EcoDep 2022
Conference on Networks Reconstruction
Time trends in atmospheric ethane

Federico Maddanu [1]
CY University
Tommaso Proietti
Università di Roma "Tor Vergata"

June 11, 2022

[1] Address for Correspondence: AGM Laboratory, Cergy University, Av. Adolphe Chauvin 95302 CERGY-PONTOISE, France. Email: federico.maddanu@gmail.com.

Introduction

1. Understanding ethane (C_2H_6) dynamics is of crucial importance in the context of climate change.
2. Ethane affects the distribution of ozone (O_3) and the formation of ground level ozone has pollution effects on the air quality and damages ecosystems.
3. Ethane is an indirect greenhouse gas, which influences the atmospheric lifetime of methane (CH_4), such that ethane emissions can be used as a measure of methane emissions.
4. Since methane is released in the atmosphere by both natural events and anthropogenic activities, while ethane emissions do not have significant natural sources, monitoring ethane levels becomes crucial to measure the human contribution to the abundance of methane in the atmosphere.

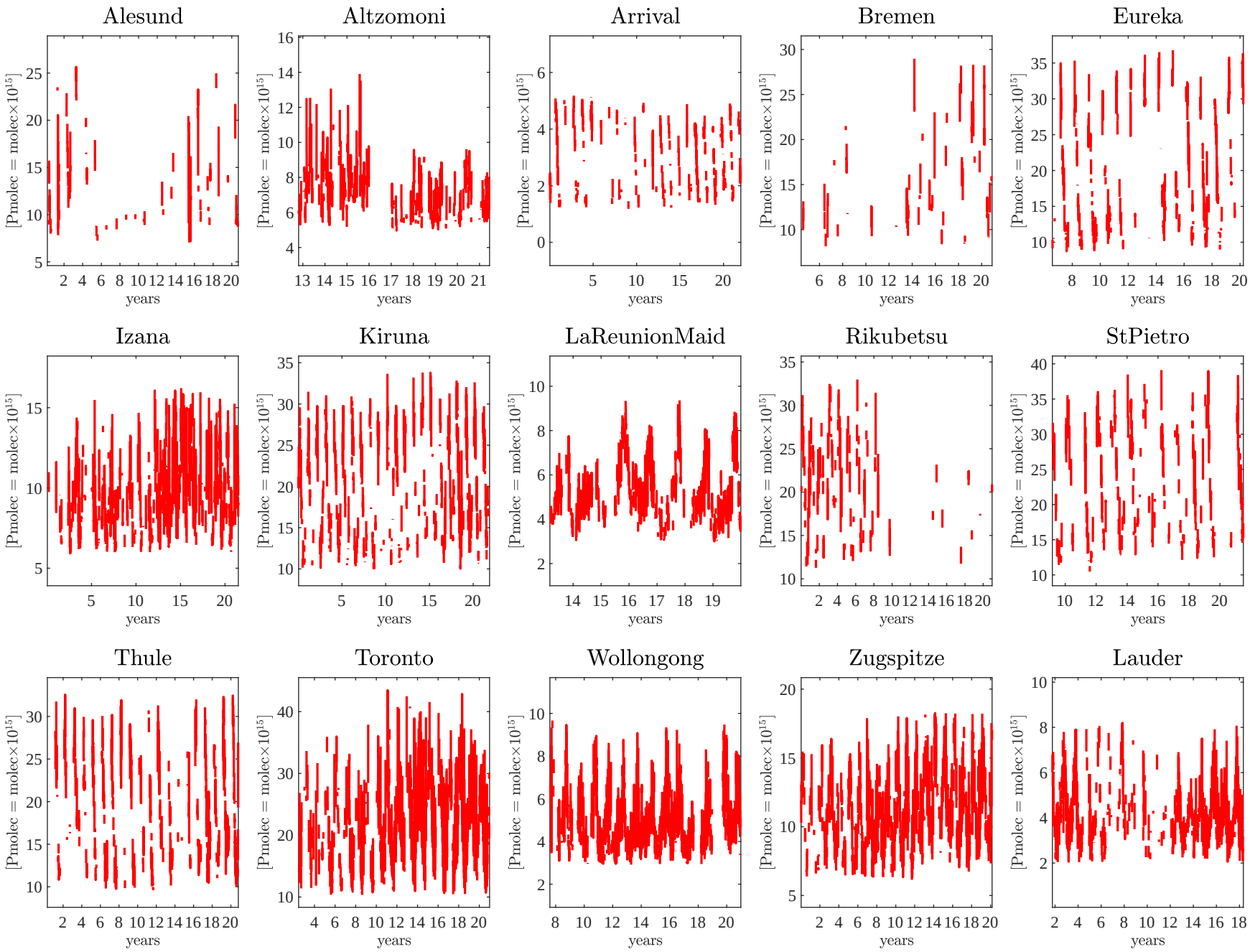
Introduction

- The main literature (see Franco et al. 2015, 2016 and Helmig et al., 2016) reports that ethane emissions in the Northern Hemisphere may be mostly attributed to the production and transport of oil and natural gas.
- According to Helmig et al. (2016), biogenic and biomass burning accounts for about 22% of global ethane emissions against the remaining 78% attributed to anthropogenic activities.
- An increasing on ethane trends has been found in the troposphere in the period from 2009 to 2015, with a reversal after some decades of declining, the latter due to the regulation policies versus the main anthropogenic emitters.
- Then, a hiatus in the ethane growth has been detected in 2015-2018, whose nature may be only temporary and still need to be fully understood.

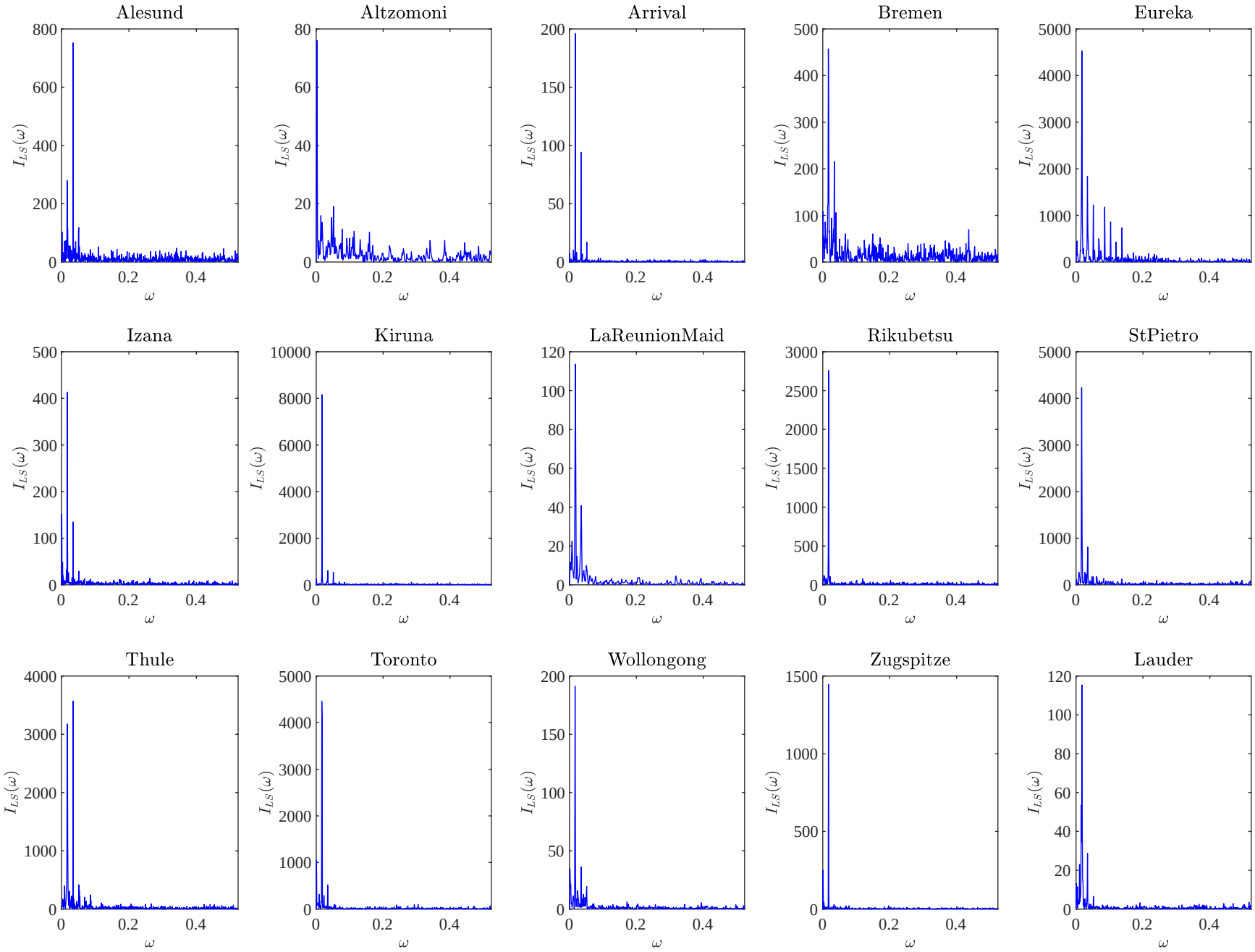
The Data

Figure: We investigate on the ground-based Fourier Transform Infrared (FTIR) solar spectra measurements of ethane abundance in the atmosphere recorded at 15 ground-stations.

Ethane time series



Ethane time series Lomb-Scargle periodograms



The Data

The ethane time series are characterized by

- **Strong seasonality** at the annual frequency, since ethane degrades faster in summer with high temperatures with respect to cold weather conditions.
- This implies that the time series is typically featured by peaks in wintertime generating seasonal **heteroskedasticity**.
- Because the measurements can only be taken under clear-sky conditions, the series share a large amount of **missing data**,

Finally, the high persistence of the annual cycle and the elevate amount of missing data, makes the detection by visual inspection of a hidden trend component a very difficult task.

Literature Review

- Previous studies in the literature (Gardiner et al. 2008, Franco et al. 2015, 2016, Hausmann et al., 2016 and Helmig et al., 2016) share the main approach in treating the ethane cycle as deterministic and proceed to remove it by regressing the data on a sum of sinusoidal terms.
- Analyses were then carried out on the residuals of such regression, by splitting the sample into two time periods and fitting a linear trend on both the subsamples, in order to detect changes in the trend direction.
- Friedrich et al. (2020) point out that the linear approach may obscure important characteristics in the data. They proposed a non-linear and non-parametric model, where the estimation of a non-linear trend allows to capture much interesting features from the series.

Paper's contribution

- There is no reason to treat the ethane cycle as deterministic. Since ethane depends on annual temperatures, it is known that phases and amplitudes of global temperatures are varying over years (see Proietti and Hillebrand, 2015 and Hillebrand and Proietti, 2017).
- This is why we propose an additive unobserved components model, where both the annual cycle and underlying trend evolve stochastically.
- The cycle results from the sum of M nonstationary fractional Sinusoidal Waveform (fSW) processes (see Proietti and Maddanu, 2022, who introduced the type I, or covariance stationary fSW process), which determine the persistence of the annual cycle at each one of the j harmonic frequencies for $j = 1, \cdots, M$.
- The model also encompasses the possibility of a deterministic cycle as a limiting case, such that the nature of the cyclical component can be suggested by the inference on the model parameters.

Paper's contribution

- Finally, the trend component is modelled as a type II fractional integrated process (cf. Marinucci and Robinson, 1999) at the zero frequency.
- Estimation is carried out by quasi-maximum likelihood and signal extraction is carried out by the Kalman filter (KF) and associated smoothing algorithm (cf. Durbin and Koopman, 2012).
- Missing values are dealt by the KF, skipping the updating operations once an observation is missing (see Durbin and Koopman, 2012 and Proietti and Luati, 2013).
- We believe that the proposed model and the associated learning methods make a valid contribution to the literature on the assessment of trend levels in atmospheric ethane abundance.
- Our analysis confirms the presence of a common trend in the Northern Hemisphere, which is increasing during the period 2009-2015, followed by a successive hiatus during 2015-2018, after which it seems to reverse its direction.

Methodology

- Let $d \in (0,1)$ denote the order of fractional integration. Define $\Delta = (1-L)$, the fractional noise (FN) process $a_t = (1-L)^{-d}\eta_t$ with $\eta_t \sim$ i.i.d. has the representation $a_t = \sum_{j=0}^{\infty} \varphi_j \eta_{t-j}$, where

$$\varphi_j = \frac{(d+j-1)!}{j!(d-1)!} \tag{1}$$

are the coefficients of the binomial expansion of Δ^{-d}, i.e., $\Delta^{-d} = \sum_{j=0}^{\infty} \varphi_j L^j$. The process is stationary if $d < 0.5$. A type II FN process is obtained by truncating the terms of the binomial expansion. In particular, define

$$\begin{aligned} a_t &= \Delta_+^{-d}\eta_t, \quad t = 1, 2, \ldots, \\ &= \eta_t + \varphi_1\eta_{t-1} + \cdots + \varphi_{t-1}\eta_1, \end{aligned} \tag{2}$$

where values of d in $(0,1)$ are admissible, such that the process is nonstationary. To allow for a nonzero mean, a_0, say, we write $a_t = a_0 + \Delta_+^{-d}\eta_t$.

Methodology: Specification

The model is specified as follows.

$$\begin{aligned} x_t &= a_{0t} + \psi_t + \varepsilon_t, & \varepsilon_t &\sim \text{i.i.d. } \mathrm{N}(0, \sigma_\varepsilon^2), \\ a_{0t} &= a_{00} + \Delta_+^{-d_0}\eta_{0t}, & \eta_{0t} &\sim \text{i.i.d. } \mathrm{N}(0, \sigma_{\eta,0}^2), \\ \psi_t &= \textstyle\sum_{j=1}^{M} \psi_{jt}. \end{aligned} \tag{3}$$

The j-th cyclical component is the nonstationary fSW process

$$\begin{aligned} \psi_{jt} &= a_{jt}\cos(\lambda_j t) + a_{jt}^*\sin(\lambda_j t), \\ a_{jt} &= a_{j0} + \Delta_+^{-d_j}\eta_{jt}, & \eta_{jt} &\sim \text{i.i.d. } \mathrm{N}(0, \sigma_{\eta j}^2), \\ a_{jt}^* &= a_{j0}^* + \Delta_+^{-d_j}\eta_{jt}^*, & \eta_{jt}^* &\sim \text{i.i.d. } \mathrm{N}(0, \sigma_{\eta j}^2). \end{aligned} \tag{4}$$

The disturbances ε_t, η_{0t}, η_{jt} and η_{jt}^* for $j = 1, 2, \cdots, M$, are assumed to be mutually uncorrelated.

Methodology: Inference

- Estimation of the model parameters is carried out by maximising an approximation to the true Gaussian likelihood. Direct maximization of the true likelihood is unfeasible, due both to the presence of missing values and since the inversion of the variance covariance matrix is computationally intensive in the nonstationary case.

- The model can be represented in state space and the likelihood evaluated with the support of the KF. However, notice that the components do not possess a markovian representation and the number of states needed to represent them increases with the sample size. See Chan and Palma (1998).

Methodology: Inference

- A finite state representation can be obtained by approximating the type II FN processes $a_t = \Delta_+^{-d}\eta_t$ by an ARMA process, following the approach proposed by Hartl and Jucknewitz (2022), who consider the ARMA(p, p):

$$\tilde{a}_t = \sum_{i=1}^{p} \phi_i(d)\tilde{a}_{t-i} + \eta_t + \sum_{i=1}^{p} \beta_i(d)\eta_{t-i} + \eta_{j,t}\,,\ \eta_t \sim \text{i.i.d. } \mathrm{N}(0, \sigma_\eta^2), \quad (5)$$

where the AR and MA coefficients are obtained by minimizing the mean square approximation error

$$\frac{1}{n}\sum_{j=1}^{n}(n-j+1)(\tilde{\varphi}_j(d) - \varphi_j)^2, \quad (6)$$

where $\tilde{\varphi}_j(d)$ depend on the ARMA coefficients via

$$\left(1 - \sum_{i=1}^{p}\phi_i(d)L^i\right)\left(\sum_{i=0}^{\infty}\tilde{\varphi}_i(d)L^j\right) = 1 + \sum_{i=1}^{p}\beta_i(d)L^i. \quad (7)$$

Methodology: Inference

The corresponding state space approximation of the main model has measurement and transition equations:

$$\begin{aligned} x_t &= \mathsf{z}_t'\boldsymbol{\alpha}_t + \mathsf{w}_t'\boldsymbol{\delta} + \varepsilon_t, \\ \boldsymbol{\alpha}_{t+1} &= \mathsf{T}\boldsymbol{\alpha}_t + \mathsf{h}\boldsymbol{\eta}_t. \end{aligned} \tag{8}$$

where $\mathsf{w}_t = (1, \cos(\lambda_1 t), \sin(\lambda_1 t), \ldots, \cos(\lambda_M t), \sin(\lambda_M t))'$, $\boldsymbol{\delta} = (a_{00}, a_{10}, a_{10}^*, \ldots, a_{M0}, a_{M0}^*)'$, $\mathsf{z}_t' = (\mathsf{e}_{p+1}' \otimes \mathsf{w}_t')$, $\mathsf{e}_{p+1}' = (1, \mathbf{0}_p')$, the state vector has dimension $(p+1)(2M+1)$, $\boldsymbol{\alpha}_t = (\boldsymbol{\alpha}_{0t}', \boldsymbol{\alpha}_{1t}', \boldsymbol{\alpha}_{1t}^{*\prime}, \cdots, \boldsymbol{\alpha}_{Mt}', \boldsymbol{\alpha}_{Mt}^{*\prime})'$, with the j-th block being

$$\boldsymbol{\alpha}_{jt} = \begin{bmatrix} \tilde{a}_{jt} \\ \sum_{i=1}^{p-1} \phi_{1+i}(d_j)\tilde{a}_{j,t-i} + \sum_{i=1}^{p} \beta_i(d_j)\eta_{j,t-i} \\ \sum_{i=2}^{p-1} \phi_{1+i}(d_j)\tilde{a}_{j,t-i} + \sum_{i=1}^{p} \beta_{1+i}(d_j)\eta_{j,t-i+1} \\ \vdots \\ \beta_p(d_j)\eta_{j,t-1} \end{bmatrix}, \tag{9}$$

Methodology: Inference

while $\mathsf{T} = \text{diag}(\mathsf{T}_0, \mathsf{I}_2 \otimes \mathsf{T}_1, \cdots, \mathsf{I}_2 \otimes \mathsf{T}_M)$, $\mathsf{h} = (\mathsf{h}_0', \mathsf{i}_2' \otimes \mathsf{h}_1', \cdots, \mathsf{i}_2' \otimes \mathsf{h}_M')'$, with

$$\mathsf{T}_j = \begin{bmatrix} \phi_1(d_j) & 1 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ \phi_p(d_j) & 0 & \ddots & 1 \\ 0 & & \cdots & 0 \end{bmatrix}, \qquad \mathsf{h}_j = \begin{bmatrix} 1 \\ \beta_1(d_j) \\ \vdots \\ \beta_p(d_j) \end{bmatrix}, \tag{10}$$

and $\boldsymbol{\eta}_t = \left((\eta_{0,t}, \eta_{1,t}, \eta^*_{1,t}, \cdots, \eta_{M,t}, \eta^*_{M,t})' \otimes \mathsf{i}'_{p+1} \right)'$, with $\eta_{j,t} \sim$ i.i.d. $\mathsf{N}(0, \sigma^2_{\eta j})$. The initial state vector, $\boldsymbol{\alpha}_1$, has a Gaussian distribution centred around $\mathbf{0}$ with variance-covariance matrix $\mathsf{h}\text{Var}(\boldsymbol{\eta}_t)\mathsf{h}'$.

Methodology: Inference

- Maximum likelihood estimation of the model parameters is performed with the support of the **augmented Kalman filter**, see De Jong (1991) and Proietti and Luati (2013). The parameter vector $\boldsymbol{\delta}$ is concentrated outside of the likelihood and the profile diffuse likelihood is maximised with respect to the parameters in the vector $\boldsymbol{\theta} = (d_0, \sigma^2_{\eta 0}, d_1, \sigma^2_{\eta 1}, \ldots, d_M, \sigma^2_{\eta M}, \sigma^2_{\epsilon})'$.
- Missing values are handled by the augmented KF by skipping the updating operations. Point and interval estimates of the latent states, conditional on the observed time series and the parameter estimates, are obtained by the fixed interval smoothing algorithm by De Jong (1989).

Main analysis and results

- Let us present the main analysis and estimation results, when the structural model (3)-(4) is fitted to the ethane time series.
- One of the main issues concerning the specification is the selection of M, the number of harmonics that are used to model the annual seasonal pattern. A common finding is that the fundamental cycle and lower harmonics explain most of the variability of the annual cycle.
- The identification of the relevant frequencies can be aided by the Lomb-Scargle periodogram (cf. Lomb, 1976), which is used to detect periodicities in irregularly spaced and incompletely observed time series. The basic idea is to regress the time series on $(1, \cos(\omega t), \sin(\omega_t))$ and rescaling the coefficient of determination to obtain the periodogram at frequency $\omega \in (0, \pi)$.

Main analysis and results

For choosing M we adopted the following recursive procedure.

- We estimate the model featuring $M = j + 1$ fSW components for $j = 0, 1, 2, \cdots, 9$, adding the λ_{j+1} harmonic at each iteration, such that if step j provides a substantial increase of the maximized log likelihood (which occurs when the log-likelihood displacement is greater than 3), then we keep the relative harmonic and we proceed forward, otherwise we discard it from the analysis.
- The maximum number of iterations is equal to 9, since the Lomb-Scargle periodogram did not evidence any relevant harmonic beyond $j = 9$. Finally, we select the model for which the Akaike Information Criterion (AIC) is minimized.
- As far as the Markovian approximation of the long memory processes is concerned, the order p is set equal to 3, as suggested by Hartl and Jucknewitz (2022).

Results

Table 2: Quasi maximum likelihood estimates of the parameters of the model at the Alesund, Altzomoni, Arrival Heights, Bremen, Eureka, Izana, Kiruna, La Reunion, Rikubetsu, Saint Petersburg, Thule, Toronto, Zugspitze, Wollongong and Lauder ground stations.

Alesund

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.9353	0.1019
1	0.0172	0.2608	1.1733
2	0.0344	0.2614	1.5935
3	0.0516	0.3148	1.1136

Altzomoni

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	1.0000	0.0310
1	0.0172	0.5124	0.0677
3	0.0516	0.4425	0.8118

Arrival Heights

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.9639	0.0080
1	0.0172	0.4678	0.1209
2	0.0344	0.2182	0.1427
3	0.0516	0.1602	0.1956

Bremen

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.9832	0.0288
1	0.0172	0.3949	0.5304
4	0.0688	0.3823	0.8825
5	0.0860	0.2453	0.0113
6	0.1032	0.2456	0.0030
10	0.1720	0.3045	1.4754

Eureka

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.9999	0.0433
1	0.0172	0.3596	0.9599
3	0.0516	0.3488	0.9966
4	0.0688	0.3103	0.7267
5	0.0860	0.2774	0.2357

Izana

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.7738	0.0784
1	0.0172	0.3047	0.9928
2	0.0344	0.3525	0.9307

Kiruna

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.8531	0.1253
1	0.0172	0.3202	1.5664
2	0.0344	0.1673	0.3029
3	0.0516	0.2480	0.6003

La Reunion

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.7978	0.0953
1	0.0172	0.3328	0.4462
2	0.0344	0.3105	0.3065
3	0.0516	0.2874	0.2691

Rikubetsu

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.9471	0.0908
1	0.0172	0.1595	0.3948
2	0.0344	0.3597	0.2951
4	0.0688	0.2267	0.3243
8	0.1376	0.2265	1.8837

St. Petersburg

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.8800	0.1423
1	0.0172	0.2780	1.5563
2	0.0344	0.2403	0.8876
3	0.0516	0.2080	0.6917

Thule

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.9900	0.0299
1	0.0172	0.3723	1.5997
2	0.0344	0.0531	0.0000

Toronto

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.7032	0.2736
1	0.0172	0.1172	3.5930
2	0.0344	0.2136	1.6337

Wollongong

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.9447	0.0420
1	0.0172	0.3139	0.0416
2	0.0344	0.6168	0.6848
4	0.0688	0.0400	0.2669

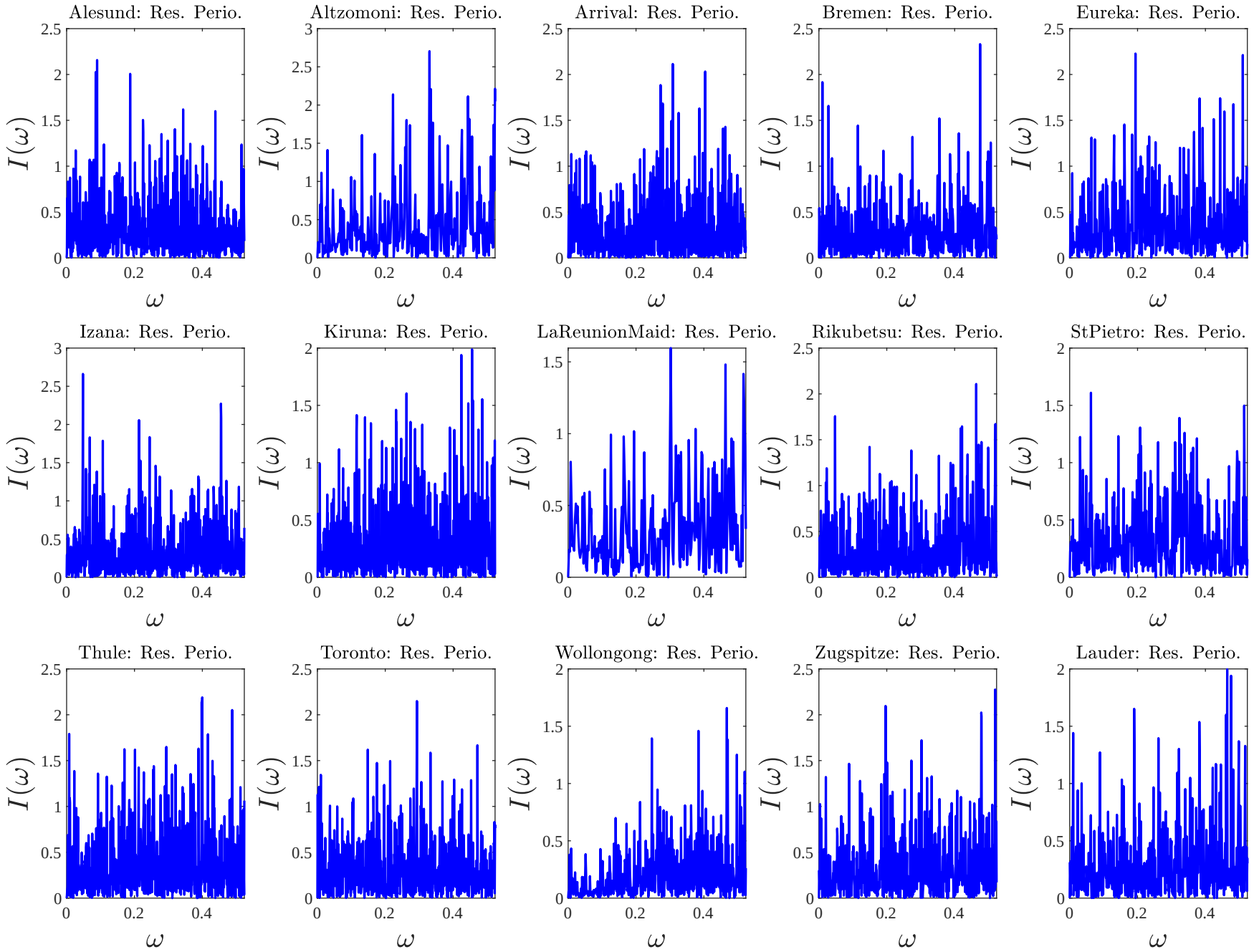
Zugspitze

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.8104	0.0829
1	0.0172	0.2182	0.7972
2	0.0344	0.2234	1.1731

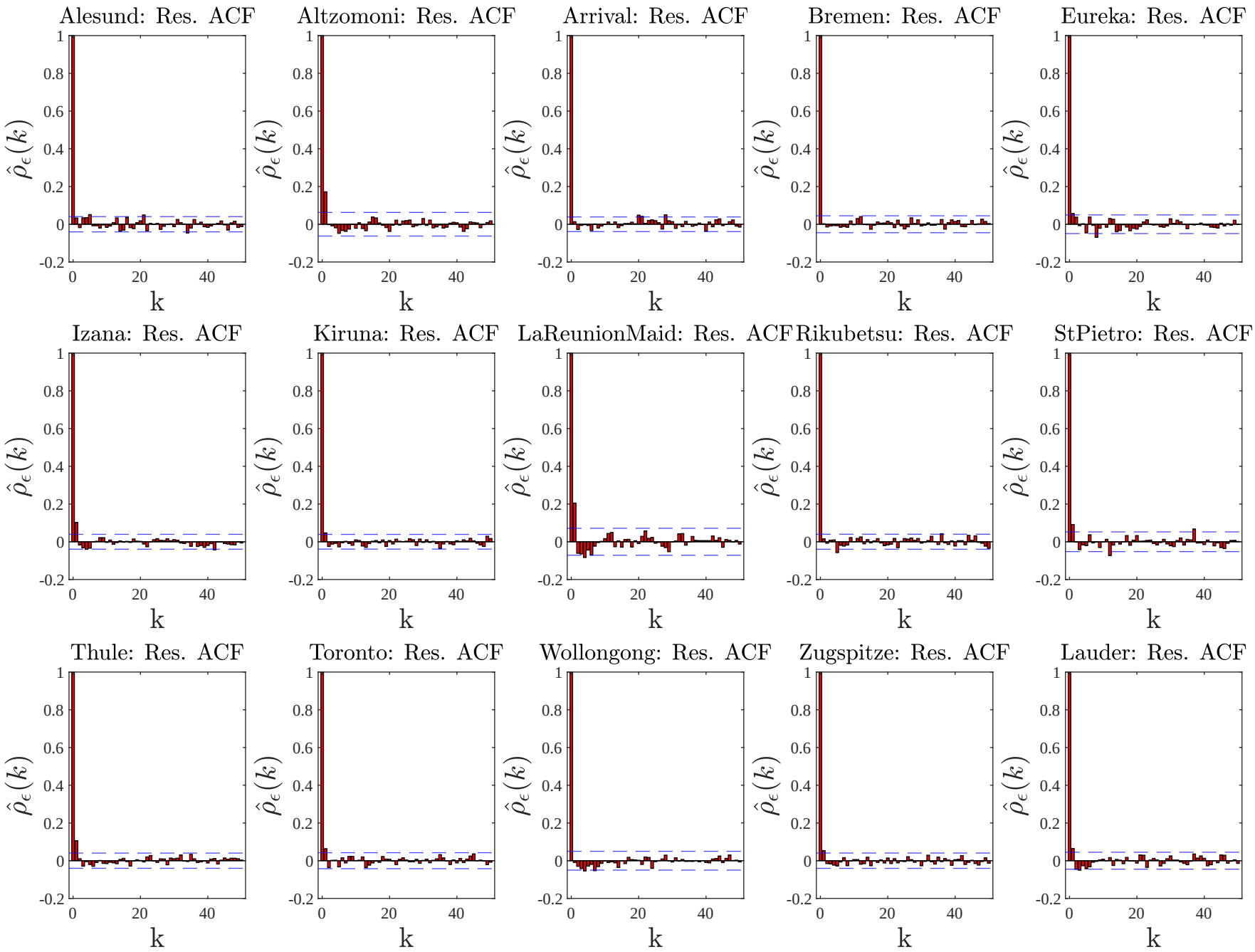
Lauder

j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\eta j}$
0	0.0000	0.7360	0.0742
1	0.0172	0.4288	0.3735
2	0.0344	0.2614	0.1537
3	0.0516	0.2434	0.2600
4	0.0688	0.2515	0.2559
5	0.0860	0.2406	0.3081

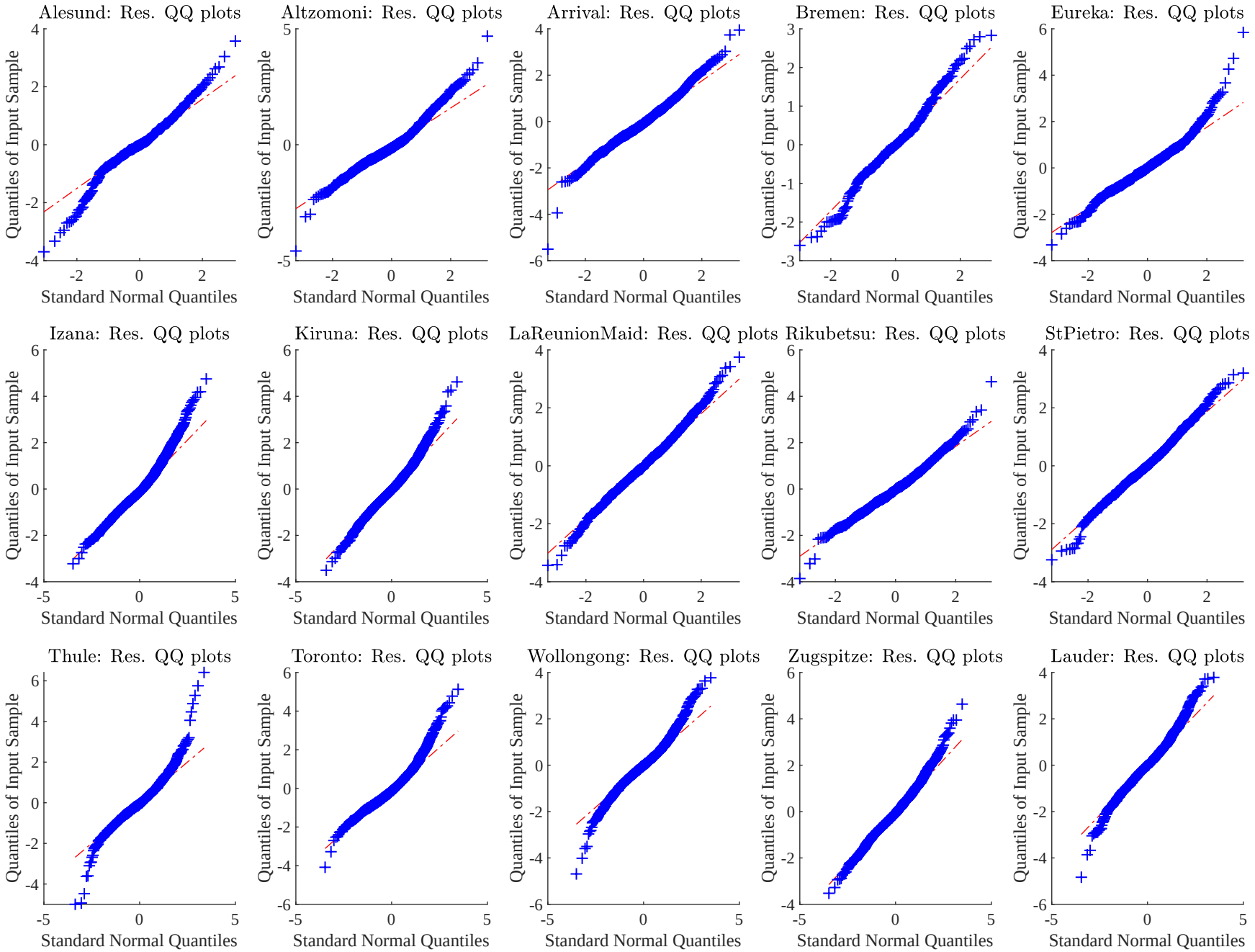
Residuals Lomb-Scargle periodograms



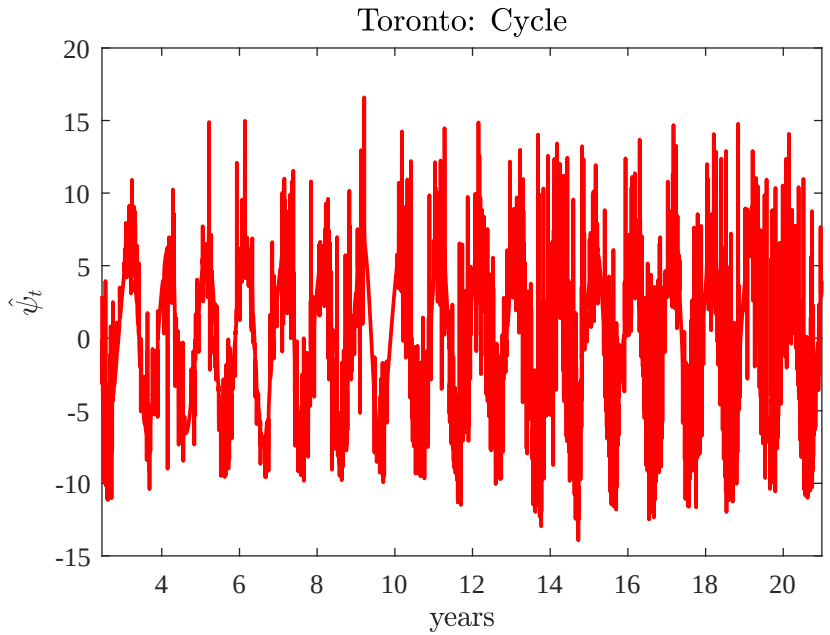
Residuals sample Autocorrelation (inverted Lomb-Scargle periodograms)

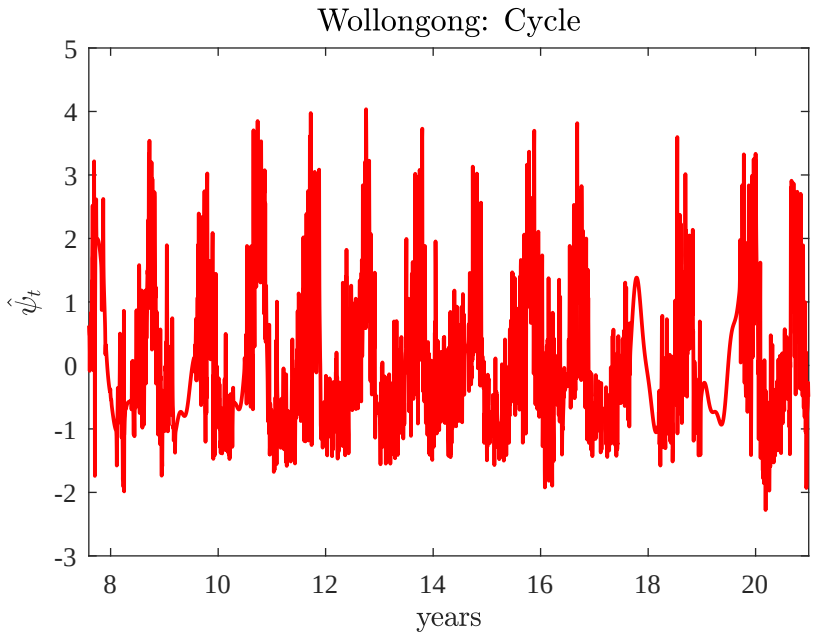


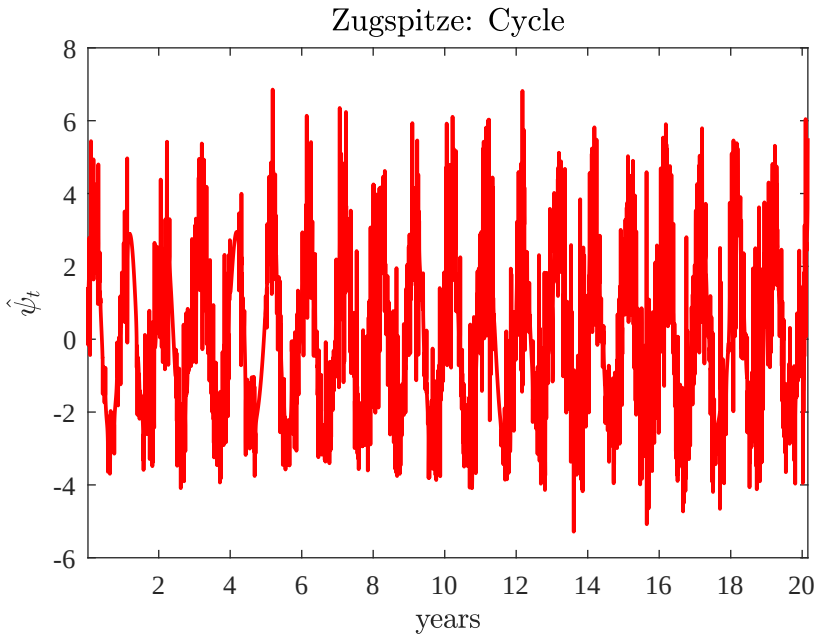
Residuals QQ-plots

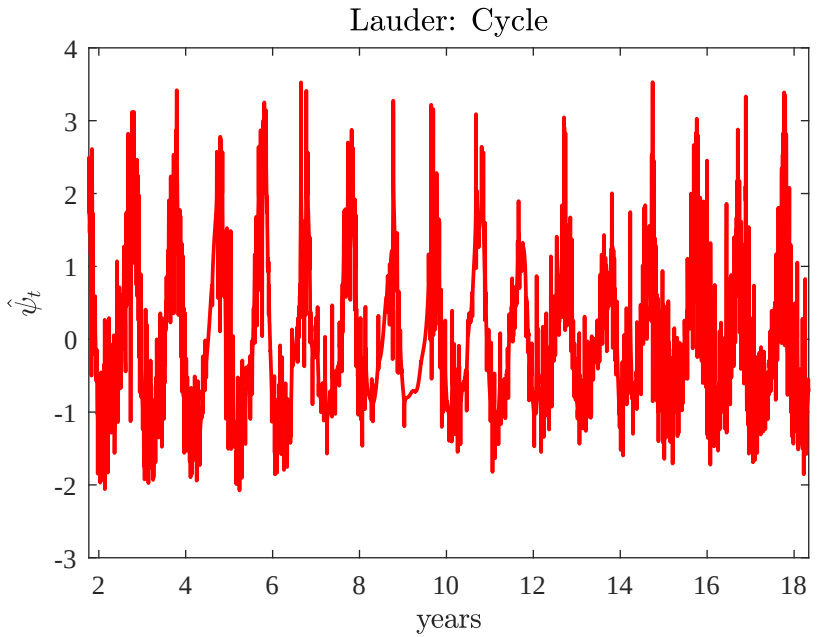


Cyclical component estimates: the annual cycle is ROUGH!!

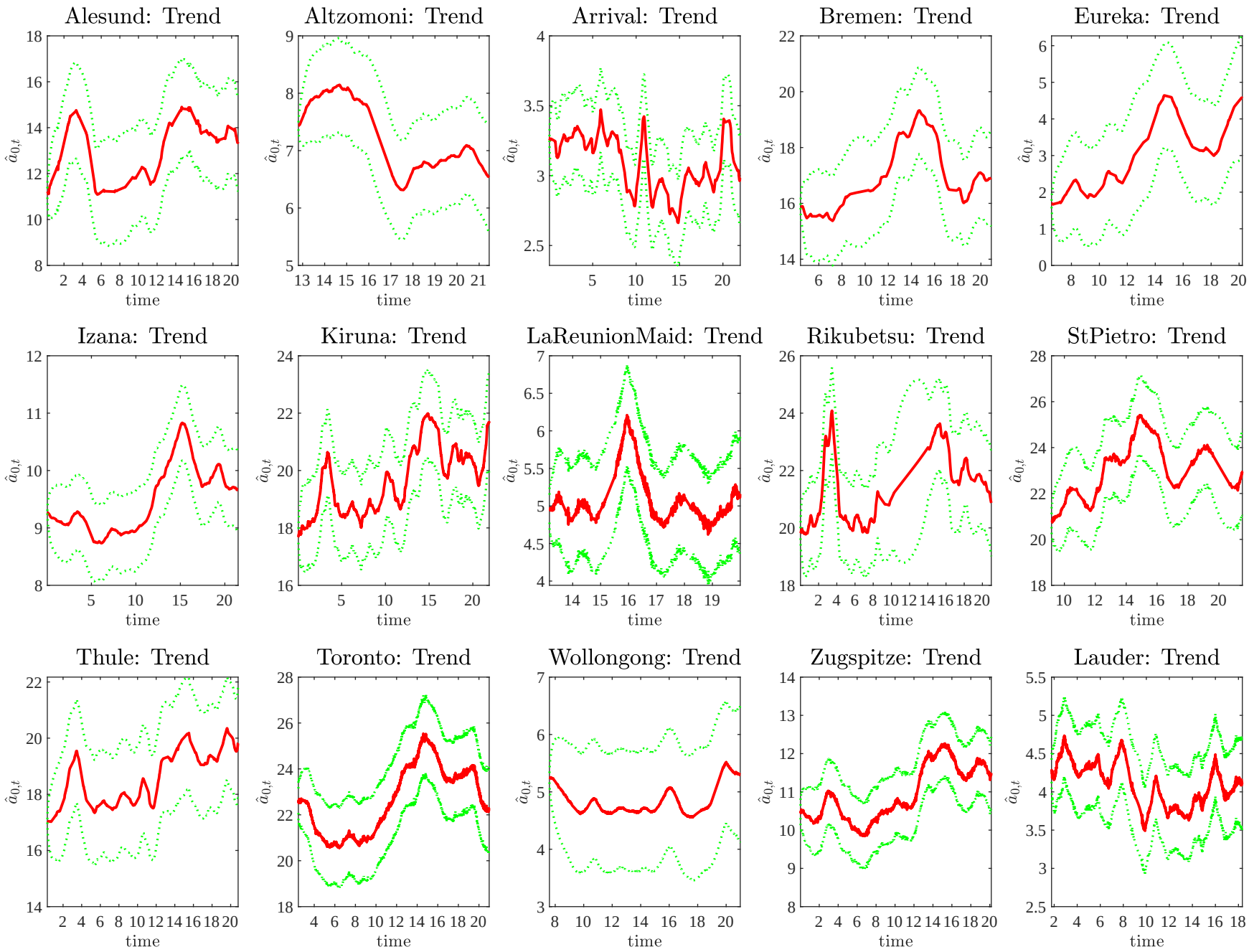








Trend components estimates

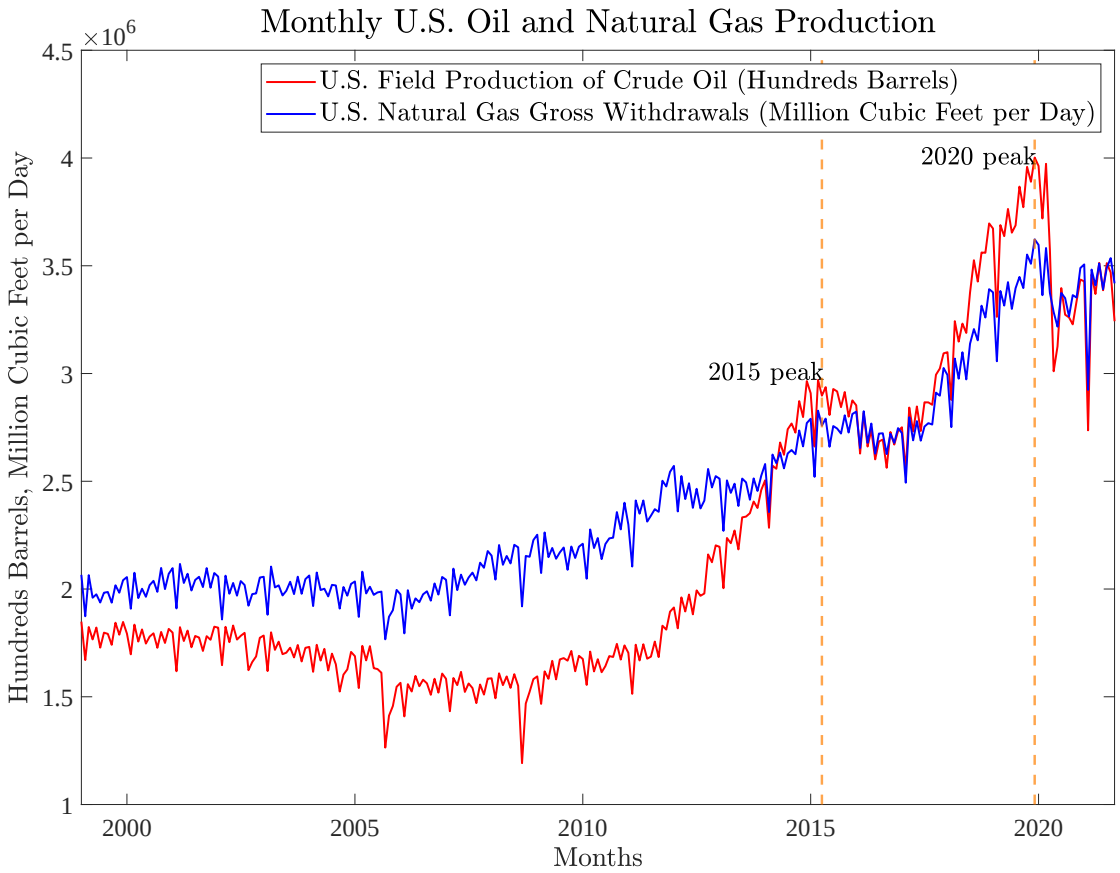


Summary

1. The inference on the parameters suggests that the ethane series may be driven by stochastic cycles and a non-stationary fractional component at the long run frequency.
2. The variability of the cyclical components is mostly depended by the first harmonic.
3. The Alesund, Kiruna, Thule and Zugspitze trends shows a peak around 2003-2004, which was also reported by Friedrich et al. (2020) for the Jungfraujoch (Switzerland) series and attributed to the boreal forest fires in Russia during such period.
4. The Northern Hemisphere ethane trends show a growth in the period 2009-2014, with a successive hiatus in 2014-2018, confirming previous studies in the literature.
5. In particular, the common hiatus around 2015 in the Northern Hemisphere may be attributed to the oil and natural gas production in US.

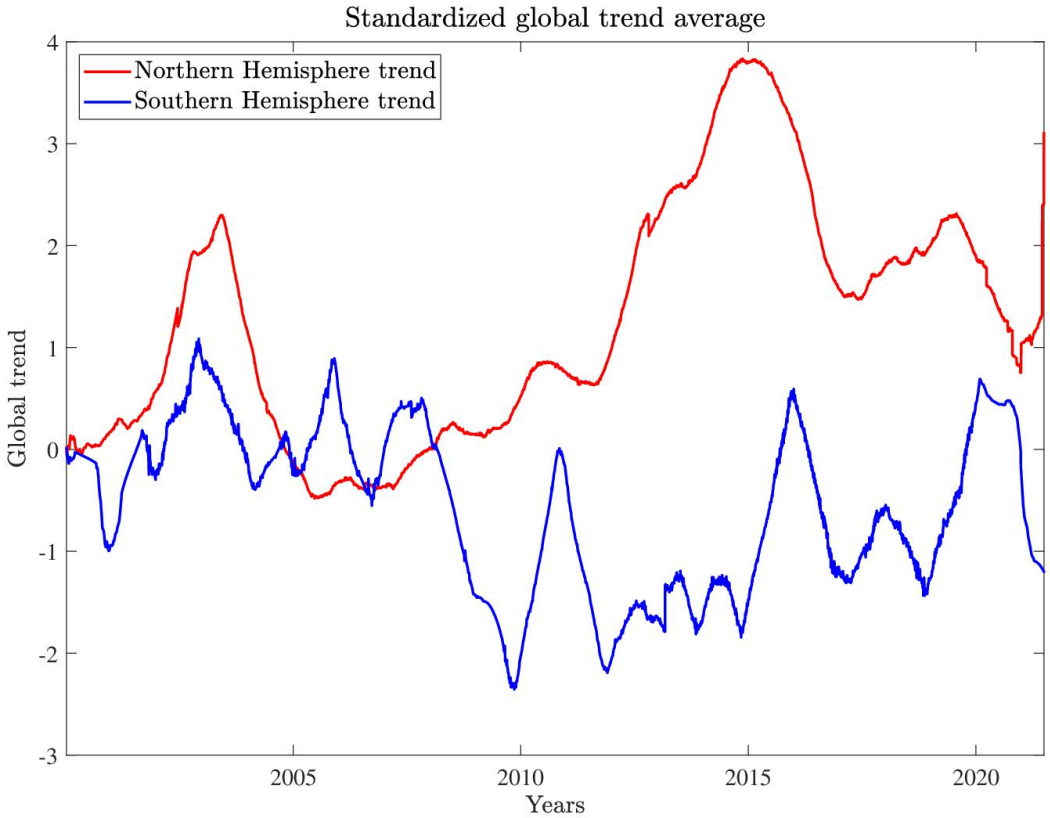
Summary

Figure: Monthly United States natural gas production gross withdrawals (million cubic feet per day, blue line) and field production of crude oil (hundred barrels, red line). Source: U.S. Energy Information Administration, https://www.eia.gov.



Summary

Figure: Standardized global **ethane trends average** in the Northern (Alesund, Altzomoni, Bremen, Eureka, Izana, Kiruna, Rikubetzu, St. Petersburg, Thule, Toronto and Zugspitze ground-stations) and Southern (Arrival Heights, Lauder, La Reunion and Wollongong ground-stations) Hemispheres.



Summary

1. The common pattern in the **Northern Hemisphere** trends supports the thesis that emissions associated with hydraulic fracturing and shale gas operations in North America are affecting Europe (supporting the hypothesis of Franco et al., 2015).
2. The **Southern Hemisphere** ethane trends require also attention.
3. For the Lauder and Wollongong records, they are mostly likely driven by biomass burning emissions, as the Australian bushfire season.
4. The huge peak in 2020 in the Wollongong trend can be associated to the last devastating bushfire season, while the peak around 2006-2007 in the Lauder trend estimate is probably attributable to the Great Divide bushfires.

Conclusions

1. We provide a suitable tool to investigate on the trend dynamics of the ethane time series.
2. The linearity assumption is preserved in our model allowing for an easier implementation of the analysis with respect to other approach available in the literature (as the non-linear methodology proposed by Friedrich et al., 2020), without any drawback in terms of accuracy of the results.
3. Our analysis suggests that the ethane series may be driven by a stationary and stochastic annual cycle. This result is something of new in the literature referred in the paper, where the annual cycle was model as deterministic since now.
4. Finally, we found a common pattern in the Northern Hemisphere ethane series, which seem mostly driven by the crude oil production and natural gas gross withdrawals in US, supporting the hypothesis by Franco et al. (2015) for which the exploitation of crude oil and natural gas in US is affecting ethane trends in Europe.

Merci Beaucoup pour Votre Attention!