]. Further, let $d_i := s_{i+1} - s_i$, $i = 1, 2, \ldots, n - 1$, and $\delta_j := t_{j+1} - t_j$, $j = 1, 2, \ldots, m - 1$, be the directional distances between two adjacent design points. With the help of this representation one can prove the following theorem.

Theorem 2.1. *Consider the OU model (2.1) with covariance structure (1.5) observed in points $\{(s_i, t_j),\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m\}$ and assume that the only parameter of interest is the trend parameter θ. In this case*

$$M_\theta(n, m) = \left(1 + \sum_{i=1}^{n-1} \frac{1 - p_i}{1 + p_i}\right)\left(1 + \sum_{j=1}^{m-1} \frac{1 - q_j}{1 + q_j}\right), \tag{2.2}$$