Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series 2012-62

Modeling methane emission by the infinite moving average process

Pavlina Jordanova[a], Jiří Dušek[b], Milan Stehlík

July 2012

[a]Shumen University
[b]Global Change Research Centre AS CR

Modeling methane emission by the infinite moving average process

Jordanova, P.[a], Dušek, J.[b], Stehlík, M.[c,*]

[a]*Faculty of Mathematics and Informatics, Shumen University, Bulgaria*
[b]*Global Change Research Centre AS CR, České Budejovice, Czech Republic*
[c]*Department of Applied Statistics,Johannes Kepler University in Linz, Austria*

Abstract

The aim of this paper is to provide guidelines for the statistically efficient estimation of the methane emission in the sedge-grass marsh, South Bohemia, Czech Republic. Such a study is important for better understanding of the rate and dynamics of methane emissions from wetland ecosystem and in particular to build up a model of local methane flux. We model a time series of the data by an infinite moving average process with Pareto tails of the exceedances. To insure the consistency of such an estimator, we use both Hill and t-Hill estimators. We provide the proof of the weak consistency of t-Hill estimator for the case of dependent data. This estimator has a simple form and it provides a nice trade-off between efficiency and robustness. Finally we present a discussion on its robustness.

Keywords: Environmental Chemistry, Pareto tails, t-Hill estimator, Hill estimator, weak consistency, moving average process, Methane emission model, robust statistics, wetland ecosystem, sedge-grass marsh, fen

1. Introduction

Several studies revealed central importance of understanding of methane emission recently (see e.g. [17] for the case of methane emissions from natural wetlands and references therein). Methane is recognized as one of the most important greenhouse gases (see e.g. [27]). It is in natural ecosystems product of anaerobic decomposition processes of organic matter. These processes require the synergy of anaerobic bacteria and methanogenic archaea. Organic matter is firstly hydrolyzed and fermented, and formed products are compounds that are used for methanogenesis (see [16]). Bacteria which are capable produced methane gas are strictly anaerobic (see [12]). Anaerobic conditions suitable for methane production are in water saturated soils during long time period. This conditions and soils we can find in wetland ecosystems where is enough of organic matter from primary production and water level is in or above soil surface. Measured methane emissions from the wetlands are the results of the biogeochemical process: (a) methane production and (b) methane consumption by the methanotrophic bacteria. Resulted methane emissions are important in relation to the global climatic changes.

*Corresponding author. Institut für angewandte Statistik, Altenberger Straße 69, A-4040 Linz, Austria. Email: Milan.Stehlik@jku.at Tel.:+4373224686808; Fax: +4373224686800

Email addresses: `pavlina_kj@abv.bg` (Jordanova, P.), `dusek.j@czechglobe.cz` (Dušek, J.), `Milan.Stehlik@jku.at` (Stehlík, M.)

Besides an importance of the optimal design (see recent work of [24]) it is very important to learn the nature of extremes in methane emission. This is the main motif of this paper. Emission of methane in several parts of atmosphere are of interest (see e.g. for the case of troposphere [28] or [13]). Since troposphere, extensively studied because of greenhouse gas emission, provides a very exotic environment for kinetics of chemical reactions normally studied at the Earth, it can be argued that correlation parameter and modified Arrhenius equation can play a crucial role by its modeling. It was applied to the study of the methane removal in the atmosphere [24].

However, models of methane emission in a given ecosystems are much more complicated, involving several explanatory variables and they are not completely understood. In particular, they are not developed for all countries and this vacancy harms a proper harmonizations of obtained results. The model we use here is infinite Moving Average (MA) with heavy tailed distribution function of exceedances of the error terms which can be expressed as $X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}$, supposing that at least one of the real numbers c_j is positive and there exist a majorization constant $\delta \in (0,1), \delta < \alpha$ such that $\sum_{j=0}^{\infty} |c_j|^{\delta} < \infty$. If the model underlying the data, is a regularly varying function with index $-1/\alpha$ it is usually supposed that the top scaled order statistics are Pareto distributed. [15] derived a procedure of Pareto tail estimation by the MLE. Later on, many authors tried to robustify the Hill estimator, but they still rely on maximum likelihood. However, the influence function of Hill estimator is slowly increasing, but unbounded. Hill procedure is thus no robust and many authors tried to make the original Hill robust (see [2] and references therein). In [8] a new method of score moment estimators has been proposed. It appeared that these score moment estimators are robust for heavy tailed distributions (see [25]). For the case of Pareto distribution, the t-Hill estimator procedure with the score moment estimator has been investigated in [26] for optimal testing for normality against Pareto tail.

The paper is organized as follows. The case study on Methane emission from the sedge-grass marsh motivates the statistical techniques developed in Section 2. In the real data example we observed that the exceedances of the trend component are not normal (as we have checked e.g. by *pp*-plot) however they can be well modeled by Pareto tails. We fit two practical modes: In subsection 2.1 only time is regressor, while in subsection 2.2 regressors are water level and time and in subsection 2.3 also two soil temperatures are added to regressors. Simulation from the established models demonstrates the goodness of fit of the initial data. We also show that for such a type of data both Hill and t-Hill estimators are behaving well. Remarkable is also the simplicity of deterministic part of the model, in comparison with the estimable trend of the corresponding response surface model. The result of weak consistency for t-Hill estimator is proven for the case of dependent data.

In Section 3 the simulation study of robustness and rate of convergence for the Hill and the t-Hill estimators is given. We have observed that the t-Hill is more robust with respect to the right outliers and more sensitive to the center of the distribution. On the other hand the Hill estimator is more efficient under the assumption of Pareto tails. The trade off between efficiency and robustness depends mainly on the tail index α. Finally Section 4 concludes and in particular discusses the obtained results from both methodological and practical point of view. To maintain the continuity of the explanation proofs and technicalities are deferred to the Appendix.

2. Case study: Sedge-grass marsh as the source of methane

Methane (CH_4) is an important greenhouse gas, whose concentration in the troposphere is steadily increasing. To estimate the flux of methane into the atmosphere and its atmospheric lifetime, its rate of removal needs to be accurately determined. The main loss process for atmospheric methane is the reaction with the hydroxyl radical, $OH + CH_4 \rightarrow CH_3 + H_2O$. This reaction has been extensively studied, and it can be expressed in the modified Arrhenius form as

$$k(t) = a \cdot \frac{1}{t^m} exp(-\frac{\beta}{t})$$

i.e. we have Generalized Exponential model with correlated errors, studied in [24]. Therein the filling designs (see [22]) for correlated version of modified Arrhenius model has been employed (for independent case see [23]). However, models of methane emission in a given ecosystem on the Earth are much more complicated, involving several explanatory variables and they are not completely understood. And for a model-based design one needs to have first the model so that the optimal designs could be established.

In this study were used data of the methane emission from the natural wetland ecosystem in the South Bohemia region, Czech Republic. The sedge-grass marsh is part of a wetland complex Mokré Louky (the Wet Meadows), situated near the town of Trebon, South Bohemia, Czech Republic. The marsh is a flat depression with an area of 450 ha, situated in the inundation area of a large human-made lake (Rožmberk fishpond, 5 km^2). Water level is fairly stable and fluctuating usually from 0.2 to +0.1 m relative to soil surface. The minimum water levels is about 0.3 m. During floods the water level culminates as high as 2 m above the soil surface (see [6]). We may expect by ecological reasons that the methane emission depends substantially on water level and soil temperature. Emissions of methane are not having a daily periodicity (in comparison of emissions of CO_2 (respiration of soil and vegetation), but there is a significant seasonality in the data.

Methods of the emission measurements

For measurement emissions of CH_4 non-transparent closed chambers were used (non-steady-state through-flow system). This system was developed at the Department of Matter and Energy Fluxes (Global Change Research Center AS Czech Republic, v.v.i.). The chamber system is assembled from three larger chambers, computer unit responsible for chamber closing and data processing and greenhouse gas concentration analyzer of CO_2 and CH_4 (DLT-100 model: 908-0007, Los Gatos Research Inc., USA The chamber body has cylindrical shape with dimensions 1 m in diameter and 1 m in height. Internal volume of chamber is 785 L. Chamber closes/opens vertically upwards to a height of 1.2 m above soil surface to the bottom chamber corner. For sealing between chamber body and soil surface, sealing collar was inserted in soil. Dimensions of this collar were 1 m in diameter and 0.3m in height and 0.15 m from the collar height was inserted to soil. On the top of collar was fixed the circumferential slot. In this slot the bottom corner of chamber body was seals by water and neoprene sealing. Measurements were carried out automatically at 2-hours intervals during whole day. Each single measurement spent about 20 minutes (closing time of the chamber). During this time actually concentration of CH_4 continuously recorded in second intervals and after following format transformation stored in database. After the measurement minute averages are calculated from the stored second data. Minute data (20 average values) are used for gasses flux calculation. Additional measurements to chamber measurements were soil temperature profile at 5-levels 0, 5, 10, 20 and 30 cm (Pt-100, EMS Brno, Czech Republic), air temperature inside chamber and air pressure (sensor-TMAG 518 N4H, Cressto, Czech Republic).

In this section we discuss modeling of the methane data by infinite moving average time series with Pareto like positive(negative) part of the innovations. Such a time series in a various statistical considerations has been studied in previous works (c.f. [20] for the consistency of the estimation procedure under the dependence or [4] for general issues). The novelty of this paper is consideration of t-Hill estimator for the tail index, which is consistent also for some cases of dependent data, as stated in the following Theorem 1.

Further on we suppose that $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ are r.v's with distribution function (d.f.) F and upper order statistics

$$\mathbf{X}_{(n,n)} \leq \mathbf{X}_{(n-1,n)} \leq ... \leq \mathbf{X}_{(1,n)}.$$

We consider the t-Hill estimator of α^{-1}

$$\frac{1}{\widehat{\alpha}_{k,n}} = \left\{ \frac{1}{k} \sum_{i=1}^{k} \frac{\mathbf{X}_{(k+1,n)}}{\mathbf{X}_{(i,n)}} \right\}^{-1} - 1, \quad k = 1, 2, ..., n. \tag{1}$$

Let

$$\mu_{X,k(n),n}(\cdot) := \frac{1}{k(n)} \sum_{i=1}^{n} \varepsilon \left\{ \frac{\mathbf{X}_i}{b(\frac{n}{k(n)})} \in (\cdot) \right\}$$

be a random element in the space $\mathbb{E}^+$ of non-negative Radon measures on $(0, \infty]$ endowed with the vague topology. Here

$$b(t) := F^{\leftarrow}(1 - \frac{1}{t}) = \left(\frac{1}{\overline{F}} \right)^{\leftarrow} (t) \tag{2}$$

is the quantile function of F.

The following three conditions are substantial for the consistency.

First, $1 - F$ is regularly varying with exponent $-\alpha, \alpha > 0$, briefly

$$1 - F \in RV_{-\alpha}. \tag{3}$$

The second is the following Mason's condition

$$\frac{k(n)}{n} \to 0, \quad k(n) \to \infty, \quad n \to \infty \tag{4}$$

(i.e. $k(n)$ is an intermediate sequence).

The third concerns the following vague convergence (see [21])

$$\mu_{X,k(n),n} \Longrightarrow \mu, \quad n \to \infty, \tag{5}$$

where $\mu : \sigma((0, \infty]) \to [0, \infty)$ and $\mu(x; \infty] = x^{-\alpha}$, $x > 0$.

Theorem 1. *Let conditions (3,4,5) hold. Then* $\frac{1}{\widehat{\alpha}_{k,n}}$ *is an weakly consistent estimator for* $\frac{1}{\alpha}$.

Further we apply the obtained results for a particular case of infinite moving average sequence.

Corollary. Suppose that at least one of the real numbers $c_j, j = 0, 1, ...$ is positive and there exists $\delta \in (0, 1)$, $\delta < \alpha$ such that

$$\sum_{j=0}^{\infty} |c_j|^{\delta} < \infty. \tag{6}$$

Consider the moving average sequence

$$\mathbf{X}_n = \sum_{j=0}^{\infty} c_j \mathbf{Z}_{n-j}, \quad -\infty < n < \infty, \tag{7}$$

where $\mathbf{Z}_i, -\infty < i < \infty$, are non-negative independent identically distributed(i.i.d.) innovations with d.f. G, such that $\overline{G} \in RV_{-\alpha}$, $\alpha > 0$. If the Mason's condition (4) holds then $\widehat{\alpha}_{k,n}^{-1}$ is weakly consistent estimator for α^{-1}.

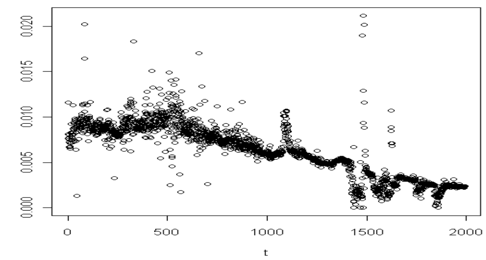

Figure 1: The initial data after 480-th observation (time in "in 2 hours").

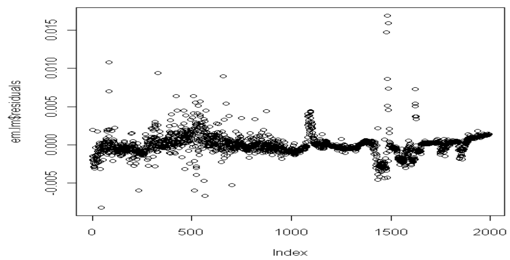

Figure 2: The plot of Em without polynomial(2) trend.

2.1. Infinite moving average model: only time is regressor

In this section we apply the infinite moving average model to the methane data. First we made the time intervals regular. Then we consider the model from the data after 480-th observation because of there were too many missing data in the first part. The initial data of methane emissions are plotted on Figure 1.

We denote the data from the sample at moment $t > 0$ by $Em(t)$, where Em stands for abbreviation of "emissions". In this section we consider quadratic model with time as entire regressor,

$$Em(t) = a + bt + ct^2 + X(t) + Z(t) - Z_-(t) + \epsilon(t).I_{[-c_1,c_2]}(\epsilon(t)), \quad t \geq 0, \tag{8}$$

where $a \in \mathbb{R}$, $b \in \mathbb{R}$, $c \in \mathbb{R}$, $X(t)$ is a moving average process, $\epsilon(t) \sim N(0,\sigma^2), Z(t) \sim Pareto(\alpha_1)$ and $Z_-(t) \sim Pareto(\alpha_2)$. The processes X, ϵ, Z and Z_- are homogeneous in time and independent. $I_{[-c_1,c_2]}$ is the identificator function of the set $[-c_1, c_2]$, c_2 is the left-end point of $Z(1)$ and c_1 is the left-end point of $Z_-(1)$. The process ϵ is a Wienner process. Z and Z_- have independent additive increments.

Notice that the process X inherit the dependent structure of the process Em without its deterministic trend. We are using the exceedances over given threshold. Here Z stand for the positive exceedances of the error term over c_2 and Z_- is for the exceedances of the minus error term above c_1. We suppose that between $-c_1$ and c_2 the errors are truncated $N(0, \sigma^2)$ distributed.

The deterministic part of Em was modeled by least-square line $\widehat{Em}(t) = a + bt + ct^2$. We obtained it by `lm` function in software R 2.15.00 ([R Development Core Team]).

$$\widehat{Em}(t) = 0,009606 - 0,000001389.t - 0,000000001488.t^2,$$

where the residual standard error is 0.001487. The plot of $Em - \widehat{Em}$ is given on Figure 2.

The partial autocorrelation function of $Em - \widehat{Em}$ is plotted on Figure 3. The estimated parameter of the autoregression is 12. Further we modeled X by causal AR(12). We use the Yule-Walker estimators (see e.g. [4]), implemented in software R 2.15.00 for the coefficients of the model.

```
Coefficients:
       ar1      ar2      ar3       ar4       ar5      ar6       ar7
     0.4537  0.0011  0.0910    0.0694    0.0797  -0.0246   -0.0498
      ar8     ar9      ar10    ar11     ar12      intercept
   -0.0017  0.0473 0.0333  0.0176  0.1136       0e+00
```

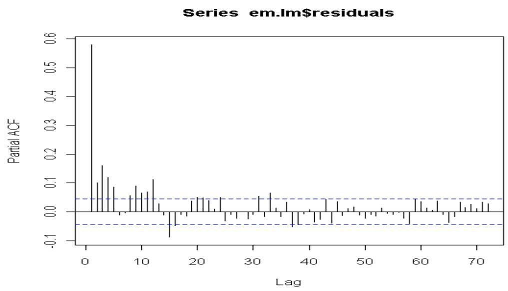

Figure 3: Partial ACF of $Em - \widehat{Em}$.

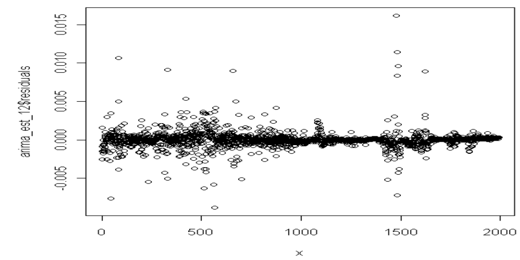

Figure 4: The plot of $Em - \widehat{Em} - \hat{X}$.

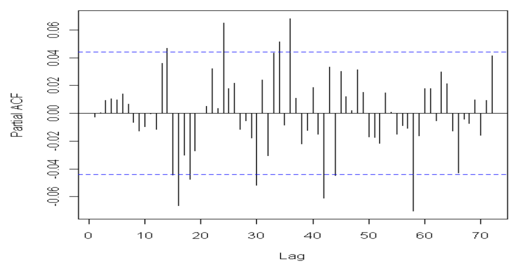

Figure 5: Partial ACF of the residuals $Em - \widehat{Em} - \hat{X}$.

```
sigma^2 estimated as 1.334e-06:  log likelihood = 10646.28,  aic = -21264.56
```

The errors of $Em - \widehat{Em} - \hat{X}$ are plotted on Figure 4. Their partial autocorrelation function, given on Figure 5, shows that these residuals are almost independent. Their histogram (Figure 6) and the mean-excess plots of the positive and negative parts of $Em - \widehat{Em} - \hat{X}$ (Figure 7 and Figure 8) show that the observed random variables have Pareto like distribution. Change of the slot of the ME-plot determines the appropriate number of upper order statistics (see [7], p. 355.)

We estimate the parameters of these Pareto distributions by Hill and t-Hill estimators.

The Hill plot consists of points with coordinates

$$(k, \frac{1}{k} \sum_{i=1}^{k} ln \frac{\mathbf{X}_{(i,n)}}{\mathbf{X}_{(k+1,n)}}), \quad k = 1, 2, ..., n.$$

For the positive and the negative parts of $Em - \widehat{Em} - \hat{X}$ they are given on Figure 9 and 11.

The t-Hill plot is defined as the set of points with coordinates

$$(k, \frac{1}{\widehat{\alpha}_{k,n}}), \quad k = 1, 2, ..., n,$$

For positive and negative parts of $Em - \widehat{Em} - \hat{X}$ they are given correspondingly on Figure 10 and Figure 12. Note that the Hill and tHill plots have similar behavior. In order to obtain weak

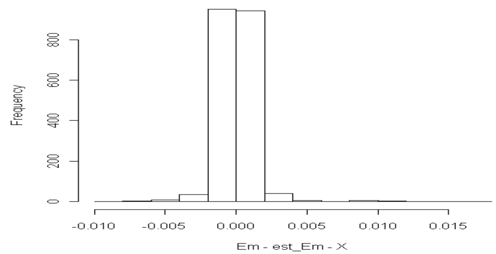

Figure 6: The histogram of $Em - \widehat{Em} - \hat{X}$.

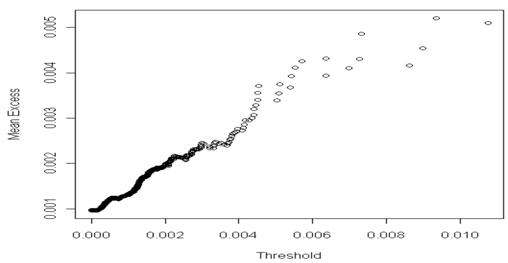

Figure 7: The mean excess plot of the positive parts of $Em - \widehat{Em} - \hat{X}$.

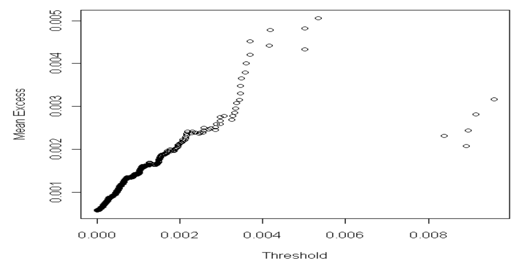

Figure 8: The mean excess plot of the negative parts of $Em - \widehat{Em} - \hat{X}$.

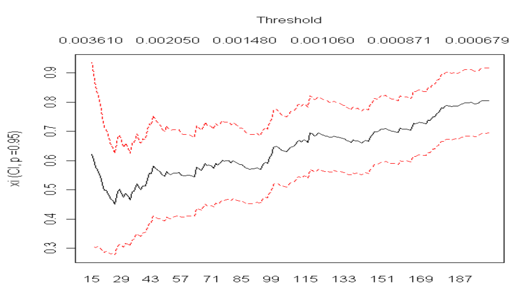

Figure 9: The Hill plot of the positive parts of $Em - \widehat{Em} - \hat{X}$.

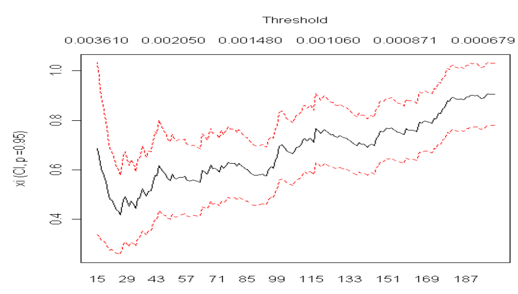

Figure 10: The t-Hill plot of the positive parts of $Em - \widehat{Em} - \hat{X}$

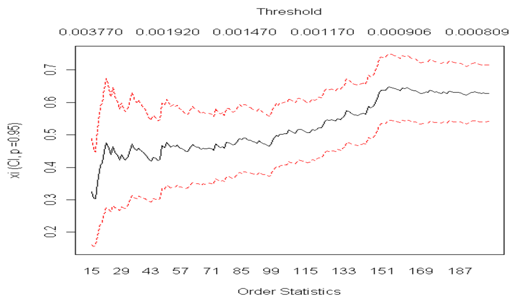

Figure 11: The Hill plot of the negative parts of $Em - \widehat{Em} - \hat{X}$

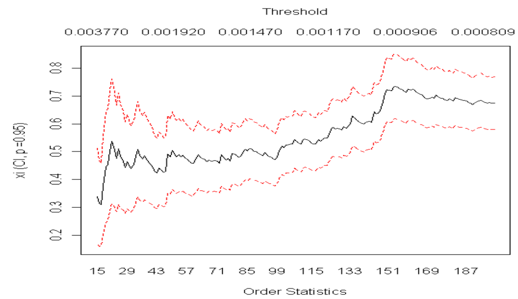

Figure 12: The t-Hill plot of the negative parts of $Em - \widehat{Em} - \hat{X}$

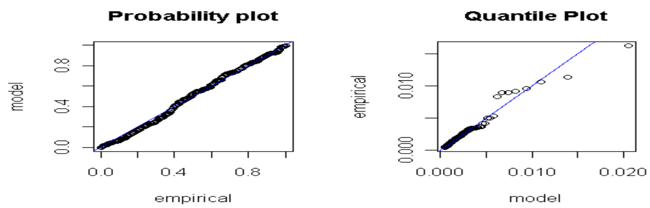

Figure 13: GPD fit of the exceedances of $Em - \widehat{Em} - \hat{X}$ over 0.0005 threshold.

consistent estimators we need conditions of the corollary to be meet. Therefore these estimators are good only for very large samples and k should satisfy the Mason's condition (4).

We made the generalized Pareto fit of the exceedancess of $Em - \widehat{Em} - \hat{X}$ over $c_2 = 0.0005$ threshold and obtained that the observed random variable is Pareto distributed, with $\alpha_1 = 2$ (see Figure 13). Analogously the generalized Pareto fit of the exceedances of $-(Em - \widehat{Em} - \hat{X})$ over $c_1 = 0.0005$ threshold shows that it is Pareto(3.33) distributed (see Figure 14).

According to [5]'s theoretical result the tail indexes of the innovation distribution functions are the same as the corresponding tail indexes of the distribution function of the exceedances of X.

2.1.1. Empirical check of the model

In this paragraph we check by simulation study that our model (8) can produce similar behavior as the initial methane emission data. In this simulation study we used symmetric thresholds $c_1 = c_2 = 0.0005$. However, $Z's$ are not symmetric, we use $\alpha_1 = 2$ and $\alpha_2 = 3.33$, as obtained in the above study. The Figure 15 shows the obtained simulated emission from this model after 480 data.

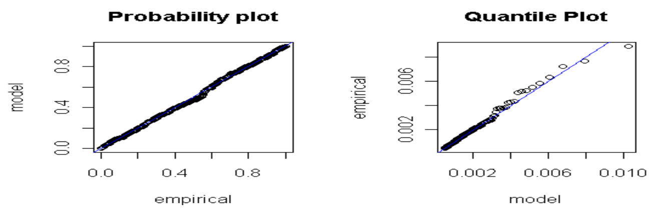

Figure 14: GPD fit of the exceedances of $-(Em - \widehat{Em} - \hat{X})$ over 0.0005 threshold.

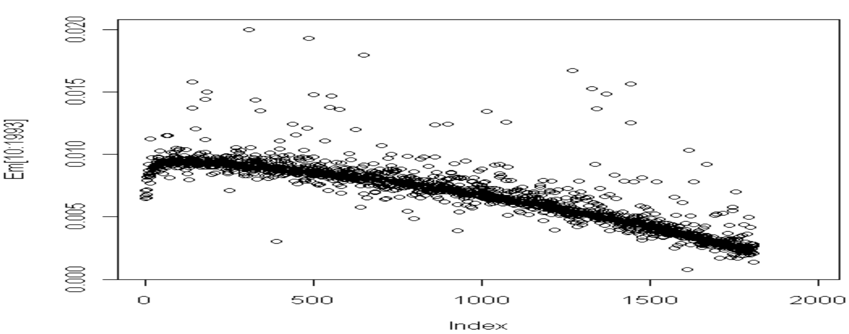

Figure 15: Empirical check of the model

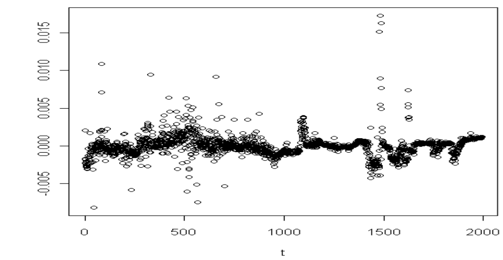

Figure 16: The plot of $Em - \widehat{Em}$.

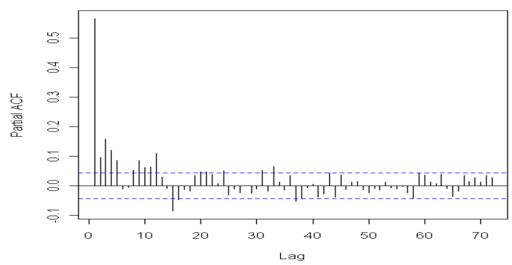

Figure 17: ACF of the process $Em - \widehat{Em}$.

2.2. Infinite moving average model, regressors: time, water level

In this section, analogously to the previous one, we estimate a regression model with deterministic part

$$\widehat{Em}(t) = a + b.t + c.t^2 + d.w + e.w^2 + f.t.w,$$

i.e.

$$Em(t) = \widehat{Em}(t) + X(t) + Z(t) - Z_-(t) + \epsilon(t).I_{[-c_1,c_2]}(\epsilon(t)), \tag{9}$$

where $t \geq 0$ stands for time, w is the water level and a, b, c, d, e, f are coefficients of the response surface model. The processes X, Z, Z_- and ϵ are as in the model (8).

First we have regularized time, than enlarged time intervals, omitted first 480 periods (because of the huge part of missing data there) and finally obtained 1716 observations. then we received the coefficients in $\widehat{Em}(t)$ by OLS procedure lm in R software.

```
Coefficients:
Intercept       t          t^2          w         w^2          t:w
9.528e-03  -1.558e-06  -1.366e-09  9.870e-06  1.822e-06  -4.168e-08
```

The residual standard error is 0.001464 and the plot of $Em - \widehat{Em}$ is given on Figure 16.

The partial autocorrelation function of X is plotted on Figure 17.

We use the Yule-Walker estimators (see e.g. [4]) for the coefficients of X and modeled it X causal AR(12) process.

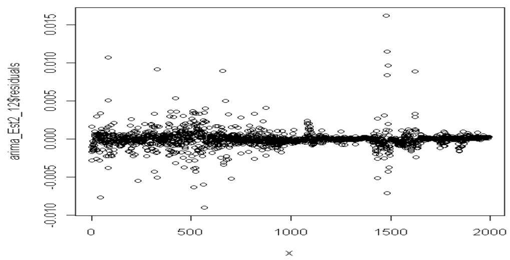

Figure 18: The plot of $Em - \widehat{Em} - \hat{X}$.

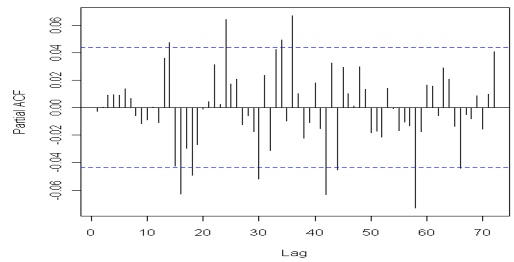

Figure 19: Partial ACF of the residuals $Em - \widehat{Em} - \hat{X}$.

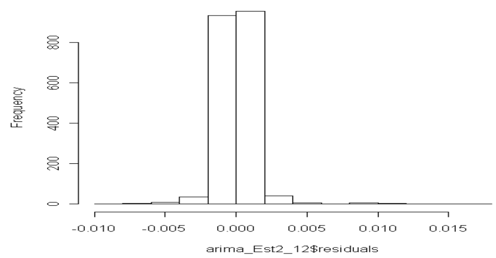

Figure 20: The histogram of the positive parts of $Em - \widehat{Em} - \hat{X}$.

```
Coefficients:
         ar1      ar2     ar3     ar4     ar5      ar6      ar7      ar8
      0.4489  -0.0009  0.0904  0.0699  0.0799  -0.0222  -0.0491  -0.0010
s.e.  0.0223   0.0245  0.0245  0.0245  0.0246   0.0246   0.0246   0.0246
         ar9     ar10    ar11    ar12  intercept
      0.0457   0.0319  0.0144  0.1112      0e+00
s.e.  0.0245   0.0244  0.0244  0.0223      1e-04

sigma^2 estimated as 1.328e-06:  log likelihood = 10650.56,  aic = -21273.12
```

The errors of $Em - \widehat{Em} - \hat{X}$ are plotted on Figure 18. Their partial autocorrelation function is given on Figure 19. It shows that these residuals are almost independent. Their histogram (Figure 20) and the mean-excess plots of the positive and negative parts of $Em - \widehat{Em} - \hat{X}$ (Figure 21 and Figure 22) show that the observed random variables have Pareto like distribution.

The Hill plots for the positive and the negative parts of $Em - \widehat{Em} - \hat{X}$ are given on Figure 23 and 25. Corresponding t-Hill plots are given on Figure 24 and Figure 26.

We made the generalized Pareto fit of the exceedancess of $Em - \widehat{Em} - \hat{X}$ over $c_2 = 0.0005$ threshold and obtained that the observed random variable is Pareto distributed, with $\alpha_1 = 1.91$ (see Figure 27). Analogously the generalized Pareto fit of the exceedances of $-(Em - \widehat{Em} - \hat{X})$ over $c_1 = 0.0005$ threshold shows that it is Pareto(3.03) distributed (see Figure 28).

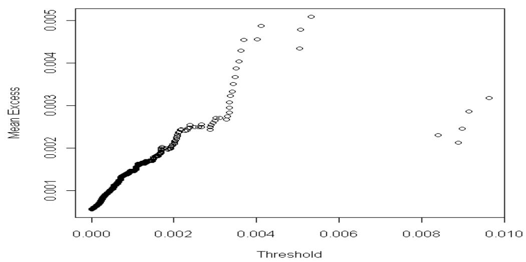

Figure 21: The mean excess plot of the positive parts of $Em - \widehat{Em} - \hat{X}$.

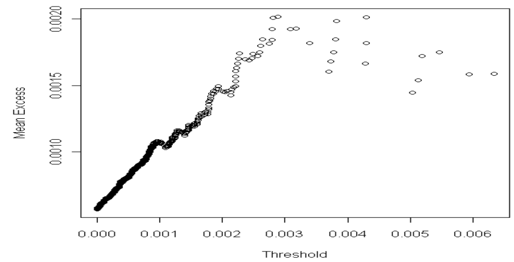

Figure 22: The mean excess plot of the negative parts of $Em - \widehat{Em} - \hat{X}$.

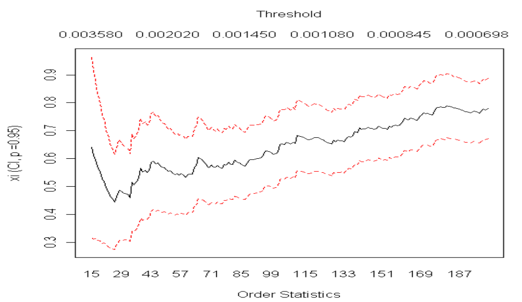

Figure 23: The Hill plot of the positive parts of $Em - \widehat{Em} - \hat{X}$.

Figure 24: The t-Hill plot of the positive parts of $Em - \widehat{Em} - \hat{X}$

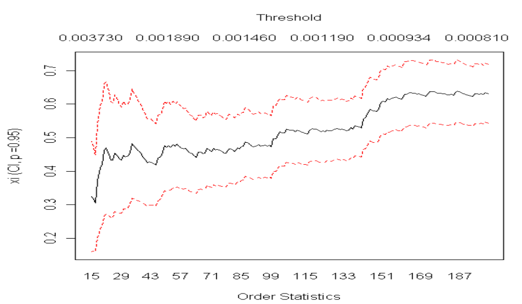

Figure 25: The Hill plot of the negative parts of $Em - \widehat{Em} - \hat{X}$

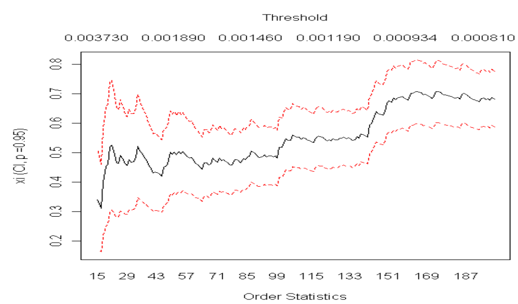

Figure 26: The t-Hill plot of the negative parts of $Em - \widehat{Em} - \hat{X}$

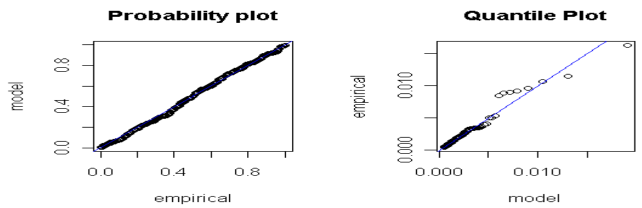

Figure 27: GPD fit of the exceedances of $Em - \widehat{Em} - \hat{X}$ over 0.0005 threshold.

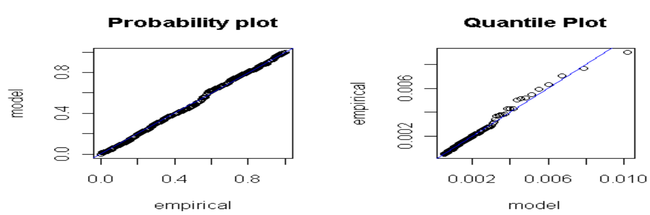

Figure 28: GPD fit of the exceedances of $-(Em - \widehat{Em} - \hat{X})$ over 0.0005 threshold.

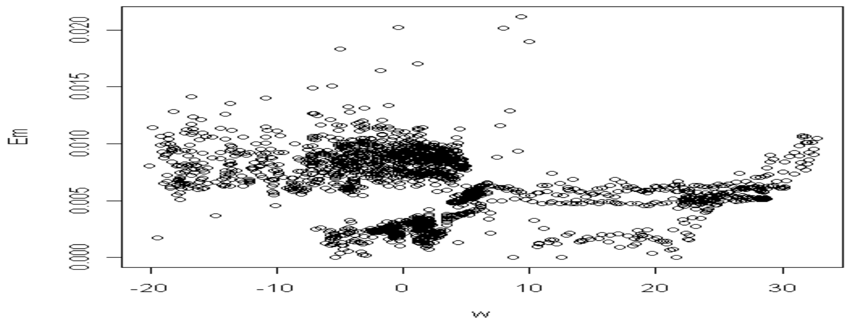

Figure 29: The scatter-plot of the dependence between the water temperature and the Emission in real data

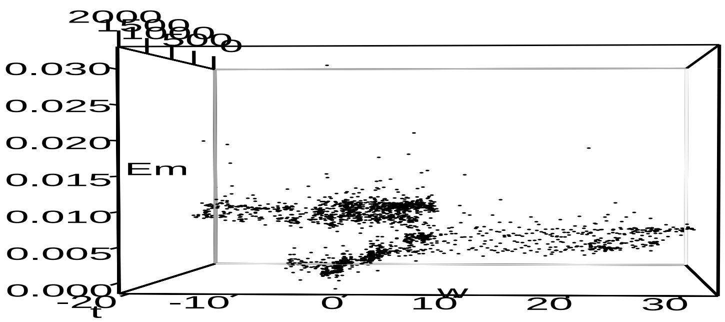

Figure 30: Empirical check of the model

2.2.1. Empirical check of the model

In this paragraph we check by simulation study that model (9) can produce similar behavior as the initial methane emission data. On Figure 29 is given the scatter-plot of the dependence between the water temperature and the emission in real data.

In this simulation study we used symmetric thresholds $c_1 = c_2 = 0.0005$. However, $Z's$ are not symmetric, we use $\alpha_1 = 1.91$ and $\alpha_2 = 3.03$, as obtained in the above study. Figure 30 shows the obtained simulated emission from this model.

2.3. Infinite moving average model, regressors: time, water level,soil temperatures

The main aim of the paper is to find a simplistic and reliable time series model of the methane emission data in ecosystem in South Bohemia. In particular, when we consider response surface model, where dependent variable is methane emission and independent variables (regressors) are water level w and 5 soil temperatures $t_0, t_5, t_{10}, t_{20}, t_{30}$. When we search for a response surface model with most of interactions to be estimated, and we consider only the first 50 observations, estimable are: intercept, cubic model in water level and t_{30}, quadratic model in t_0, t_5, t_{10}, t_{20}, several pair

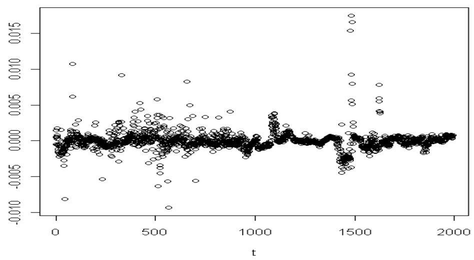

Figure 31: The plot of $Em - \widehat{Em}$.

interactions even with higher powers of water level and some of trilinear interactions are estimable. This computation was made in *Cocoa 4.7.5.* software ([1]) and it also justifies the ecological intuition that water level together with soil temperatures are very important factors. When employing first 100 of time observations, we are able to fit already Quartic model in water level, and also richer models for other regressors, but we need assurance that coefficients of response surface model does not vary significantly with time. Thus understanding of the time variation of the process, which is the main aim of this paper, is the first step for a solid model building. The latter will be further logical step of the investigation.

In this section by OLS procedure we estimate a regression model with deterministic part

$$\widehat{Em}(t) = a + b.t + c.t^2 + d.w + e.w^2 + f.t_0 + g.t_0^2 + h.t_5 + k.t_5^2 + bilinearinteractions,$$

i.e.

$$Em(t) = \widehat{Em}(t) + X(t) + Z(t) - Z_{-}(t) + \epsilon(t).I_{[-c_1,c_2]}(\epsilon(t)), \quad t \geq 0, \tag{10}$$

where the bilinear interactions are $tw, tt_0, tt_5, wt_0, wt_5, t_0t_5$. Here t stands for time, w is the water level, t_0, t_5 are the soil temperatures with levels $0, -5cm$ respectively, and $a, b, c, d, e, f, g, h, k$ are unknown coefficients of non-interactions in the response surface model. The processes X, Z, Z_{-} and ϵ are as in the model (8).

For the sake of simplicity and parsimony of the model we have not introduced soil temperatures bellow $-5cm$ since the explanation of the resting variability was negligible. We received the coefficients in $\widehat{Em}(t)$ by OLS procedure *lm* in R software.

```
Coefficients:
(Intercept)       t           t^2          w           t_0         t_5          w^2
-1.680e-03   7.524e-06  -2.812e-09  2.624e-04  4.048e-04   7.270e-04  5.681e-07

t_0^2        t_5^2       tw          tt_0        t:t_5       wt_0       wt_5       t_0:t_5
2.203e-04 2.052e-04 -1.401e-07 3.457e-07  -8.222e-07 7.237e-07 -1.484e-05 -4.506e-04
```

The residual standard error is 0.001349 and the plot of $Em - \widehat{Em}$ is given on Figure 31.

The partial autocorrelation function of X is plotted on Figure 32.

We modeled X by causal AR(12). We use the Yule-Walker estimators (see e.g. [4]) for the coefficients of the model and obtained in software R 2.15.00.

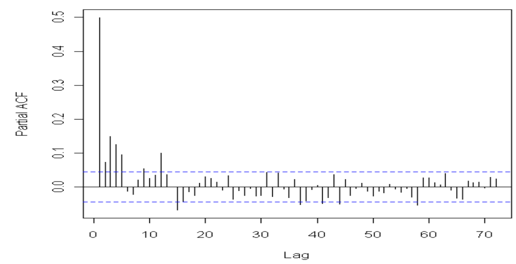

Figure 32: ACF of the process $Em - \widehat{Em}$.

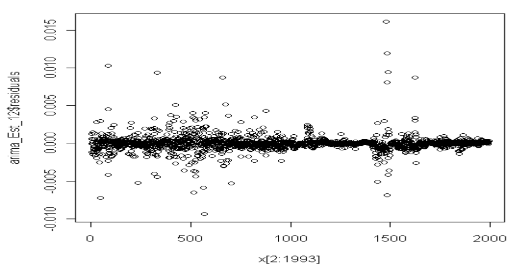

Figure 33: The plot of $Em - \widehat{Em} - \hat{X}$.

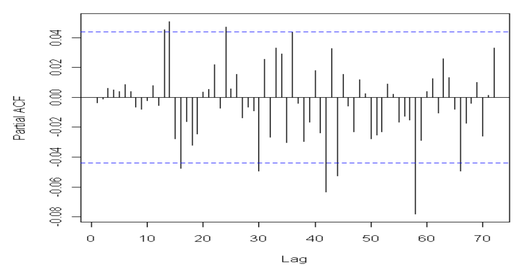

Figure 34: Partial ACF of the residuals $Em - \widehat{Em} - \hat{X}$.

```
Coefficients:
         ar1      ar2     ar3     ar4     ar5      ar6      ar7      ar8
      0.4160  -0.0028  0.0899  0.0809  0.0958  -0.0115  -0.0453  -0.0125
s.e.  0.0223   0.0242  0.0242  0.0242  0.0243   0.0244   0.0244   0.0243
         ar9     ar10     ar11    ar12  intercept
      0.0332   0.0103  -0.0069  0.0989      0e+00
s.e.  0.0242   0.0242   0.0241  0.0223      1e-04

sigma^2 estimated as 1.264e-06:  log likelihood = 10700.21,  aic = -21372.43
```

The errors of $Em - \widehat{Em} - \hat{X}$ are plotted on Figure 33. Their partial autocorrelation function shows that these residuals are almost independent (See Figure 34).

The histogram of the residuals $Em - \widehat{Em} - \hat{X}$ is given on Figure 35. The mean-excess plots of the positive and negative parts of $Em - \widehat{Em} - \hat{X}$ (Figure 36 and Figure 37) show that the observed random variables have Pareto like distribution.

The Hill plots for the positive and the negative parts of $Em - \widehat{Em} - \hat{X}$ are given on Figure 38 and 40.

The t-Hill plots for the positive and negative parts of $Em - \widehat{Em} - \hat{X}$ are given correspondingly on Figure 39 and Figure 41.

We made the generalized Pareto fit of the exceedancess of $Em - \widehat{Em} - \hat{X}$ over $c_2 = 0.0005$ threshold and obtained that the observed random variable is Pareto distributed, with $\alpha_1 = 1.925528$ (see Figure 42). Analogously the generalized Pareto fit of the exceedances of $-(Em - \widehat{Em} - \hat{X})$

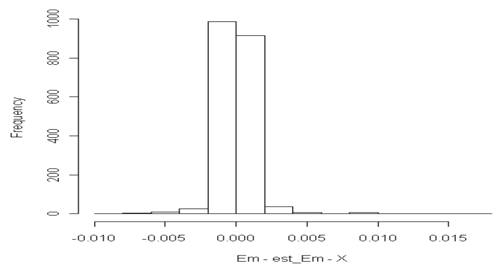

Figure 35: The histogram of the positive parts of $Em - \widehat{Em} - \hat{X}$.

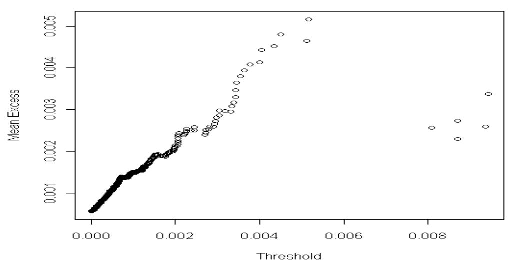

Figure 36: The mean excess plot of the positive parts of $Em - \widehat{Em} - \hat{X}$.

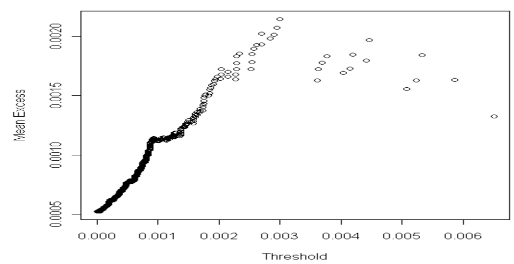

Figure 37: The mean excess plot of the negative parts of $Em - \widehat{Em} - \hat{X}$.

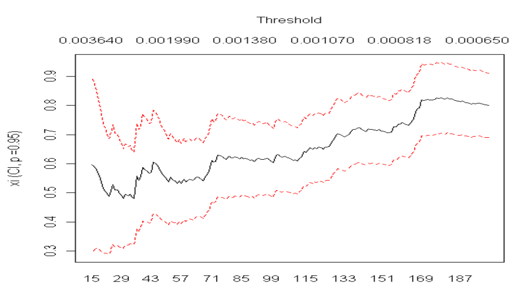

Figure 38: The Hill plot of the positive parts of $Em - \widehat{Em} - \hat{X}$.

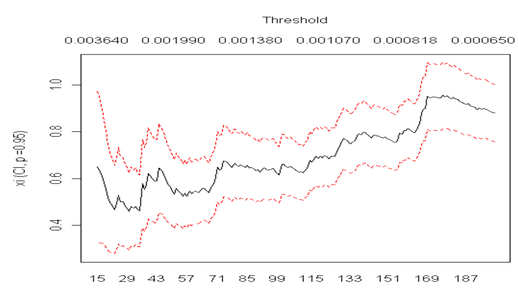

Figure 39: The t-Hill plot of the positive parts of $Em - \widehat{Em} - \hat{X}$

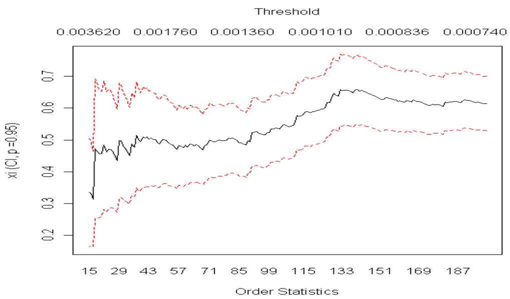

Figure 40: The Hill plot of the negative parts of $Em - \widehat{Em} - \hat{X}$

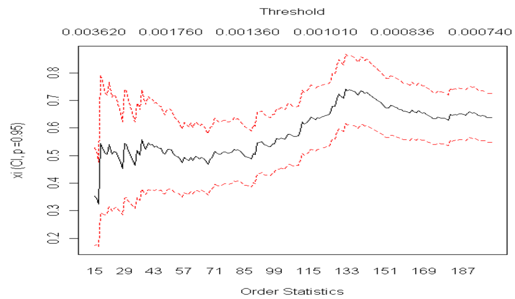

Figure 41: The t-Hill plot of the negative parts of $Em - \widehat{Em} - \hat{X}$

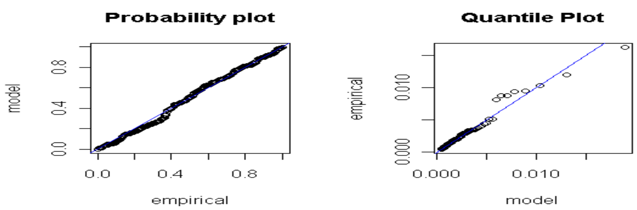

Figure 42: GPD fit of the exceedances of $Em - \widehat{Em} - \hat{X}$ over 0.0005 threshold.

over $c_1 = 0.0005$ threshold shows that it is Pareto(2.631860) distributed (see Figure 43).

3. Simulation study

In this section we consider Hill and t-Hill plots as defined above to compare the properties of both estimators. Further on n is fixed, the dotted line presents the Hill plot the solid line is for t-Hill plot. We investigate two cases.

Contaminated data from iid Pareto sample.

We simulated 1000 samples with $n = 2500$ observations of r.v. with distribution function

$$F(x) = (1 - \frac{1}{x^{\alpha}}).0,9 + (1 - \frac{1}{x^{\delta}}).0,1, \quad x > 1$$

for $\alpha = 0.3, 0.4, 1, 1.2, 1.7$ and $\delta = 0.1, 0.5, 1$. Then we calculated separately both averages, of the Hill and t-Hill estimators for $k(n) = 1, 2, ..., n$, together with the corresponding 0.95-confidence intervals.

When $\alpha < 1$ and $\delta \geq 1$ the Hill estimators are better than the t-Hill estimators. When $\alpha < 1$ and $\delta < 1$, or $\alpha \geq 1$ and $\delta > 1$ the t-Hill estimators are comparable to the corresponding Hill estimators.

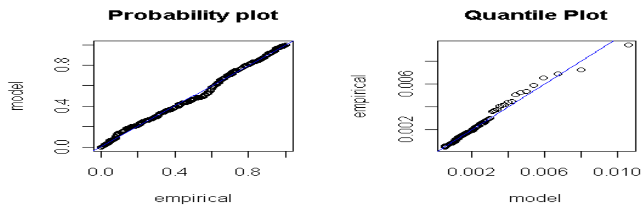

Figure 43: GPD fit of the exceedances of $-(Em - \widehat{Em} - \hat{X})$ over 0.0005 threshold.

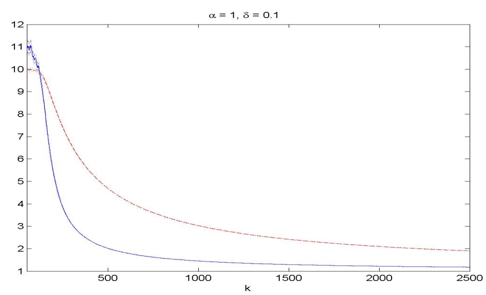

Figure 44: $\alpha = 1, \delta = 0.1$

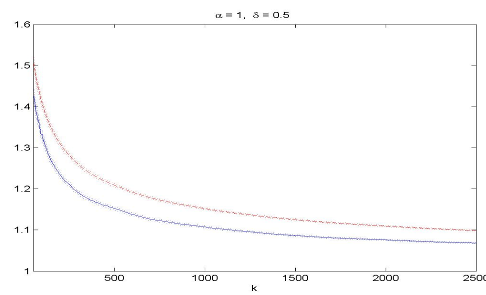

Figure 45: $\alpha = 1, \delta = 0.5$

The Figures 44-49 show the rate of convergence of the t-Hill and Hill estimators.

For $\alpha \geq 1$ and $\delta \leq 1$, the t-Hill estimators turned out to be better than the Hill estimators. This investigation refine the conclusions made by [8], [9] and [25].

Moving average process. Further on we consider the infinite moving average process with the following auto regressive form

$$\mathbf{X}_i = 1.3\mathbf{X}_{i-1} - 0.7\mathbf{X}_{i-2} + \mathbf{Z}_i, \quad i = 1, 2, ..., n.$$

This process is studied in [20] with respect to the Hill estimator.

- Pareto noise. Here $\mathbf{Z}_i$, $i = 1, 2, ...$ are independent and Pareto(α) distributed. In each of the following examples we simulated 1000 samples of $n = 2500$ such data. For each sample we calculated the moving averages $\mathbf{X}_i, i = 1, 2, ..., n$ and finally plotted the Hill and the t-Hill plots of the average of the corresponding estimators together with their *0.95*-confidence intervals.

  The Figures 50- 53 show that both estimators have similar behavior for fixed number of upper order statistics. In this case we need $k(n) = o(n)$ in order to obtain weakly consistent estimators. Therefore they are appropriate only for very large samples. On our plots we have $k(n) = 10, 11, ..., 500$.

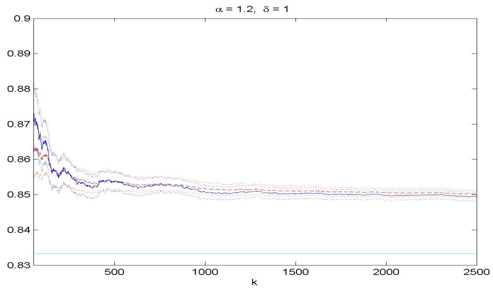

Figure 46: $\alpha = 1.2, \delta = 1$

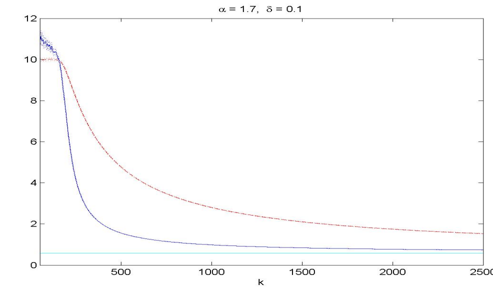

Figure 47: $\alpha = 1.7, \delta = 0.1$

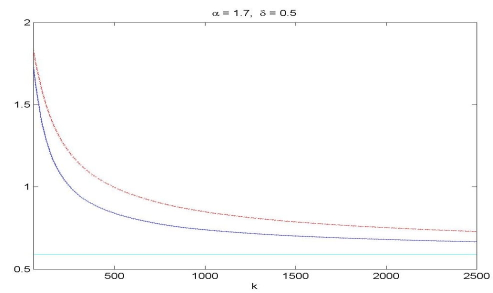

Figure 48: $\alpha = 1.7, \delta = 0.5$

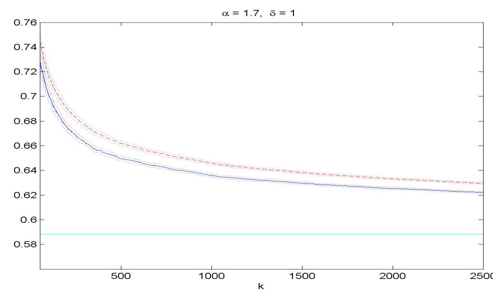

Figure 49: $\alpha = 1.7, \delta = 1$

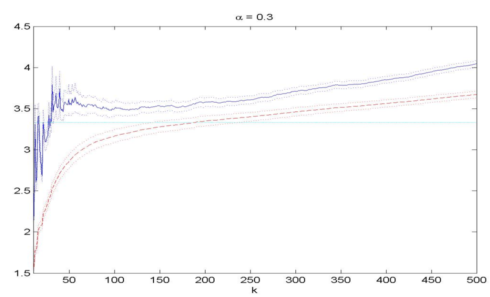

Figure 50: $\alpha = 0.3$

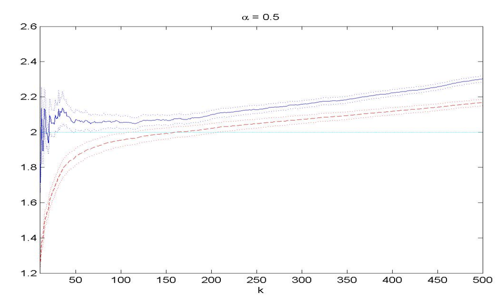

Figure 51: $\alpha = 0.5$

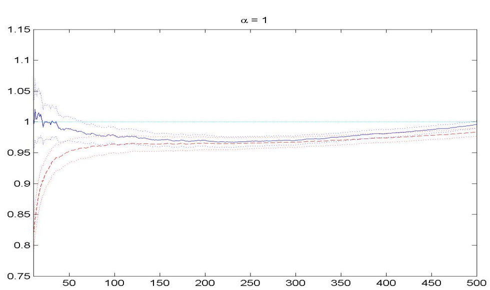

Figure 52: $\alpha = 1$

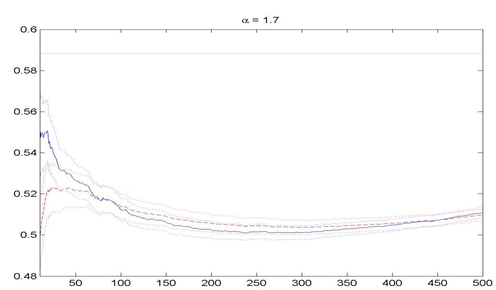

Figure 53: $\alpha = 1.7$

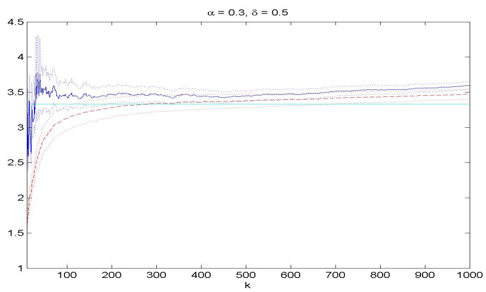

Figure 54: $\alpha = 0.3, \delta = 0.5$

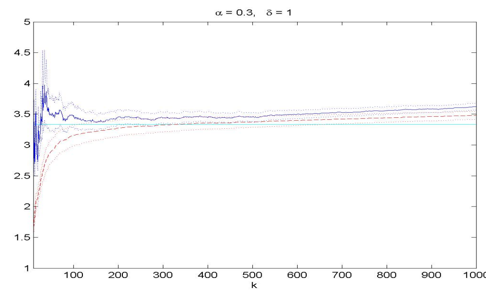

Figure 55: $\alpha = 0.3, \delta = 1$

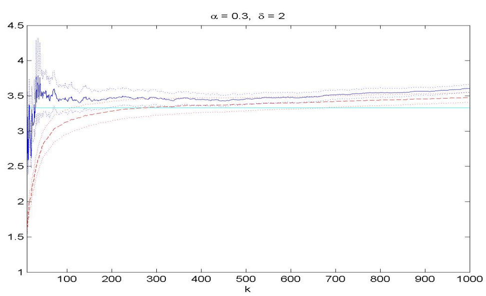

Figure 56: $\alpha = 0.3, \delta = 2$

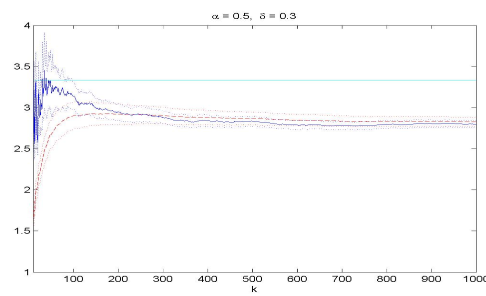

Figure 57: $\alpha = 0.5, \delta = 0.3$

- Moving average data with contaminated Pareto noise. In this case the innovations $\mathbf{Z}_i$, $i = 1, 2, ..., n$ are independent and

  $$G(x) = (1 - \frac{1}{x^\alpha}).0,9 + (1 - \frac{1}{x^\delta}).0,1, \quad x > 1.$$

  We made 100 samples of $n = 10000$ observations and calculated $\mathbf{X}_i, i = 1, 2, ..., n$. Then we plotted the Hill and the t-Hill plots of the averages of the corresponding estimators together with their *0.95*-confidence intervals for different k. The results are very different from the case of i.i.d. data.

  It is well known that the MA data are dependent. In view of the regular variation of the tail of the distribution of the noise component the largest values are determined by $\min\{\alpha, \delta\}$. If some innovation is huge it influences the following moving averages. Therefore both estimators are not robust for $1/\alpha$, however, there are good estimators for $1/\min\{\alpha, \delta\}$ (see Figures 54-59).

The rate of the convergence increases with α and therefore if our parameter is smaller we will need more observations in order to obtain good estimation.

4. Discussion and Conclusions

The main aim of the paper is to find a parsimonious and reliable time series model of the methane emission data in ecosystem in South Bohemia. In particular we consider response surface model, where independent variable is methane emission and dependent variables (regressors) are water level

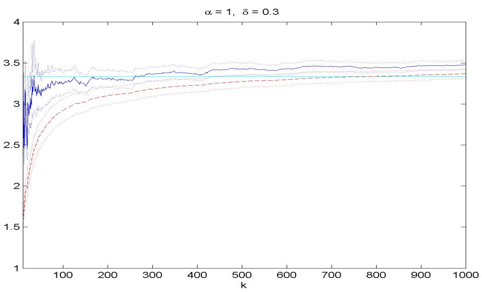

Figure 58: $\alpha = 1, \delta = 0.3$

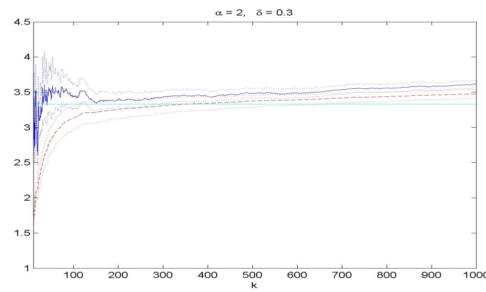

Figure 59: $\alpha = 2, \delta = 0.3$

and 5 soil temperatures in different depths. When we search for a response surface model with most of possible interactions to be estimated, for considering only the first 50 observations intercept, cubic model in water level and t_{30}, quadratic models in other soil temperatures, several pair interactions even with higher powers of water level and some of trilinear interactions are estimable. This justifies the ecological intuition that water level together with soil temperatures are very important factors. When employing first 100 of time observations, we are able to fit already Quartic model in water level, and also richer models for other regressors, but we need assurance that coefficients of response surface model does not vary significantly with time. Thus understanding of time variation of the process, which is one of the main aims of this paper, is the fundamental step for a solid model building. The latter will be further logical step of the investigation.

To summarize, we introduced the infinite moving average model with Pareto distributed positive parts of innovations for the statistically efficient estimation of the methane emission from the sedge-grass marsh, South Bohemia, Czech Republic. We provide the proof of weak consistency of t-Hill estimator for the case of dependent data. t-Hill estimator has a simple form and it provides a nice trade-off between efficiency and robustness. Precise modeling of the methane emission from the wetland ecosystems can be helpful to understand the rate and dynamics of emissions related to biometeorological factors in site. Production and potential consumption of methane and possible followed releasing from the soil are processes related to many factors. These factors are soil temperature, water level and also more complex factors, which are difficulty measured directly, such as availability of organic matter for bacteria and their competition of organic matter. We do not consider the water level as a function of time because its complicated form. The further research will be done to investigate this issue in more details. Anyhow, we have checked by simulation that both models (8) and (9) are fitting the data well. More calibration (e.g. of thresholds c_1, c_2) may give a potential for even a better fit.

Appendix

We will use the following inequality for distribution functions with regularly varying tails. If a function H is regularly varying with exponent $\gamma \in \mathbb{R}$, then for $\varepsilon > 0$ there exist $t_0(\varepsilon)$, such that for $t > t_0(\varepsilon)$ and $x \geq 1$,

$$(1-\varepsilon)x^{\gamma-\varepsilon} \leq \frac{H(tx)}{H(t)} \leq (1+\varepsilon)x^{\gamma+\varepsilon}.$$

Its proof could be found in [14].

Proof of the Theorem. Analogously to [20] we define

$$H_{X,k(n),n} := \frac{1}{k(n)} \sum_{i=1}^{k(n)} \left\{ \frac{\mathbf{X}_{(i,n)}}{\mathbf{X}_{(k(n)+1,n)}} \right\}^{-1}$$

and

$$\widehat{\mu}_{X,k(n),n}(\cdot) := \frac{1}{k(n)} \sum_{i=1}^{n} \varepsilon \left\{ \frac{\mathbf{X}_i}{\mathbf{X}_{(k(n),n)}} \in (\cdot) \right\}, \tag{11}$$

where $H_{X,k(n),n}$ is the denominator of (1). Integration by parts entails

$$H_{X,k(n),n} = \int_1^\infty \frac{1}{x} \left\{ \frac{1}{k(n)} \sum_{i=1}^{n} \varepsilon \left\{ \frac{\mathbf{X}_i}{\mathbf{X}_{(k(n)+1,n)}} \in dx \right\} \right\} =$$

$$= \int_1^\infty \frac{1}{x} \, \widehat{\mu}_{X,k(n),n}\,(dx) = -\int_1^\infty \frac{1}{x} \, d\,\widehat{\mu}_{X,k(n),n}\,(x,\infty] =$$

$$= \lim_{x\to 1} \frac{1}{x}\, \widehat{\mu}_{X,k(n),n}\,(x,\infty] - \lim_{x\to\infty} \frac{1}{x}\, \widehat{\mu}_{X,k(n),n}\,(x,\infty] + \int_1^\infty \widehat{\mu}_{X,k(n),n}\,(x,\infty]\, d\,\frac{1}{x}.$$

The positive measures $\widehat{\mu}_{X,k(n),n}$ are Radon, therefore

$$\lim_{x\to\infty} \frac{1}{x}\, \widehat{\mu}_{X,k(n),n}\,(x,\infty] = 0$$

and

$$H_{X,k(n),n} = \widehat{\mu}_{X,k(n),n}\,(1,\infty] - \int_1^\infty \widehat{\mu}_{X,k(n),n}\,(x,\infty]\, \frac{1}{x^2}\, dx.$$

[20] prove that conditions of our theorem imply

$$\hat{\mu}_{X,k(n),n} \Longrightarrow \mu, \quad n \to \infty.$$

In view of the definition of μ we have

$$\lim_{n\to\infty} \widehat{\mu}_{X,k(n),n}\,(1,\infty] = 1.$$

Then

$$\lim_{n\to\infty} H_{X,k(n),n} = 1 - \lim_{n\to\infty} \{ \int_1^t \widehat{\mu}_{X,k(n),n}\,(x,\infty]\, \frac{1}{x^2}\, dx - \int_t^\infty \widehat{\mu}_{X,k(n),n}\,(x,\infty]\, \frac{1}{x^2}\, dx \}.$$

Consider

$$\lim_{n\to\infty} \int_1^t \widehat{\mu}_{X,k(n),n}\,(x,\infty]\, \frac{1}{x^2}\, dx = \int_1^t \mu\,(x,\infty]\, \frac{1}{x^2}\, dx = \frac{t^{-\alpha-1}-1}{-\alpha-1}$$

and

$$\lim_{t\to\infty} \lim_{n\to\infty} \int_1^t \widehat{\mu}_{X,k(n),n}\,(x,\infty]\, \frac{1}{x^2}\, dx = \frac{1}{\alpha+1}.$$

In order to apply Th. 4.2. of [3] we have to check that

$$\lim_{t\to\infty}\lim_{n\to\infty} P\left\{\int_t^\infty \widehat{\mu}_{X,k(n),n}\,(x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon\right\} = 0. \tag{12}$$

We need to change $\widehat{\mu}_{X,k(n),n}$ with $\mu_{X,k(n),n}$. We choose $\delta > 0$ and consider

$$P\left\{\int_t^\infty \widehat{\mu}_{X,k(n),n}\,(x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon\right\} =$$

$$= P\left\{\int_t^\infty \mu_{X,k(n),n}\,(\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})}x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon\right\} =$$

$$= P\left\{\int_t^\infty \mu_{X,k(n),n}\,(\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})}x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon, |\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})} - 1| < \delta\right\} +$$

$$+ P\left\{\int_t^\infty \mu_{X,k(n),n}\,(\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})}x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon, |\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})} - 1| \geq \delta\right\}.$$

[20] prove that conditions (3), (4) and (5) imply

$$\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})} \xrightarrow{\mathbb{P}} 1, \quad n \to \infty. \tag{13}$$

Therefore

$$0 \leq P\left\{\int_t^\infty \mu_{X,k(n),n}\,(\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})}x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon, |\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})} - 1| \geq \delta\right\} \leq$$

$$\leq P\left\{|\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})} - 1| \geq \delta\right\}$$

converges to zero when $n \to \infty$.

In view of (13), there exists $0 < \delta_0 < 1 + \alpha$ and n_{δ_0} - large enough such that for all $t > 1$ and $n > n_{\delta_0}$,

$$P\left\{\int_t^\infty \mu_{X,k(n),n}\,(\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})}x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon, |\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})} - 1| < \delta\right\} \leq$$

$$\leq P\left\{\int_t^\infty \mu_{X,k(n),n}\,((1-\delta_0)x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon, |\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})} - 1| < \delta\right\} \leq$$

$$\leq P\left\{\int_t^\infty \mu_{X,k(n),n}\,((1-\delta_0)x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon\right\}.$$

Chebishev's inequality guarantee that

$$P\left\{\int_t^\infty \mu_{X,k(n),n}\,((1-\delta_0)x,\infty]\,\frac{1}{x^2}\,dx > \varepsilon\right\} \leq \frac{1}{\varepsilon} E\int_t^\infty \mu_{X,k(n),n}\,((1-\delta_0)x,\infty]\,\frac{1}{x^2}\,dx.$$

Identical distribution of $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ entails
$E\mu_{X,k(n),n}\left((1-\delta_0)x, \infty\right] =$

$$= \frac{1}{k(n)}\sum_{i=1}^{n} P\left\{\frac{\mathbf{X}_i}{b(\frac{n}{k(n)})} \in ((1-\delta_0)x, \infty]\right\} = \frac{n}{k(n)} P\left\{\frac{\mathbf{X}_1}{b(\frac{n}{k(n)})} \in ((1-\delta_0)x, \infty]\right\} =$$

$$= \frac{n}{k(n)}\overline{F}((1-\delta_0)x.b(\frac{n}{k(n)})) = \frac{n}{k(n)}\overline{F}(b(\frac{n}{k(n)}))\frac{\overline{F}((1-\delta_0)x.b(\frac{n}{k(n)}))}{\overline{F}(b(\frac{n}{k(n)}))}.$$

Definition of b imply that when (4) is satisfied, then $b(\frac{n}{k(n)}) \to \infty$ and

$$\lim_{n\to\infty} \frac{n}{k(n)}\overline{F}(b(\frac{n}{k(n)})) = 1.$$

Thus, in view of (3) and de Haan's inequality for regularly varying functions, there exists $0 < \delta_1 < 1 + \alpha$ and n_{δ_1} - large enough such that for all $x > 1$ and $n > n_{\delta_1}$,

$$E\mu_{X,k(n),n}\left((1-\delta_0)x, \infty\right] \le (1+\delta_1)((1-\delta_0)x)^{-\alpha+\delta_1}$$

and for all $t > 1$,

$$\int_t^\infty E\mu_{X,k(n),n}\left((1-\delta_0)x, \infty\right] \frac{1}{x^2}\, dx \le (1-\delta_0)^{-\alpha+\delta_1}\int_t^\infty (1+\delta_1)x^{-\alpha+\delta_1-2}\, dx =$$

$$= -(1-\delta_0)^{-\alpha+\delta_1}(1+\delta_1).\frac{t^{-\alpha+\delta_1-1}}{-\alpha+\delta_1-1} = (1-\delta_0)^{-\alpha+\delta_1}(1+\delta_1).\frac{t^{-\alpha+\delta_1-1}}{\alpha-\delta_1+1} < \infty.$$

We apply Fubini's theorem and obtain that for all $t > 1$ and $n > n_\delta$,

$$P\left\{\int_t^\infty \mu_{X,k(n),n}\left((1-\delta_0)x, \infty\right] \frac{1}{x^2}\, dx > \varepsilon\right\}$$

$$\le \frac{1}{\varepsilon} E\int_t^\infty \mu_{X,k(n),n}\left((1-\delta_0)x, \infty\right] \frac{1}{x^2}\, dx =$$

$$= \frac{1}{\varepsilon}\int_t^\infty E\mu_{X,k(n),n}\left((1-\delta_0)x, \infty\right] \frac{1}{x^2}\, dx \le \frac{(1-\delta_0)^{-\alpha+\delta_1}(1+\delta_1)}{\varepsilon}.\frac{t^{-\alpha+\delta_1-1}}{\alpha-\delta_1+1}.$$

That is why

$$\lim_{t\to\infty}\lim_{n\to\infty} P\left\{\int_t^\infty \mu_{X,k(n),n}\left(\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})}x, \infty\right] \frac{1}{x^2}\, dx > \varepsilon, |\frac{\mathbf{X}_{(k(n),n)}}{b(\frac{n}{k(n)})} - 1| < \delta\right\} = 0$$

and (12) is satisfied.

$\square$

Proof of the Corollary. We check the conditions of the previous theorem. The random, variables $\mathbf{X}_n$, $-\infty < n < \infty$, are identically distributed.

[20] prove that

$$\mu_{X,k(n),n} \Longrightarrow \mu, \quad n \to \infty.$$

[5] proves that under these settings

$$\overline{F}(x) = P\left(\sum_{j=0}^{\infty} c_j \mathbf{Z}_j > x\right) \sim \sum_{j:c_j>0}^{\infty} c_j^{\alpha} \overline{G}(x) \in RV_{-\alpha}.$$

In view of the Mason's condition we complete the proof. $\square$

Acknowledgement

Monitoring of the methane emissions are supported by the project: Role of plants in the greenhouse gas budget of a sedge fen (P504/11/1151) from the Grant Agency of the Czech Republic. We acknowledge Kathrin Muck for typing the part of Methane data to Cocoa 4.7.5.

The first author is partially supported by Grant No. RD-05-278/2012 of Shumen University, Bulgaria.

References

[1] J. Abbott and A. Bigatti and M. Kreuzer and L. Robbiano, Computing Ideals of Points, JSYMC 30 (2000) 4: 341–356.

[2] J. Beran, D. Schell, On robust tail index estimation, Computational Statistics and Data Analysis, doi:10.1016/j.csda.2010.05.028 (2010).

[3] P. Billingsley, Convergence of Probability Measures, John Wiley and Sons, 1977.

[4] P.J. Brockwell, R.A. Davis, Introduction to Time series and forecasting, 2nd Edition, Springer texts in Statistics, 2002.

[5] D. Cline, Infinite series of random variables with regularly varying tailes, Tech.Report No 83-24 (1983) Indtitute of Applied Mathematics and Statistics, University of British Columbia.

[6] J. Dušek, H. Cížkov, R. Czerný, K. Taufarová, M. Šmídov, D. Janouš, Influence of summer flood on the net ecosystem exchange of CO2 in a temperate sedge-grass marsh, Agricultural and Forest Meteorology 149 (2009) 1524–1530.

[7] P. Embrechts, C. Klueppelberg, T. Mikosch, Modeling extremal events for insurance and finance, Springer Verlag, 1997.

[8] Z. Fabián, Induced cores and their use in robust parametric estimation, Communication in Statistics, Theory Methods 30 (2001) 537-556.

[9] Z. Fabián, M. Stehlík, A note on favorable estimation when data is contaminated, Communications in Dependability and Quality Management 11:4 (2008) 36-43.

[10] Z. Fabián, M. Stehlík, On robust and distribution sensitive Hill like method, IFAS res. report.43 (2009).

[11] P.D. Feigin and S. I. Resnick, Limit dsitributions for linear programming time series estimators, Stoch. Processes Appl. 51, 135–165.

[12] J.L. Garcia, B.K.C.Patel, B. Ollivier, Taxonomic, phylogenetic, and ecological diversity of methanogenic Archaea. Anaerobe. 6:4 (2000) 205-226.

[13] T. Gierczak, R.K. Talukdar, S.C. Herndon, G.L. Vaghjiani, A.R. Ravishankara, Rate coefficients for the reactions of hydroxyl radical with methane and deuterated methanes, J. Phys. Chem. A 101 (1997) 3125-3134.

[14] L. de Haan, On Regular Variation and Its Application to the Weak Convergence of Sample Extremes, Mathematical Centre Tract 32,(1970) Mathematics Centre, Amsterdam, Holland.

[15] B. Hill, A Simple General Approach to Inference About the Tail of a Distribution, The Annals of Statistics 3 (1975) 1163-1173.

[16] 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.

[17] T. Li, Y. Huang, W. Zhang, Ch. Song, CH4MODwetland: A biogeophysical model for simulating methane emissions from natural wetlands, Ecological Modelling 221 (2010) 666-680.

[18] D.M. Mason, Laws of Large Numbers for Sums of Extreme Values, Ann. Probab. 10 (1982) 754-764.

[R Development Core Team] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, ISBN 3-900051-07-0, http://www.R-project.org

[20] S. Resnick, C. Starica, Consistency of Hill's Estimator for Dependent Data. Technical report No. 1077, School of Operational Research and Industrial Engeneering Cornell University Itaca NY 14853, (1993) 43 - 50.

[21] S.I. Resnick, Heavy-Tail Phenomena, Springer, New York, 2007.

[22] L.J. Rodríguez-Aragón, J. López-Fidalgo, Optimal designs for the Arrhenius equation, Chemometr. Intell. Lab. Syst. 77 (2005) 131-138.

[23] J.M. Rodríguez-Díaz, M.T. Santos-Martín, Study of the best designs for modifications of the Arrhenius equation, Chemometr. Intell. Lab. Syst. 95 (2009) 199-208.

[24] J.M. Rodríguez-Díaz, M.T. Santos-Martín, H. Waldl, M. Stehlík, Filling and D-optimal designs for the correlated Generalized Exponential models, Chemometrics and Intelligent Laboratory Systems 114 (2012) 10-18.

[25] M. Stehlík, R. Potocký, H. Waldl, Z. Fabián, On the favourable estimation of fitting heavy tailed data,Computational Statistics 25 (2010) 485-503.

[26] M. Stehlík, Z. Fabián, L. Střelec, Small sample robust testing for Normality against Pareto tails, Communications in Statistics - Simulation and Computation 41:7 (2012) 1167-1194.

[27] US Environmental Protection Agency (EPA) (Ed.), Inventory of US greenhouse gas emissions and sinks: 19902005. US Environmental Protection Agency, Washington, (2007) 1-16.

[28] G.L. Vaghjiani, A.R. Ravishankara, New measurement of the rate coefficient for the reaction of OH with methane, Nature 350 (1991) 406-409.