Entropy based statistical inference for methane emissions released from wetland

Sabolová, R.[a], Sečkárová, V.[b,a], Dušek, J.[c], Stehlík, M.[d,e,*]

[a] *Department of Probability and Mathematical Statistics, Charles University, Czech Republic*
[b] *Institute of Information Theory and Automation, Czech Republic*
[c] *Global Change Research Centre AS CR, České Budějovice, Czech Republic*
[d] *Department of Applied Statistics, Johannes Kepler University in Linz, Austria*
[e] *Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile*

Abstract

We concentrate on paradigm question how much stochasticity and how much chaos is present in the methane emission model. In particular we analyze the residua from the process of methane emissions from wetlands in the sedge-grass marsh, in South Bohemia, Czech Republic. Relation to entropy and a specific version of Kullback-Leibler divergence will be given. Graphical tool to assess the amount of entropy in the system is developed and illustrated on real data from the sedge-grass marsh methane emission.

Keywords: chaos, entropy, Kullback-Leibler (KL) divergence, Pareto distribution, saddlepoint approximation, wetland ecosystem

1. Introduction

According to [2] life in general can be characterized as a stream of order. Biogeochemical processes on the different time scales are not moving towards the direction from disorder to order and change of entropy has thermodynamic meaning ([2], [3]). Highly organized structures are gradually decomposed up to individual organic and inorganic compounds (disorder) at the end of life of individual organisms (or ecosystem as the super organism). Real life proceeds in many different cycles. These include important cycles of biogenic compounds/elements such as carbon, nitrogen, phosphorus and water. Other necessary elements and microelements are also (re)cycled [4]. Cycling and continuous re-cycling of elementary compounds does occur due to often sharp limitation by nutrients (macro elements) and microelements [4]. Most ecosystems must be very efficient to manage them. Methane is a product of anaerobic decomposition processes of the organic matter cycles mainly in water-saturated soils. These processes require the synergy or syntrophic cooperation between anaerobic bacteria and methanogenic archaea. Organic matter is firstly hydrolysed and fermented, and the products that are formed are the compounds used for methanogenesis [5]. Methane production can be described as dissipative process of entropy in which highly organized

*Corresponding author. Departamento de Matemática, Universidad Técnica Federico Santa María, Casilla 110-V, Valparaíso, Chile. Email: Milan.Stehlik@jku.at Tel.:+4373224686808; Fax: +4373224686800
Email addresses: sabolova@karlin.mff.cuni.cz (Sabolová, R.), seckarov@utia.cas.cz (Sečkárová, V.), dusek.j@czechglobe.cz (Dušek, J.), Milan.Stehlik@jku.at (Stehlík, M.)

organic structures are decomposed to basic simple compounds. The process of releasing the methane from soil and plant stand is highly determined by its production. Methane production depends on the occurrence of methanogenic bacteria, the amount of decomposed organic matter and suitable anaerobic conditions. Methane releasing from deeper soil layers is influenced by many processes, including methane oxidation by the methanetrophic bacteria [6] and also by its means of being released from the soil environment.

Methane emissions are typically modelled via trend process fitting, which makes the amount of stochasticity and chaos present in the system questionable. In [1] we have modelled this dependence by a time series model. The trend component is estimated by the Ordinary Least Squares technique. The noise component is represented by sum of an infinite moving average model with Pareto-like positive and negative parts of the innovations and independent identically distributed (i.i.d.) innovations with similar tail behaviour. Pareto tails have been also justified by robust tests for normality against Pareto tails (see [7]). Such moving average time series could be considered as born by a point process, which is not homogeneous (see [8]). The process of methane release from soil is both chaotic and stochastic. In [1] we outlined the relation between stochastic and chaotic model: The parameters typically associated with chaos (both deterministic and stochastic) are measures of dimension, rate of information generated (entropy), and the Lyapunov spectrum. Entropies, as a measure of self-organization (or level of chaos of the system) in our case correspond to heavy tail parameters (in our stochastic model these are Pareto tails of lower and upper exceedances over thresholds, respectively). In [1] we have observed large tail parameters of the underlying stochastic process for both upper and lower exceedances. This confirms that the amount of chaos measured through correlation dimension or entropy is not large.

The behaviour of methane-producing bacteria is highly self-organized. Nevertheless, there is a vast amount of randomness in the process of methane flux from bacteria's point of view, as they produce the methane into an external space through a biological process. We tried to model the methane exchange process between the soil covered by plants and the surrounding atmosphere. The emission of methane or other gases from bare soils and plant stands is a result of many partial processes that continuously happen inside soils ([9]). These processes are often classified as deterministic or stochastic in an appropriate scale (see [3]). From the point of view of the methanogenic bacteria the occurrence of methane in the soil is a dissipative process determined by their metabolisms. Bacteria are also a highly self-organized system and the amount of chaos in the soil system is not too large. Biological processes in soil are strictly determined by physical and environmental parameters with stochastic behaviour (see [10]).

As we have already mentioned the Pareto tails are proved to be suitable for modelling the release of methane by bacteria. To measure the amount of unpredictability or chaos involved in such system we exploit the information entropy (2). If we assume strict parametric assumption of the Pareto distribution $Pareto(x_m, \alpha)$ with scale parameter x_m and shape parameter α having the probability density function (pdf)

$$f_X(x) = \alpha \frac{x_m^\alpha}{x^{\alpha+1}}, \quad x_m > 0, \quad \alpha > 0$$

then we can derive the theoretical relationship between the tail parameter and its entropy (see [11]):

$$H_X = \ln\left(\frac{x_m}{\alpha}\right) + \frac{1}{\alpha} + 1. \tag{1}$$

This is always decreasing function for increasing α independent of the value of x_m:

$$\frac{\partial H_X}{\partial \alpha} = -\frac{1}{\alpha} - \frac{1}{\alpha^2} < 0 \quad (\alpha > 0).$$

In [1] we have observed large values of tail parameters α confirming moderate level of chaos in the system, justifying underlying biochemical intuition.

If only semiparametric/nonparametric approach to tails is available, we shall rely on extreme value theory or dynamical systems, or their interplays (see [12]). From biochemical point of view we can define three main ways in which the methane moves from below ground soil layers to the atmosphere: 1) Methane emissions from the bare soil surface, 2) Methane emissions through plant tissue (plant growing on the methane source surfaces), 3) Ebullition from methane source surfaces (Spontaneous releasing of methane mainly from the flooded sediments). The competition between three above-mentioned possible means introduces naturally dynamic system which can have chaotic behavior.

A chemometrician having this data will need some tool to measure the amount of chaos vs. stochasticity in his model, therefore the I-divergence construction shall be of great help. I-divergence is based on KL-divergence, called also cross-entropy, which already includes the amount of chaos, for detailed description see [13], [14]. It also reflects the amount of dissimilarity from one distribution to another, which is useful for testing in such cases as, for example, when the expert's opinion and hypothesis about the unknown model are available.

The methodological novelty of the current paper is that we provide convenient analytical approach to empirically assess the amount of chaos in methane releasing measured via tail entropy. We construct a graphical tool to check for fit of the models based on I-divergence deconvolution. The relationship using the KL-divergence has been studied in [15], from a Bayesian point of view. A Jeffrey's prior is used to obtain posterior predictive distribution of data. Then the I-divergence is plotted. In this paper we develop testing tool based on I-divergence and tool measuring the dissimilarity for different scenarios in different parameter setups. Having an exact Pareto model and knowing its threshold it is possible to simply transform the Pareto variable into exponential one and therefore the natural start of this exposition is the exponential distribution setup. In [16] the exact distributions have been derived for the setup. In the Appendix we provide relations to other approaches via I-divergence and give the formulas used in text.

2. Empirical testing of entropy for the exponential case

Let us assume that observations $\{y_i\}_{i=1}^N$ are distributed according to the exponential distribution $Exp(\gamma_i)$ with pdf $h(y_i|\gamma) = \gamma e^{-\gamma y_i}$, $y_i \geq 0$.

For testing the hypothesis about the unknown parameter $\gamma = (\gamma_1, \ldots, \gamma_N)$

$$H_0 : \gamma = \gamma_0 \quad (H_1 : \gamma \neq \gamma_0)$$

based on available observations $y = (y_1, \ldots, y_N)$ we exploit the I-divergence introduced in [16]:

$$I_N(y, \gamma_0) = N \ln \sum_i y_i - \sum_i \ln y_i + \gamma_0 \sum_i y_i - N \ln(\gamma_0 \sum_i y_i) - N$$

The KL-divergence forming $I_N(y, \gamma_0)$ measures the divergence of the exponential distribution pdf with vector of MLE estimates $(\hat{\gamma}_1, \ldots, \hat{\gamma}_N)$ from the pdf of the exponential with vector of hypotheses $(\gamma_0, \ldots, \gamma_0)$.

2.1. Graphical method for testing $H_0 : \gamma = \gamma_0$ in the exponential distribution

For testing the hypothesis $H_0 : \gamma = \gamma_0$ we introduce a graphical method based on the cumulative distribution function (cdf). We generate M random samples from $Exp(\gamma_0)$ each of sample size N and compute the $I_N(y, \gamma_0)$. We plot the quantiles for the cdf of $I_N(y, \gamma_0)$ against the quantiles for cdf of uniform distribution $U(0, 1)$. If $\{y_i\}_{i=1}^N$ is distributed according to $Exp(\gamma_0)$, the result should form a diagonal in this plot. Thus we do not reject the hypothesis when the points lie on or nearby the diagonal. The more the results differ from a diagonal, the more likely we are going to reject H_0.

For the exponential case, the exact form of cdf for I_N, $N \leq 4$ is derived in [17]. For $N = 1$ and $y \sim Exp(\gamma)$ the cdf of $I_1(\hat{\gamma}, \gamma)$ is equal to

$$F_1(x) = \begin{cases} \exp\{\hat{\gamma}^{-1}\gamma W_0(-\exp\{-1-x\}\} - \exp\{\hat{\gamma}^{-1}\gamma W_{-1}(-\exp\{-1-x\}\} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0, \end{cases}$$

where W_0, W_{-1} are the two real-valued branches of Lambert-W function. Quantile plot of $F_1(I_1)$ under $H_0 : \gamma = 1$ against quantiles of uniform distribution U(0,1) is displayed in Fig. 1. In this example we generated $M = 1000$ random samples each consisting of 1 random variable from the distribution $Exp(1)$. The plotted points form the diagonal in the plot, thus we do not reject the hypothesis.

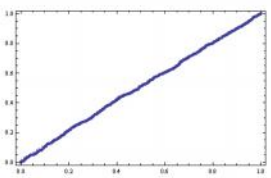


Figure 1: Simulations for exponential case with exact F_1, number of samples 1000 and $\gamma = 1$

As formulas for $N \geq 2$ are rather complicated, we exploit the deconvolution of I_N to two independent parts R_N and S_N as follows:

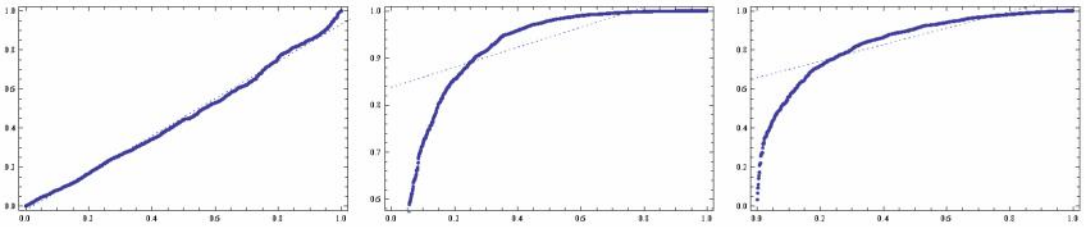
$$I_N(y, \gamma_0) = R_N + S_N$$

$$R_N = \gamma_0 \sum_i y_i - N \ln(\gamma_0 \sum_i y_i) - N + N \ln N$$
$$S_N = N \ln \sum_i y_i - N \ln N - \sum_i \ln y_i.$$

To construct the cdf of I_N and thus the quantile plot we use asymptotic distribution of R_N and S_N. According to [16], R_N is asymptotically χ^2_1-distributed. Asymptotic distribution of S_N was derived in [18] and is equal to

$$\frac{1}{2}\left(1 + \frac{1 + \frac{1}{N}}{6}\right)\chi^2_{N-1}.$$

Quantile plots of $\hat{F}_N(I_N)$ for $\hat{F}_N$ being the asymptotic cdf of I_N for $M = 1000$ samples of size $N = 50$ from $Exp(1)$ against quantiles of uniform distribution U(0,1) for different hypotheses $H_0 : \gamma = \gamma_0$ can be found in Fig. 2a, 2b and 2c.



(a) $\gamma = 1$, $\gamma_0 = 1$ (b) $\gamma = 1$, $\gamma_0 = 2$ (c) $\gamma = 1$, $\gamma_0 = \frac{1}{2}$

Figure 2: Simulations for exponential case for 1000 samples with $N = 50$.

In case when the data were drawn from $Exp(1)$ and the hypothetical value for the unknown parameter γ was set to 1, the plotted points form a diagonal in the quantile plot for asymptotic cdf of I_N against cdf of uniform distribution $U(0, 1)$. When testing the hypothesis that the unknown parameter γ equals 2 or $1/2$ (see Fig. 2b and 2c), the results in the plots differ significantly from the straight line forming diagonal. Thus we reject hypotheses in both cases.

In this section the basics of graphical method for testing hypotheses about unknown parameters based on I-divergence were introduced. Since no real data was available in the case of exponential distribution (tails in methane emission were heavier, see [1]) we generated the data samples from exponential distribution with various choices of parameter γ.

3. Empirical testing for the Pareto case

Based on results in previous sections, we will now propose a graphical procedure for testing the hypothesis about the parameter α in Pareto distribution. For real data assumed to be Pareto distributed the pdf does not depend not only on this unknown parameter γ, but also on the value of x_m. Thus we first focus on the case where the value of x_m is known. Later on we proceed to more sophisticated method for unknown value of x_m.

3.1. Graphical procedure for known value of x_m

Let us now have N independent observations $\{y_i\}_{i=1}^N$ from Pareto distribution with shape and scale parameters x_m and $\alpha_i = -\gamma_i - 1$, $i = 1, \ldots, N$, where x_m is known. The I-divergence then looks as follows:

$$I_N(y, \gamma_0) = \sum_{i=1}^{N} \left(\frac{1}{\ln x_m - \ln y_i} - 1 \right) \ln(y_i) + \sum_{i=1}^{N} \ln \left\{ \left(-\frac{1}{\ln x_m - \ln y_i} \right) x_m^{\left(-\frac{1}{\ln x_m - \ln y_i} \right)} \right\}$$
$$- \gamma_0 \sum_i \ln(y_i) - \sum_i \ln \left\{ (-\gamma_0 - 1) x_m^{(-\gamma_0 - 1)} \right\}$$

With the use of the above I-divergence we now introduce the hypothesis testing for Pareto distributed data based on data y. Here, the monotonicity of this I-divergence depends on the value of γ_0. For the fixed value of hypothesis γ_0 in testing hypothesis

$$H_0 : (\gamma_1, \ldots, \gamma_N) = (\gamma_0, \ldots, \gamma_0)$$

against alternative

$$H_1 : (\gamma_1, \ldots, \gamma_N) \neq (\gamma_0, \ldots, \gamma_N)$$

we can compute the exact value of the I-divergence and conclude whether we accept or reject the hypothesis.

To help the chemometrician to distinguish between these two results we introduce a graphical method for testing the above hypothesis. We exploit the uniform transform of cumulative distribution function. Then we can compare the cumulative distribution function of the I-divergence against the cumulative distribution function of the continuous uniform distribution $U(0,1)$. If the plotted values form a diagonal in quadrat $[0,1] \times [0,1]$ thus the data are likely to origin in the distribution $Pareto(x_m, \alpha_0)$ and we accept the hypothesis. If the plotted values differ from the diagonal, the I-divergence reaches more often higher values, then the $Pareto(x_m, \alpha_0)$ do not describe our data well. Thus we reject the hypothesis H_0.

To illustrate the method we generate a random sample $z_1, \ldots, z_N$ of size N from Pareto distribution with known x_m and hypothetical γ_0 representing a desired or expected value of the parameter chosen by experimentator. We compute the I-divergence for $z_1, \ldots, z_N$ and define the "empirical distribution function" of $I_N(z_1, \ldots, z_N; \gamma_0)$ as

$$\hat{F}_{I_N}(z) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{I}[z_i \leq z].$$

For given data $y_1, \ldots, y_N$ one should compute $\hat{F}_{I_N}(y_i), i = 1, \ldots, N$ and construct a quantile plot of these values against uniform distribution $U(0,1)$. Under the null hypothesis the points are supposed to form a straight line connecting points $(0,0)$ and $(1,1)$, see Fig. 3 (for general discussion on QQ Plots see e.g. [19]). Under alternative hypothesis the points decline from the straight line towards axis x or y, see Fig. 4 and Fig. 5.

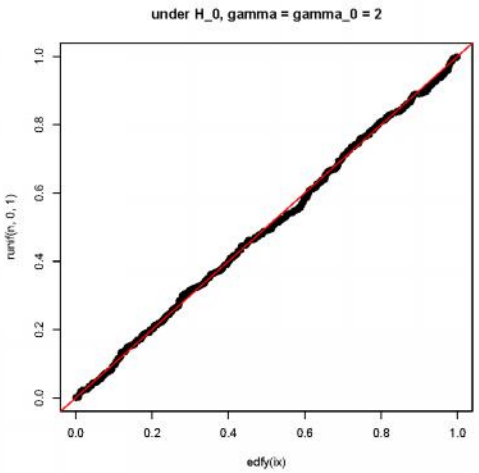


Figure 3: Simulations for Pareto case: under H_0

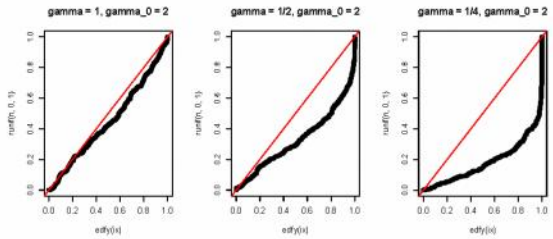


Figure 4: Simulations for Pareto case: H_1, real value of α is less than α_0

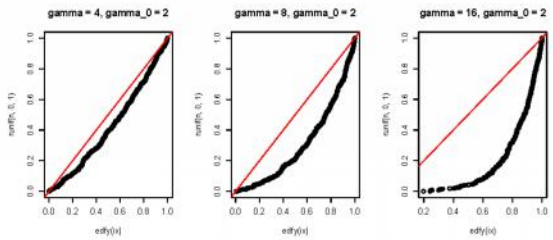


Figure 5: Simulations for Pareto case: H1, real value of α is greater than α_0

3.2. Graphical procedure for unknown value of x_m

For real data on methane emissions the value of threshold x_m is usually not provided explicitly. We shall treat x_m as a nuisance parameter, which is however substantial for interpretation since it distinguishes between normal diffusion and anomalous diffusion (chaos). Our conjecture for the source of the chaos are various interactions (e.g. way of escaping of methane). The process of diffusion and other non-specific ways of methane releasing (e.g. ebullition) occur simultaneously. Normal diffusion relates to stochasticity, non-specific ways of releasing of methane relate to chaotical behavior of the system.

Both parts, stochastic and chaotic are hardly separable. Therefore the simulated process compared via I-divergence to real data process can be taken only in upper tails of real data. Thus peaks of methane emissions can correspond to various spontaneous releases of methane which we understand to be chaotic in its nature.

3.2.1. Real methane data

We will analyze the residuals $Z, Z-, Z+$ of methane emissions taken from infinite moving average model (8), see [1], where only time is taken as a regressor. We trimmed the original data sets with sample sizes $M_{Z+} = 998$, $M_{Z-} = 971$ by 30%, thus we obtained Pareto distributed samples of sample sizes M_+, M_-.

Before we proceed to test the null hypothesis about the parameter $\gamma = (\gamma_1, \ldots, \gamma_N)$ it is necessary to deal with the unknown nuisance parameter x_m. To obtain the desired value of x_m we exploit a maximum entropy principle [20], [21]. This principle states that from the set of all possible distributions, now represented by a set of Pareto distributions with fixed α and unknown x_m, we should choose the distribution with the highest entropy. For fixed γ, ($\gamma = -\alpha - 1$), and x_m on the

interval $(0, y_{min})$, where y_{min} is the minimal value from the sample $y_1, \ldots, y_N$, the entropy is an increasing function of x_m, see (1). Thus, the value of x_m chosen for the proposed graphical method should be close to the minimal value of the sample $y_1, \ldots, y_N$. Since data are Pareto distributed, $x_m < y_i \quad \forall i$, we set $x_m = 0.99 \times y_{min}$.

Let us still assume the shape parameter α is known. Thus the uncertainty of the current system, represented by entropy (1), now depends on the value of the scale parameter x_m. If the chemometrician is expecting the system to be more deterministic, one should choose the value of x_m close to zero. This coincides with the fact that then the exceedances over threshold x_m are assigned with lower probability. On the other hand, if the system can be more chaotic, that is the exceedances over the threshold x_m occur with high probability, the value of x_m should be close to the minimal value of the sample $y_1, \ldots, y_N$. For our Pareto data $Z+$ with $\alpha = 1.3$ and the minimal value of the sample 2.05×10^{-6} the relation between entropy and value of x_m is described by (1). For lower values of entropy the value of x_m is really close to zero, the maximum achievable entropy is obtained for x_m close to y_{min}.

For chosen hypotheses $H_0 : \gamma = \gamma_0$ we generated hypothetical Paretian sample $\{z_i\}_{i=1}^K$ with parameters $Pareto(x_m, \alpha_0)$, where $\alpha_0 = -\gamma_0 - 1$. We also computed corresponding I-divergences $I_1(z_i, \gamma_0)$, based on which we estimated the empirical cumulative distribution function $\hat{F}_{I_1}$. Then we plotted the quantiles of ecdf $\hat{F}_{I_1}$ evaluated in points $I_1(y_i, \gamma_0)$, $i = 1, \ldots, M_+$ or M_- against quantiles of uniform distribution $U(0, 1)$.

Previous data analyses suggested value $\alpha_0 = 1.3$ for trimmed data set $Z+$ and $\alpha_0 = 1.2$ for trimmed data set $Z-$. Proposed graphical method confirms these hypotheses as we can see in Fig. 6a and Fig. 6b.

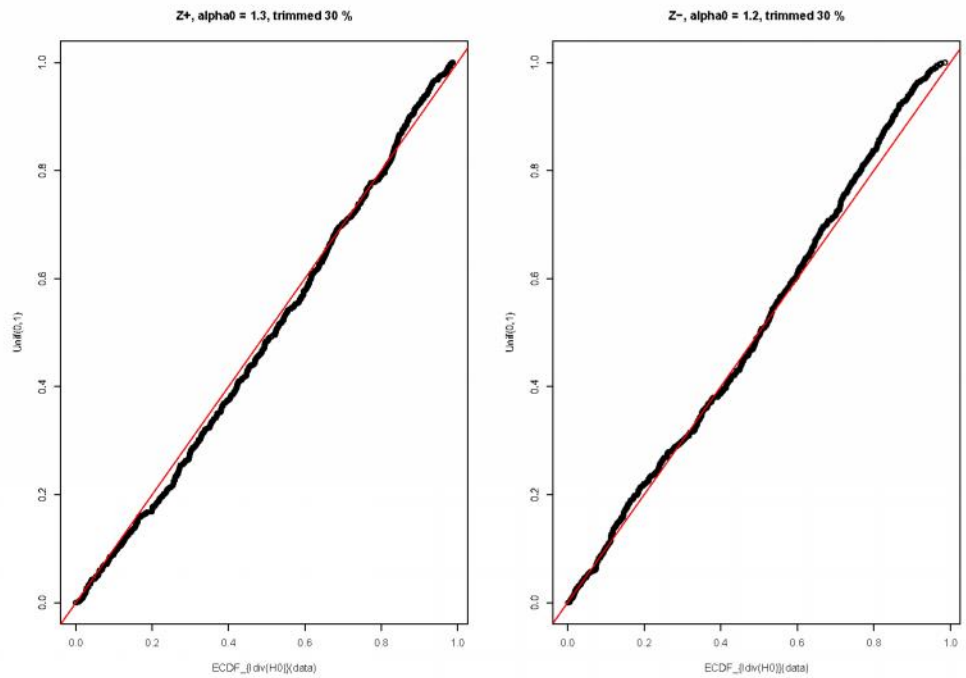
4. Saddlepoint approximations for the density of MLE

Apart from testing hypotheses about unknown parameter γ, another area of interest is the distribution of maximum likelihood estimator of the tail parameter for Pareto distribution. As it is usually not possible to derive analytical formula for the density of an estimator, various approximations might be used. One of the widely used approaches is central limit theorem and asymptotic theory for maximum likelihood estimators based on CLT. Unfortunately, approximations for the density based on CLT tend to be inaccurate in the tails of the distribution. As saddlepoint techniques proved to be very useful in this area, we will now concentrate on this approach. In this section we will present saddlepoint approximation for the density of MLE based on I-divergence and accompany theoretical results with plots for two-dimensional case using real data from previous section.

In order to derive a saddlepoint approximation it is first necessary to compute density of a sufficient statistic t, that for Pareto distribution has the form $t = \log y$. Using results from [22] we get the saddlepoint approximation for the density of t

$$q_T(t|\gamma) = (2\pi)^{-N/2} \prod_i (-\gamma_i - 1) x_m^{-\gamma_i - 1 + \frac{1}{\ln x_m - t_i}} \exp\left\{ t_i \left(\gamma_i - \frac{1}{\ln x_m - t_i} + 1 \right) \right\}.$$

Saddlepoint aproximation for the density of MLE is based on the density of the sufficient statistic



(a) Pareto methane data - Z+ (b) Pareto methane data - Z-

Figure 6: Pareto real data case: $\gamma_0 = 1.3$ for Z+, $\gamma_0 = 1.2$ for Z-.

(see [22])

$$\hat{\gamma}_i = \frac{1}{\ln x_m - t_i} - 1, \qquad i = 1, \ldots, N$$

and its formula reads

$$q_T(\hat{\gamma}|\gamma) = (2\pi)^{-N/2} \prod_i (\ln x_m - t_i)^2 (-\gamma_i - 1) x_m^{-\gamma_i - 1 + \frac{1}{\ln x_m - t_i}} \exp\left\{ t_i \left(\gamma_i - \frac{1}{\ln x_m - t_i} + 1 \right) \right\} |_{t=t(\hat{\gamma})}$$
$$= (2\pi)^{-N/2} \prod_i \frac{-\gamma_i - 1}{(\hat{\gamma}_i + 1)^2} x_m^{-\gamma_i + \hat{\gamma}_i} \exp\left\{ (\gamma_i - \hat{\gamma}_i) \left(\ln x_m - \frac{1}{\hat{\gamma}_i + 1} \right) \right\}.$$

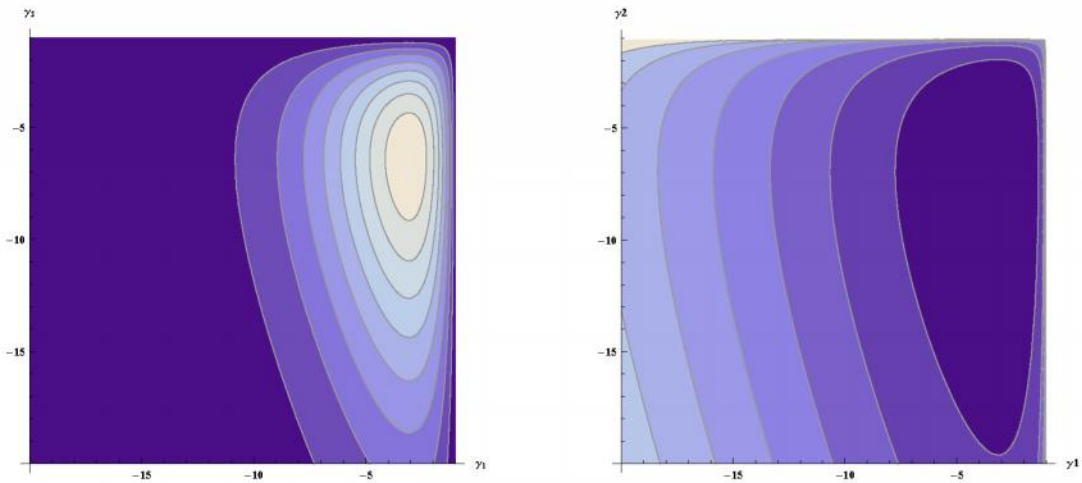
4.1. Application to real data - Pareto

In order to illustrate the applicability of the above derived formulas we exploit the trimmed data $Z+$ and $Z-$ used in Subsection 3.2.1.

Firstly, we concentrate on the density of MLE estimate and the I-divergence for trimmed data $Z+$. We will study the form of saddlepoint approximation of the density of MLE and I-divergence as

a function of 2 variables (we will deal with two Pareto distributed random variables with parameters γ_1, γ_2) based on available observations. Since we have derived saddlepoint approximation for the density of MLE based on I-divergence, it is not surprising that both can be used in inference on hypothetical value of parameter. The only difference between saddlepoint approximation for the sufficient statistic and MLE is a multiplicative constant, therefore we will not plot the approximation for the density of a sufficient statistic.

We will consider two independent Pareto distributed variables y_1, y_2 with scale parameter $x_m = 2.03 \times 10^{-6}$. To study aforementioned properties with respect to changes in the unknown parameters γ_i, $i = 1, 2$, we randomly chose two observations $y_1 = 1.45 \times 10^{-5}$ and $y_2 = 4.24 \times 10^{-6}$ within the admissible area: $\gamma_i < -1$, $i = 1, 2$. Results are shown in Fig. 7a and 7b.



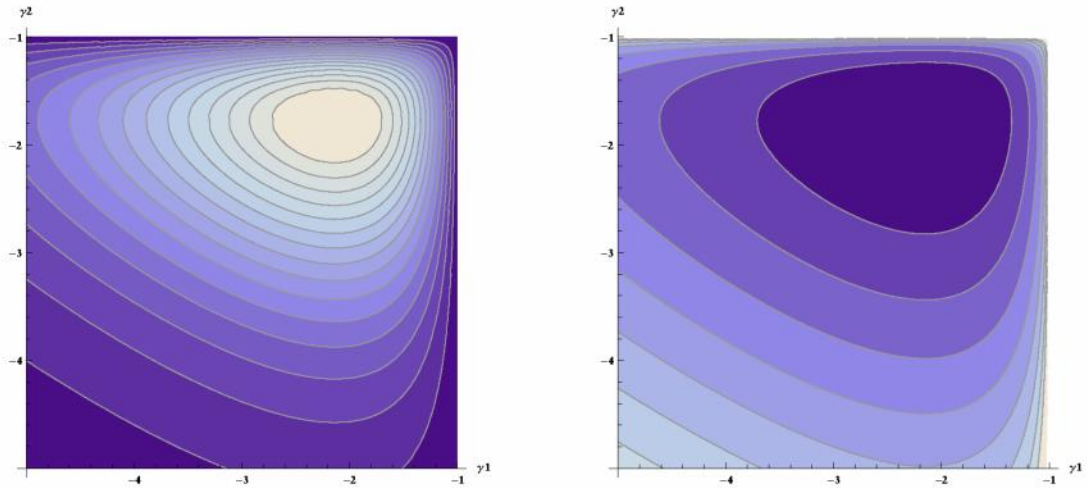
(a) Density of MLE (b) I-divergence

Figure 7: Pareto real data Z+ with saddlepoint densities for MLE estimate and I-divergence.

In case of the I-divergence $I_2(\hat{\gamma}_y, \gamma)$, we expect that as a function of two variables it reaches its minimum value, zero, for the arguments: $\hat{\gamma}_i = \frac{1}{\ln y_i - \ln x_m} - 1$, $i = 1, 2$ (see properties of the KL-divergence). Thus, in this case in the point $\hat{\gamma}_1 = -1.51$ and $\hat{\gamma}_2 = -2.36$, which can be seen in 7b also brings the view on the changes of the I-divergence with respect to γ_1, γ_2. We can see that plotted I-divergence reaches minimum value for values $\hat{\gamma}_1$ and $\hat{\gamma}_2$. From the principle of maximum likelihood, we expect that the approximated density reaches its maximum values for these values of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ (see figures).

The similar analysis was carried out for trimmed data set $Z-$ with values $x_m = 2 \times 10^{-6}$, $y_1 = 4.85 \times 10^{-6}$ and $y_2 = 7.38 \times 10^{-6}$. The results are shown in Fig. 8a, 8b and corresponding MLE are $\hat{\gamma}_1 = -2.13$, $\hat{\gamma}_2 = -1.77$.

Once we have derived the approximation for the density of $\hat{\gamma}$, it enables us to do more statistical inference on this parameter. One of the biggest advantages of saddlepoint approximation based on I-divergence is that it provides analytical formula for approximation of the density of MLE. Therefore there is no need for any other numerical procedures and the density may be used straighforwardly. In this case not only the density is tractable in explicit form, but also distribution function possesses



(a) Density of MLE (b) I-divergence

Figure 8: Pareto real data Z- with saddlepoint densities for MLE estimate and I-divergence.

this desirable property. Among other applications, we will concentrate on estimating the probability that the values of parameter exceed certain values and finding the confidence interval for given probability.

Let us get back to the case of $Z+$ and randomly chosen observations $y_1 = 1.45 \times 10^{-5}$ and $y_2 = 4.24 \times 10^{-6}$. As we are dealing with an approximation, it may happen that it will not integrate exactly to 1. In our case

$$\int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} q_T(\hat{\gamma}|\gamma)\, \mathrm{d}\hat{\gamma}_1\, \mathrm{d}\hat{\gamma}_2 = 1.18.$$

The way to solve this small inconvenience is to divide the formula for density by this number and thus ensure this important property of density. The cumulative distribution function $Q(x, y)$ is a result of integration

$$Q(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} q_T(\hat{\gamma}|\gamma)\, \mathrm{d}\hat{\gamma}_1\, \mathrm{d}\hat{\gamma}_2.$$

Suppose we would like to compute the probability that both γ_1 and γ_2 will exceed the value -3 (or equivalently that both α_1 and α_2 will be less than 2). Therefore we need to compute

$$\int_{-3}^{-1} \int_{-3}^{-1} q_T(\hat{\gamma}|\gamma)\, \mathrm{d}\hat{\gamma}_1\, \mathrm{d}\hat{\gamma}_2$$

and the resulting probability equals 0.39.

In order to compute a confidence interval for γ_1 we have to obtain the marginal density of γ_1

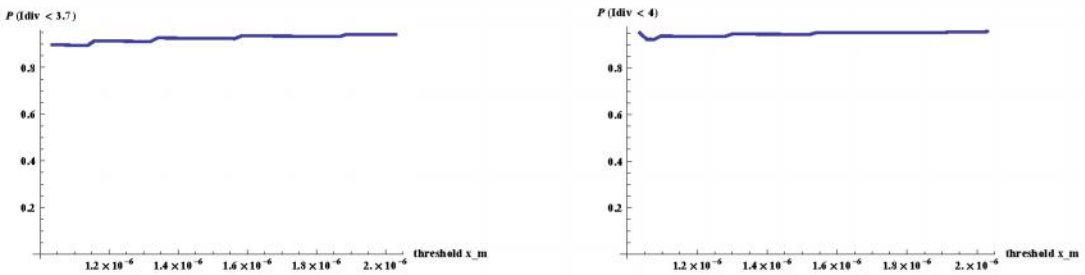
$$f_1(\hat{\gamma}_1) = \int_{-\infty}^{-1} q_T(\hat{\gamma}_1, \hat{\gamma}_2|\gamma)\, \mathrm{d}\hat{\gamma}_2$$

and find a, b such that

$$\int_a^b f_1(\hat{\gamma}_1)\,\mathrm{d}\hat{\gamma}_1 = 0.95.$$

As f_1 is continuous and not symmetric, more tuples a, b satisfying this equation can be found. One of them is $a = -4.25$ and $b = -1.15$, so the interval (a, b) will cover the real value of parameter γ_1 with probability 0.95.

Computation of confidence level of I-divergence contours in the case of known threshold x_m is in detail elaborated in [17]. The case of unknown x_m needs special attention. In the plots Fig. 9a and Fig. 9b we plotted the probability that I-divergence does not exceed value c for different values of threshold x_m and for values $\hat{\gamma}_1$ and $\hat{\gamma}_2$. The chosen values of c approximately yield the most widely used probabilities for confidence intervals in statistics - 0.9 and 0.95 respectively. We may observe that this probability in our case changes only imperceptibly for various choices of threshold x_m, that means that the value of I-divergence depends on the shape parameter and not on the threshold. The probabilities were approximated in Mathematica by numerical integration using global adaptive integration strategy.



(a) $c = 3.7$ (b) $c = 4$

Figure 9: Probability of event $I < c$ for various choices of threshold x_m

5. Discussion

The complexities of methane emissions from wetlands, especially many dependencies and interactions may lead an experimenter to conclusion that it is a complicated and chaotic system. This chaos can be particularly related to the potential ebullition of methane, which is by definition (see e.g. [23]) a spontaneous releasing of methane bubbles from flooded sediments (for a possible parametric modelling of ebullition see e.g. [8]). In this paper we try to introduce simple graphical tool for assessing the entropy in the system. One of valuable further research questions can be a practical validation of the relationship between chaos and ebullition by introduced entropy dissimilarity measure.

First, our method is parametric, and we can rigorously justify the parametric/semiparametric form of a Pareto tails (see [1, 8]). There exist nonparametric methods for nonlinear filtering, e.g. [24]. Thus, nonparametric framework is allowing more freedom in the model, however, it decreases the precision of measurement of the amount of chaos present in the system. But the latter, e.g. measurement of the amount of chaos present in the methane wetland system, is the main address of our paper. The goal of [24] in particular, and thus the goal of the so-called "Probability

Integral Transform" in general is the evaluation of the filtering probability density *as a whole*, not dependently on any particular point-estimator. However, the goal of the current measurement is to find a *good point estimate* (or credible interval) of *upper quantiles* of a complex stochastic process (not a one density), which mixes both chaos, stochasticity and deterministic part. From the point of view of statistical theory of information, the problems aimed by "Probability Integral Transform" and the problem of Entropy measured by upper quantiles of process are different-in the 1st case ("Probability Integral Transform") we need to use part of useful statistical information for the filtering probability density *as a whole*, in our case we take all useful statistical information entirely for quantiles, i.e. estimation of Pareto tails.

Chaos in biochemical systems can be measured in a different ways (e.g. by different complexity measures proposed in the literature like Shannon entropy, Relative entropy, Lempel-Ziv, Kolmogrov and Algorithmic complexity, etc.) and can be interpreted in a different set-ups. For example, one can be interested in uncertainty quantification in chemical system (see e.g. [25]). However, in the current manuscript we measure Entropy driven by heavy-tailed character of ebullition (or anomalous diffusion), which is ecologically interpretable, e.g. as releasing of methane by bubbles. Thus we are interested not in arbitrary technical entropy, which can be associated to our wetland system (and there is definitely not the only one), but to the entropy related to ecological (biological) fact of ebullition.

The study of chaos enables an important insight for ecologists since most natural systems are too complex to be easily comprehensible. The comprehension and prediction of behavior of ecological systems is difficult due to its irreducibility and persistent changes with none stable equilibrium (see [26]). Chaos was firstly recognized in ecology in simple population models ([27]). These models described changes of the population with deterministic way and chaos represents inaccurate model predictions ([28, 29]). Chaos can refer to the apparent unpredictability of deterministic systems, where the unpredictability is driven by gradual growth of errors in the specification of the initial state. The earliest methods used to determine and study chaos in empirical time series were developed for physical applications and assume noise-free dynamics and perfectly accurate data. Ecologists unfortunately cannot make these assumptions for living systems.

Acknowledgement
Monitoring of the methane emissions are supported in projects: Role of plants in the greenhouse gas budget of a sedge fen (P504/11/1151) by the Grant Agency of the Czech Republic, CzechGlobe project CZ.1.05/1.1.00/02.0073. First and second author have been partially supported by Aktion grant Austria-Czech Republic and by the grant SVV 2014-260105. The first author has been also partially supported by Czech Republic Grant 201/12/0083, the second author has been also partially supported by Czech Republic Grant 13-13502S. The authors are grateful to Editor and both referees.

6. Appendix

The entropy of a continuous random variable X having pdf $f(x)$ with respect to Lebesgue measure is defined as

$$H(f(x)) = -\int_X f(x) \ln f(x) dx. \tag{2}$$

The Kullback-Leibler (KL) divergence between two possible pdfs $f_1(x)$ and $f_2(x)$ of random variable X and measures the divergence of $f_2(x)$ from $f_1(x)$. KL-divergence is non-negative, anti-symmetric and reaches its minimum when $f_1(x) = f_2(x)$ a.e.:

$$\mathrm{KLD}(f_1(x)|f_2(x)) = \int_X f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx = -H(f_1(x)) - \int_X f_1(x) \ln f_2(x) dx. \tag{3}$$

The pdf of the Pareto distribution $Pareto(x_m, \alpha)$ with scale parameter x_m and shape parameter α:

$$f_X(x) = \alpha \frac{x_m^\alpha}{x^{\alpha+1}} \quad x_m > 0, \quad \alpha > 0 \tag{4}$$

Next we provide relations to other approaches via I-divergence. For recent developments on deconvolutions and decompositions of I-divergence see [30].

6.1. Relation to the divergence score

We focus on divergence score DS defined in [31] based on Brier score. This divergence score measures the mean KL-divergence (KLD) from observation distribution to forecast distribution. The observation distribution is represented by Kronecker delta, probability forecast (can be interpreted as the best estimate of the unknown true outcome) by alternative distribution. With each new observation a probability of rainfall is forecasted, the occurrence of the same forecast probabilities is possible. This score is then decomposed onto three parts: reliability, resolution, uncertainty. Reliability part stands for the expected divergence of the observed probability distribution from the forecast probability distribution. Resolution part contains the amount of information in observations with respect to the stratification based on values of forecasts. Uncertainty part contains uncertainty relative to the observation.

In order to generalize this decomposition, let us now consider N events represented by N independent Pareto distributed random variables. The observation distribution will now be Pareto with known scale parameter x_m and the unknown cannonical parameter γ replaced by a subjective MLE estimate $\hat{\gamma}_i$, which is a function of a observation y_i, $i = 1, \ldots, N$. The forecast distribution will be Pareto with the same scale parameter x_m and unknown parameter set to γ_0, $i = 1, \ldots, N$. Following the setup in [31], we obtain that:

$$\begin{aligned}\mathrm{DS} =& \mathrm{KLD}(Pareto(x_m, \hat{\gamma}_{MLE})||Pareto(x_m, \gamma_0)) - \mathrm{KLD}(Pareto(x_m, \hat{\gamma}_{MLE})||Pareto(x_m, \hat{\gamma}_{MLE})) \\ &+ \frac{1}{N}\sum_{i=1}^{N} \mathrm{KLD}(Pareto(x_m, \hat{\gamma}_i)||Pareto(x_m, \sum_i \hat{\gamma}_i/N)) = \frac{1}{N} I_N(\hat{\gamma}_y, \gamma_{0,N}),\end{aligned}$$

where $\hat{\gamma}_{MLE}$ is the maximum likelihood estimate of γ based on $y_1, \ldots, y_N$. First part of this decomposition, the expected divergence from the observation distribution to the hypothesis distribution, coincides with reliability part explained above. The second part, expressing the resolution part, disappears based on the properties of KL-divergence. This result was expected, since in this case we have just one subgroup (the forecast is same for each i). Third term, uncertainty, can be viewed as the uncertainty related to the observation, but also as the expected divergence from the observation distribution to the distribution with unknown parameter set as MLE estimate based on all observations.

6.2. Relation to the density power divergence

Let us have a random sample $x_1, \ldots, x_s$, independent identically distributed random variables, from the distribution with pdf depending on the unknown parameter γ. This coincides exactly with the I-divergence when $N = 1$ and $y_1 = (x_1, \ldots, x_s)$. Thus γ is one-dimensional. The joint pdf of y_1 is then

$$h(y_1|\gamma) = \exp\{-\psi(y_1) + \gamma T(y_1) - \kappa(\gamma)\}$$

The I-divergence $I_1(y_1, \gamma_0)$ for testing $H_0 : \gamma = \gamma_0$ against $H_1 : \gamma \neq \gamma_0$ has the following form:

$$I_1(\hat{\gamma}_1, \gamma_0) = I_1(\hat{\gamma}_{MLE}, \gamma_0) + I_1(\hat{\gamma}_1, \hat{\gamma}_{MLE}) = I_1(\hat{\gamma}_{MLE}, \gamma_0) + 0$$

where the maximum likelihood estimate $\hat{\gamma}_{MLE}$ coincides with subjective maximum likelihood estimate $\hat{\gamma}_i$. Using the minimum density power estimator of γ instead of maximum likelihood estimator we obtain the density power divergence $R_1(h_{\hat{\gamma}}, h_{\gamma_0})$ (see [32]).

The test statistics $2I_1(\hat{\gamma}_x, \gamma_0) = 2R_1(h_{\hat{\gamma}}, h_{\gamma_0})$ composed from the considered divergences are asymptotically distributed according to χ^2_1 for $s \to \infty$ (see [16]) as well as test statistic $2sd(\hat{\gamma}_{MLE}, \gamma_0)$ derived in [33].

6.3. Relation to decision procedures based on scaled Bregman distance surfaces

Let us consider a probability density function of a distribution belonging to the exponential family:

$$p_\gamma(y) = \exp\left\{g_0(y) + \sum_{i=1}^{N} g_i(y)T_i(\gamma) - b(T(\gamma))\right\} \tag{5}$$

The canonical version of the considered pdf is:

$$p_\gamma(y) = \exp\left\{g_0(y) + \sum_{i=1}^{N} g_i(y)\gamma_i - b(\gamma)\right\}$$

The coinciding Bregman power distance $B_\alpha(P_{\gamma_1}, P_{\gamma_2}|P_{\gamma_0})$ can be found in [34], formula (5). If we consider $\alpha = 1$, then the Bregman distance coincides with KL-divergence. The formula for scaled Bregman power distance then reads:

$$\begin{aligned} B_1(P_{\gamma_1}, P_{\gamma_2}|P_{\gamma_0}) =& b(T(\gamma_2)) - b(T(\gamma_1)) - \Delta b(T(\gamma_1))(T(\gamma_2) - T(\gamma_1)) \\ =& b(\gamma_2) - b(\gamma_1) - \Delta b(\gamma_1)(\gamma_2 - \gamma_1) \end{aligned} \tag{6}$$

The above given probability density function (5) can be rewritten as follows:

$$\begin{aligned} p_\gamma(y) =& \prod_{i=1}^{N} f(y_i)h(\gamma_i)\exp\{t(y_i)\gamma_i\} \\ =& \exp\left\{\sum_{i=1}^{N} \ln f(y_i) + \sum_{i=1}^{N} t(y_i)\gamma_i + \sum_{i=1}^{N} \ln h(\gamma_i)\right\} \\ =& \exp\{-\psi(y) + t(y)\gamma - \kappa(\gamma)\} \end{aligned}$$

Thus the relatione between the scaled Bregman power distance (6) and I-divergence is:

$$\begin{aligned} B_1(P_{\gamma_1}, P_{\gamma_2}|P_{\gamma_0}) =& \kappa(\gamma_2) - \kappa(\gamma_1) - \frac{\delta\kappa(\gamma_1)^T}{\delta\gamma_1}(\gamma_2 - \gamma_1) \\ =& \sum_{i=1}^{N} \ln h(\gamma_{1,i}) - \sum_{i=1}^{N} \ln h(\gamma_{2,i}) + \sum_{i=1}^{N} t(y_i)(\gamma_{1,i} - \gamma_{2,i}) \\ =& I_N(\gamma_1, \gamma_2) \end{aligned}$$

References

[1] P. Jordanova, J. Dušek, M. Stehlík, Modeling methane emission by the infinite moving average process, Chemometrics and Intelligent Laboratory Systems 122(2013) 40–49.

[2] E. Schrödinger, What is Life?, Cambridge University Press, Cambridge, 1967.

[3] T.M. Addiscott, Entropy, non-linearity and hierarchy in ecosystems, Geoderma 160(2010) 57–63.

[4] W.H. Schlesinger, Biogeochemistry: An Analysis of Global Change, Academic Press, Amsterdam, 1997, p. 558.

[5] J. Le Mer, P. Roger, Production, oxidation, emission and consumption of methane by soils: A review, Eur. J. Soil Biol. 37(1)(2001) 25–50.

[6] P.N. Shukla, K.D. Pandey, V.K. and Mishra, Environmental Determinants of Soil Methane Oxidation and Methanotrophs, Critical Reviews in Environmental Science and Technology 43(2013) 1945–2011.

[7] M. Stehlík, L. Střelec, M. Thulin, On robust testing for normality in chemometrics, Chemometrics and Intelligent Laboratory Systems 130(2003) 98–108.

[8] P. Jordanova, J. Dušek, M. Stehlík, Microergodicity effects on ebullition of methane modelled by Mixed 2 Poisson process with Pareto mixing variable, Chemometrics and Intelligent Laboratory Systems 128(2013) 124–134.

[9] J. Richardson, M.J. Vepraskas, Wetland Soils: Genesis, Hydrology, Landscapes and Classification, Lewis Publishers, USA, 2000.

[10] H.J. Laanbroek, Methane emission from natural wetlands: interplay between emergent macrophytes and soil microbial processes, A mini-review, Annals of Botany 105(2010) 141–153.

[11] G.H. Yari, G.R. Mohtashami Borzadaran, Entropy for Pareto-types and its order statistics distributions, Commun. Inf. Syst. 10(3)(2010) 193–202.

[12] M.P. Holland, R. Vitolo, P. Rabassa, A.E. Sterk, H. Broer, Extreme value laws in dynamical systems under physical observables, (1982) arXiv:1107.5673v1 28 July 2011.

[13] S. Kullback, R.A. Leibler, On Information and Sufficiency, Ann. Math. Statist. 22(1)(1951) 79–86.

[14] S. Kullback, Information Theory and Statistics, Dover Publications, 1997.

[15] A. Verster, D.J. De Waal, S. Van Der Merwe, Selecting an optimum threshold with the Kullback-Leibler deviance measure. Available online.

[16] M. Stehlík, Distributions of exact tests in the exponential family, Metrika 57(2003) 145–164.

[17] M. Stehlík, P. Economou, J. Kiselak, W.D. Richter, Kullback - Leibler life time testing, Applied Mathematics and Computation, DOI: 10.1016/j.amc.2014.04.027.

[18] M.S. Bartlett, D.G. Kendall, The statistical analysis of variance-heterogeneity and the logarithmic transformation, Supplement to the Journal of the Royal Statistical Society 8(1)(1946) 128–138.

[19] J.I. Marden, Positions and QQ Plots, Statistical Science 19(2004) 606–614.

[20] J.E. Shore, R.W. Johnson, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, IEEE Trans. Inf. Theory 26(1980) 26–37.

[21] T. Jaynes, On The Rationale of Maximum Entropy Methods, Proc. IEEE 70(1982) 939–952.

[22] A. Pázman, Nonlinear statistical Models, Kluwer Acad. Publ., Dordrecht, 1993.

[23] J.P. Chanton, G.J. Whiting, Trace gas exchange in freshwater and coastal marine environments: ebullition and transport by plants, in P.A. Matson, R.C. Harriss (Eds.), Biogenictrace gases: measuring emissions from soil and water, Blackwell, Oxford, 1995.

[24] L. Chen, C. Lee, R. K. Mehra, How to tell a bad filter through Monte Carlo simulations, IEEE Transactions on Automatic Control 52 (2007) 1302-1307

[25] H.N. Najm, B. J. Debusschere, Y. M. Marzouk, S. Widmer and O. P. Le Maître. Uncertainty quantification in chemical systems, International Journal For Numerical Methods In Engineering, 85(5), (2011) 670-670.

[26] B. Beckage, L.J. Gross, S. Kauffman (2011). The limits to prediction in ecological systems. Ecosphere 2:art125.

[27] R.M. May (1976). Simple mathematical models with very complicated dynamics. Nature 261:459-67.

[28] R.M. May (1974). Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. Science 186:6457.

[29] A. Hastings, C. Hom, S. Ellner, P. Turchin, H. Godfray (1993). Chaos in ecology - is mother-nature a strange attractor. Annual Review Of Ecology And Systematics 24:133.

[30] M. Stehlík, Decompositions of information divergences: recent development, open problems and applications, AIP Conf. Proc. 1493(2012) 972–976.

[31] S.V. Weijs, N. Van De Giesen, Accounting for Observational Uncertainty in Forecast Verification: An Information-Theoretical View on Forecasts, Observations, and Truth, Mon. Weather Rev. 139(2011) 2156–2162.

[32] A. Basu, A. Mandal, N. Martin, L. Pardo, Testing statistical hypotheses based on the density power divergence, Ann Inst Stat Math 65(2013) 319–348.

[33] S. Bedbur, E. Beutner, U. Kamps, Generalized order statistics: an exponential family in model parameters, Statistics: A Journal of Theoretical and Applied Statistics 46(2)(2012) 159–166.

[34] A.L. Kisslinger, W. Stummer, Some Decision Procedures Based on Scaled Bregman Distance Surfaces, in F. Nielsen, F. Barbaresco (Eds.): GSI 2013, LNCS 8085, 2013, 479-486. Springer-Verlag

List of Figures