EcoDep Seminar
Time trends in atmospheric ethane

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

February 26, 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 exact human contribution to the abundance of methane.

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

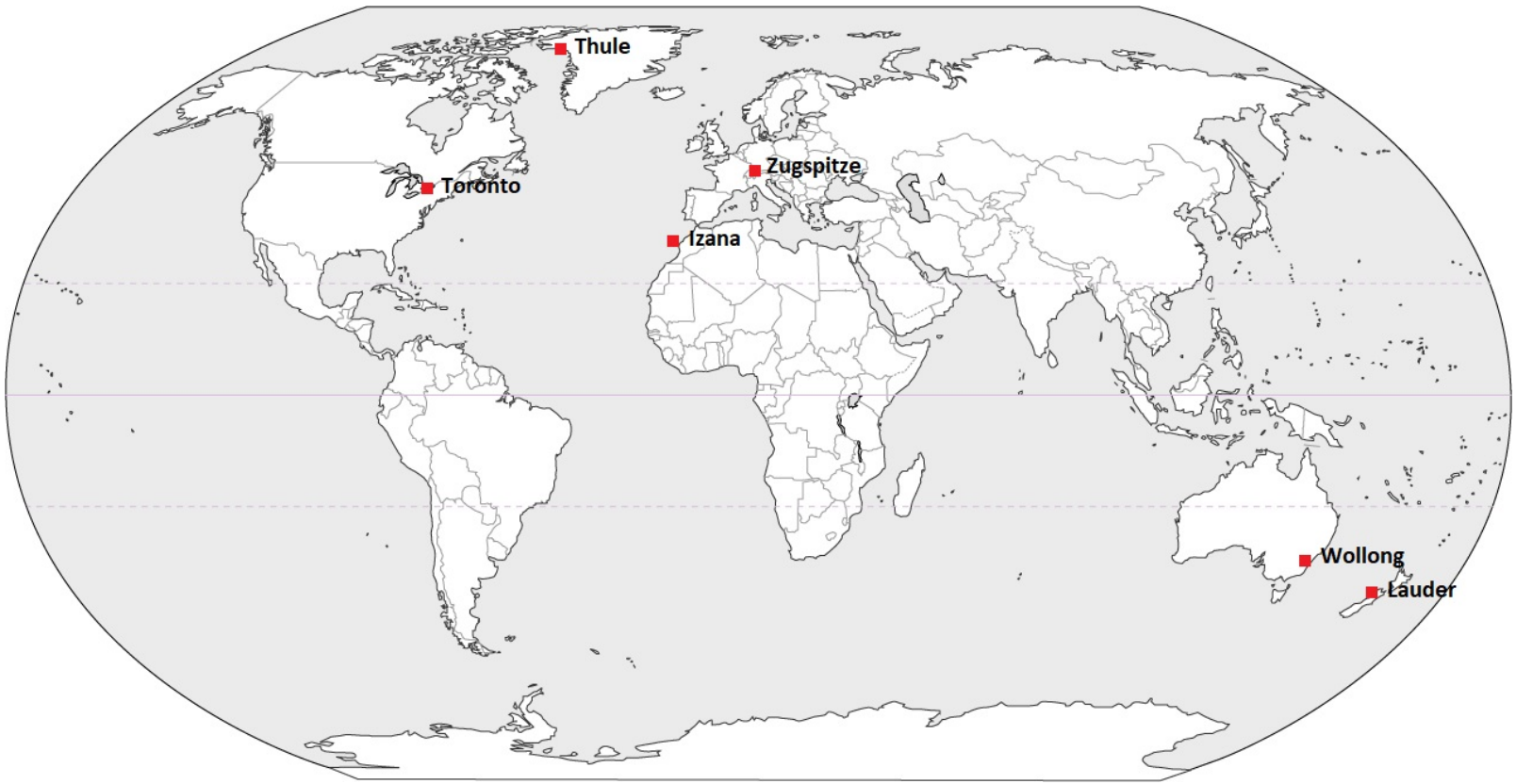
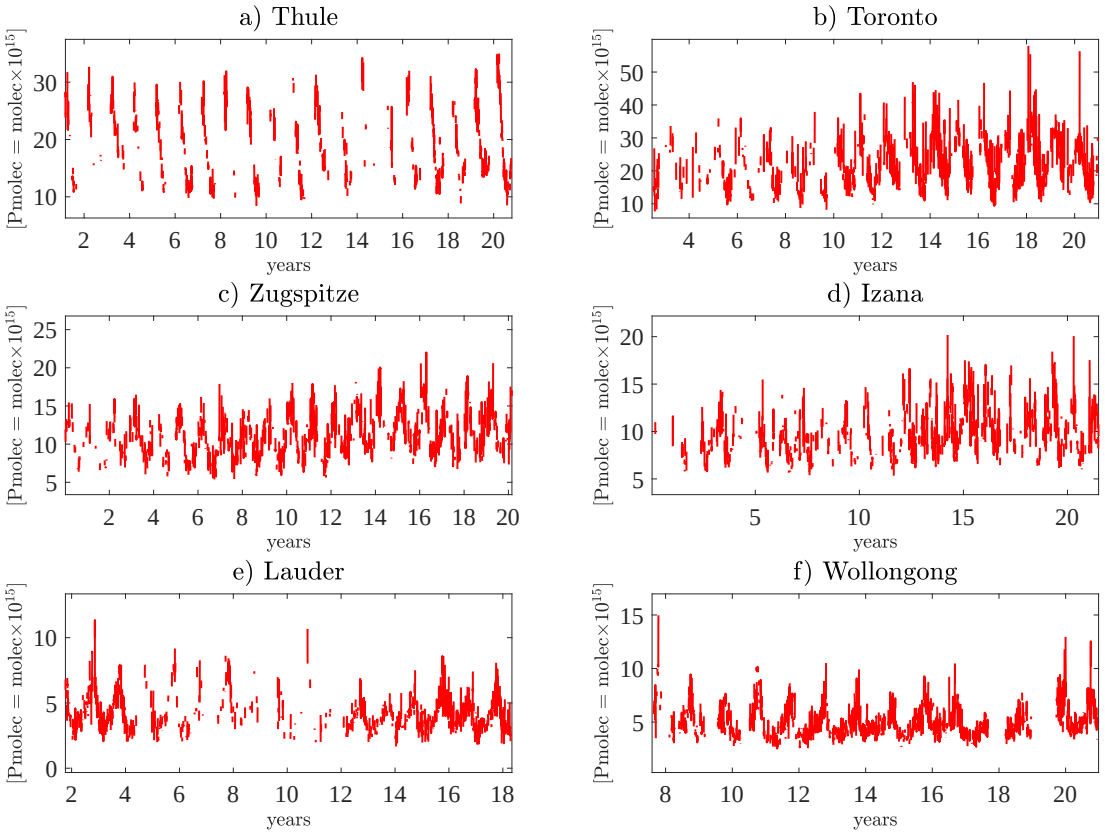


Figure: We investigate on the ground-based Fourier Transform Infrared (FTIR) solar spectra measurements of ethane abundance in the atmosphere recorded at six ground-stations.

The Data

Figure: FTIR solar spectra measurements of atmospheric ethane (C2H6) recorded at the Thule, Toronto, Zugspitze, Izana, Lauder and Wollongong ground-stations.



The Data

The six 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**, about 70% in average of missing points for each series.

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.
- The 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.

Literature Review

- In our opinion, 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 suggest to model the cycle as stochastic, imposing the fundamental frequency and the harmonics fixed at the annual frequency, while the parameters regulating the phase and amplitude evolve stochastically according to two independent FN processes.
- 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.
- Finally, the trend component is modelled by a standard random walk.

Methodology

At the end, we consider the following structural model:

$$
\begin{aligned}
x_t &= \mu_t + \psi_t \\
\mu_t &= \mu_{t-1} + \zeta_t && \zeta_t \sim \text{i.i.d.} N(0, \sigma^2_\zeta) \\
\psi_t &= \textstyle\sum_{j=1}^{M} \psi_{j,t} && M \in \mathbb{N} \\
\psi_{j,t} &= \cos(\lambda j t) a_{jt} + \sin(\lambda j t) a^*_{jt} && \lambda \in (0, \pi) \\
(1-L)^{d_j} a_{jt} &= \eta_{jt} && \eta_{jt} \sim \text{i.i.d.} N(0, \sigma^2_{j\eta}) \\
(1-L)^{d_j} a^*_{jt} &= \eta^*_{jt} && \eta^*_{jt} \sim \text{i.i.d.} N(0, \sigma^2_{j\eta})
\end{aligned}
\tag{1}
$$

where all the sources of disturbance, ζ_t, η_{jt} and η^*_{jt} for $j = 1, 2, \cdots, M$, are assumed to be mutually independent.

Methodology

- The process $\psi_{j,t}$ is defined as the fractional Sinusoidal Waveform (fSW) process (cf. Proietti and Maddanu, 2022). The fSW process defines a sinusoidal wave, whose time-varying parameters, a_{jt} and a^*_{jt}, regulate the phase ($\varphi_{jt} = \arctan(a^2_{jt}/a^{*2}_{jt})$) and amplitude ($\mathrm{A}_{jt} = \sqrt{a^2_{jt} + a^{*2}_{jt}}$) evolve stochastically as two independent FN processes.
- If $d_j \in (0, \frac{1}{2})$ the process is stationary and displays long memory with spectral density

$$f_{\psi,j}(\omega) = \frac{\sigma^2_{j\eta}}{4\pi}\left\{\left|2\sin(\frac{\omega-\lambda j}{2})\right|^{-2d_j} + \left|2\sin(\frac{\omega+\lambda j}{2})\right|^{-2d_j}\right\} \quad (2)$$

unbounded at λ_j and ACF

$$\gamma_{\psi,j}(k) = \sigma^2_{j\eta}\frac{\Gamma(1-2d_j)\Gamma(d_j+k)}{\Gamma(1+k-d_j)\Gamma(d_j)\Gamma(1-d_j)}\cos(\lambda j k), \quad k \in \mathbb{Z}, \quad (3)$$

Methodology

- The fSW process encompasses deterministic cycles as a limiting case when both $\sigma^2_{j\eta}$ and d_j approach to zero.
- Indeed, considering the harmonic process $z_t = a_j \cos(\lambda j t) + a^*_j \sin(\lambda j t)$ with $a_j, a^*_j \sim$ i.i.d.$N(0, \sigma^2_{ja})$ and ACF given by $\gamma_z(k) = \sigma^2_{ja} \cos(\lambda j k)$. Rewriting $\sigma^2_{j\eta} = \sigma^2_{ja}\pi(1-2d_j)$, if $d_j \to \frac{1}{2}$ from the left, then $\gamma_{\psi,j}(k) \to \sigma^2_{ja} \cos(\lambda j k)$ because of $\lim_{u\to 0^+} u\Gamma(u) = 1$ with $u = 1 - 2d_j$.
- It follows that the fSW process encompasses a deterministic cycle as a limiting case. By rewriting $f_{\psi,j}(\omega)$ in terms of $\sigma^2_{\psi j} = \sigma^2_{\eta j} \frac{\Gamma(1-2d_j)}{\Gamma^2(1-d_j)}$, we find that $\lim_{d_j \to \frac{1}{2}^-} f_{\psi j}(\omega) = \frac{\sigma^2_{\psi j}}{2}[\delta(\omega - \lambda j) + \delta(\omega + \lambda j)]$ where $\delta(\cdot)$ is Dirac's Delta function. As a consequence, we obtain a discrete spectrum, since $f_{\psi j}(\omega)$ degenerates to a point with mass $\sigma^2_{\psi j}/2$ at $\omega = \pm\lambda j$ and is 0 elsewhere, implying that the cyclical component ψ_t, with spectral density $f_\psi(\omega) = \sum_{j=1}^{M} f_{\psi j}(\omega)$, may present mixed spectra.

Methodology

- Notice that, due to additivity and independence, the ACF of ψ_t can be derived as

$$\gamma_\psi(k) = \sum_{j=1}^{M} \sigma^2_{\psi j} \frac{\Gamma(1-d_j)\Gamma(d_j+k)}{\Gamma(1+k-d_j)\Gamma(d_j)} \cos(\lambda j k), \quad k \in \mathbb{Z}, \tag{4}$$

where $\sigma^2_{\psi j} = \sigma^2_{\eta j} \frac{\Gamma(1-2d_j)}{\Gamma^2(1-d_j)}$ such that $\gamma_\psi(0) = \sum_{j=1}^{M} \sigma^2_{\psi j}$ allows for the decomposition of the total variability of the cyclical component in the single contributions of each ones of its harmonics.

Methodology

The derivation of an approximated state space representation of the structural model in (1) follows immediately as

$$\begin{aligned} x_t &= Z\alpha_{t|t-1} + G\epsilon_t \qquad (\textit{measurement equation}) \\ \alpha_{t+1|t} &= T\alpha_{t|t-1} + H\epsilon_t \qquad (\textit{state equation}) \end{aligned} \tag{5}$$

with

$$Z = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \end{bmatrix} \quad , \quad G = \begin{bmatrix} \sigma_\zeta & \tilde{\sigma} \end{bmatrix} \quad , \quad \epsilon_t = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \sim N(0, I_2)$$

$$\alpha_{t|t-1} = \begin{bmatrix} \mu_{t|t-1} \\ \psi_{t|t-1} \\ \psi_{t+1|t-1} \\ \vdots \\ \psi_{t+p-1|t-1} \end{bmatrix} \quad T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \tilde{\phi}_{p1} & 1 & & 0 \\ 0 & \tilde{\phi}_{p2} & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & & 1 \\ 0 & \tilde{\phi}_{pp} & 0 & & 0 \end{bmatrix} \quad H = \begin{bmatrix} \sigma_\zeta & 0 \\ 0 & \tilde{\sigma}\tilde{\phi}_{p1} \\ 0 & \tilde{\sigma}\tilde{\phi}_{p2} \\ \vdots & \vdots \\ 0 & \tilde{\sigma}\tilde{\phi}_{pp} \end{bmatrix}$$

Methodology

- The coefficients $\tilde{\phi}_{pj}$ and $\tilde{\sigma}$ comes from the AR(p) representation of the ψ_t process. In fact, given the ACF $\gamma_\psi(k)$ and provided $\frac{1}{n}\sum_{t=1}^{n}\psi_t^2 > 0$, it follows from the Yule-Walker equation (cf. Brockwell and Davis, 1986) that the autoregressive process of order $p < n$ can be fitted for $(\psi_1, \psi_2, \cdots, \psi_n)'$ according to

$$\psi_t = \sum_{j=1}^{p} \tilde{\phi}_{pj}\psi_{t-j} + \varepsilon_t \quad , \quad \varepsilon_t \sim N(0, \tilde{\sigma}^2) \tag{6}$$

where the coefficients $\tilde{\phi}_{pj}$ for $j = 1, 2, \cdots, p$ and $\tilde{\sigma}$ are computed recursively via the Durbin-Levinson (DL) algorithm (see proposition 8.2.1 in Brockwell and Davis, 1986).

Methodology

Estimation and prediction of x_t in (1) can be carried out via the **Kalman filter** (KF), which computes recursively the innovations $v_t = x_t - x_{t|t-1}$, the prediction error variance and the one-step-ahead predictor of the states, $\alpha_{t|t-1}$ and their mean square error matrix, according to

$$v_t = x_t - \mathsf{Z}\hat{\alpha}_{t|t-1} \;,\; f_t = \mathsf{Z}\mathsf{P}_{t|t-1}\mathsf{Z}' + \mathsf{G}\mathsf{G}' \;,\; \mathsf{K}_t = (\mathsf{T}\mathsf{P}_{t|t-1}\mathsf{Z}' + \mathsf{H}\mathsf{G}')/f_t \quad (7)$$
$$\hat{\alpha}_{t+1|t} = \mathsf{T}\hat{\alpha}_{t|t-1} + \mathsf{K}_t v_t \;\;,\;\; \mathsf{P}_{t+1|t} = \mathsf{T}\mathsf{P}_{t|t-1}\mathsf{T}' + \mathsf{H}\mathsf{H}' - f_t \mathsf{K}_t \mathsf{K}_t'$$

for $t = 1, 2, \cdots, n$, where $\hat{\alpha}_{t|t-1} = E(\alpha_t|\mathcal{F}_{t-1})$, $\mathsf{P}_{t|t-1} = \mathrm{Var}(\alpha_t|\mathcal{F}_{t-1})$, $f_t = \mathrm{Var}(x_t|\mathcal{F}_{t-1})$ and K_t is knowns as the *Kalman gain*. The algorithm is initialized via

$$\hat{\alpha}_{1|0} = \begin{bmatrix} \mu_0 \\ 0 \end{bmatrix} \quad , \quad \mathsf{P}_{1|0} = \begin{bmatrix} \sigma_\zeta^2 & 0' \\ 0 & \mathbf{\Gamma}_{\psi,n} \end{bmatrix} \quad (8)$$

where 0 is a $p \times 1$ vector of zeros, $\mathbf{\Gamma}_{\psi,n}$ has ij-elements equal to $\gamma_\psi(|i-j|)$ and μ_0 comes from $\mu_t = \mu_0 + \sum_{j=0}^{t-1} \zeta_{t-j}$.

Methodology

- The KF recursion returns the pseudo innovation v_t, such that if the AR(p) representation was the true model, $v_t|\mathcal{F}_{t-1} \sim N(0, f_t)$, so that the quasi log likelihood follows as

$$L_n(\theta) = -\frac{1}{2}\left(n(\ln 2\pi) + \sum_{t=1}^{n} \ln f_t + \sum_{t=1}^{n} \frac{v_t^2}{f_t} \right) \tag{9}$$

where $\theta \in \mathbb{R}^m$ is the true parameter. The resulting Quasi Maximum Likelihood Estimator (QMLE) $\hat{\theta}_n = \text{argmax}_{\theta \in \mathbb{R}^m} L_n(\theta)$ has been proved to be consistent and asymptotically normal distributed by Chan and Palma (1998) in the case of standard long memory processes (under stationarity).

Methodology

- *Extraction of the trend and cycle components*
- The **updates (real-time) estimates** of the state vector, defined as $E(\alpha_t|\mathcal{F}_t) = \hat{\alpha}_{t|t}$, and their covariance matrix $\text{Var}(\alpha_t|\mathcal{F}_t) = \text{P}_{t|t}$ can be computed via

$$\hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + \text{P}_{t|t-1}\text{Z}'\frac{\nu_t}{f_t}\,,\ \text{P}_{t|t} = \text{P}_{t|t-1} - \text{P}_{t|t-1}\text{Z}'f_t^{-1}\text{Z}\text{P}_{t|t-1} \quad (10)$$

- The **smoothed estimates** of the state vector, defined as $E(\alpha_t|\mathcal{F}_n) = \hat{\alpha}_{t|n}$, and their covariance matrix $\text{Var}(\alpha_t|\mathcal{F}_n) = \text{P}_{t|n}$ are obtained by the following backwards recursive formulae, starting at $t = n$, with initial values $r_n = 0$ and $\text{N}_n = 0$, for $t = n-1, \cdots, 1$

$$\begin{array}{rclcrcl} r_{t-1} & = & \text{L}_t'r_t + \text{Z}'f^{-1}v_t & & \text{N}_{t-1} & = & \text{L}_t'\text{N}_t\text{L}_t + \text{Z}'f^{-1}\text{Z} \\ \hat{\alpha}_{t|n} & = & \hat{\alpha}_{t|t-1} + \text{P}_{t|t-1}r_{t-1} & & \text{P}_{t|n} & = & \text{P}_{t|t-1} - \text{P}_{t|t-1}\text{N}_{t-1}\text{P}_{t|t-1}. \end{array} \quad (11)$$

where $\text{L}_t = \text{T} - \text{K}_t\text{Z}'$.

Handling Missing Data

- We deal with the missing data problem by defining the partially observed process

$$y_t = x_t C_t$$

where x_t is the process of interest, independent from C_t, which is not completely observed, and C_t is a binary process

$$C_t = \begin{cases} 1, & \text{if } x_t \text{ is observed} \\ 0, & \text{if } x_t \text{ is missing} \end{cases}$$

- The parameters estimation and prediction of x_t in the presence of missing data is particularly easy via the KF methodology. Indeed, missing values are handled just by skipping the updating operations at each t iteration in the KF for which $C_t = 0$, such that we do not have *Kalman gain* and we only compute

$$\hat{\alpha}_{t+1|t} = \mathrm{T}\hat{\alpha}_{t|t-1} \quad , \quad \mathrm{P}_{t+1|t} = \mathrm{T}\mathrm{P}_{t|t-1}\mathrm{T}' + \mathrm{H}\mathrm{H}' \tag{12}$$

Handling Missing Data

- Similarly, the updates estimates reduce to

$$\hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} \quad , \quad \mathsf{P}_{t|t} = \mathsf{P}_{t|t-1} \tag{13}$$

- while the smoothed estimates become

$$\begin{aligned} r_{t-1} &= \mathsf{T}' r_t & \mathsf{N}_{t-1} &= \mathsf{T}' \mathsf{N}_t \mathsf{T} \\ \hat{\alpha}_{t|n} &= \hat{\alpha}_{t|t-1} + \mathsf{P}_{t|t-1} r_{t-1} & \mathsf{P}_{t|n} &= \mathsf{P}_{t|t-1} - \mathsf{P}_{t|t-1} \mathsf{N}_{t-1} \mathsf{P}_{t|t-1}. \end{aligned} \tag{14}$$

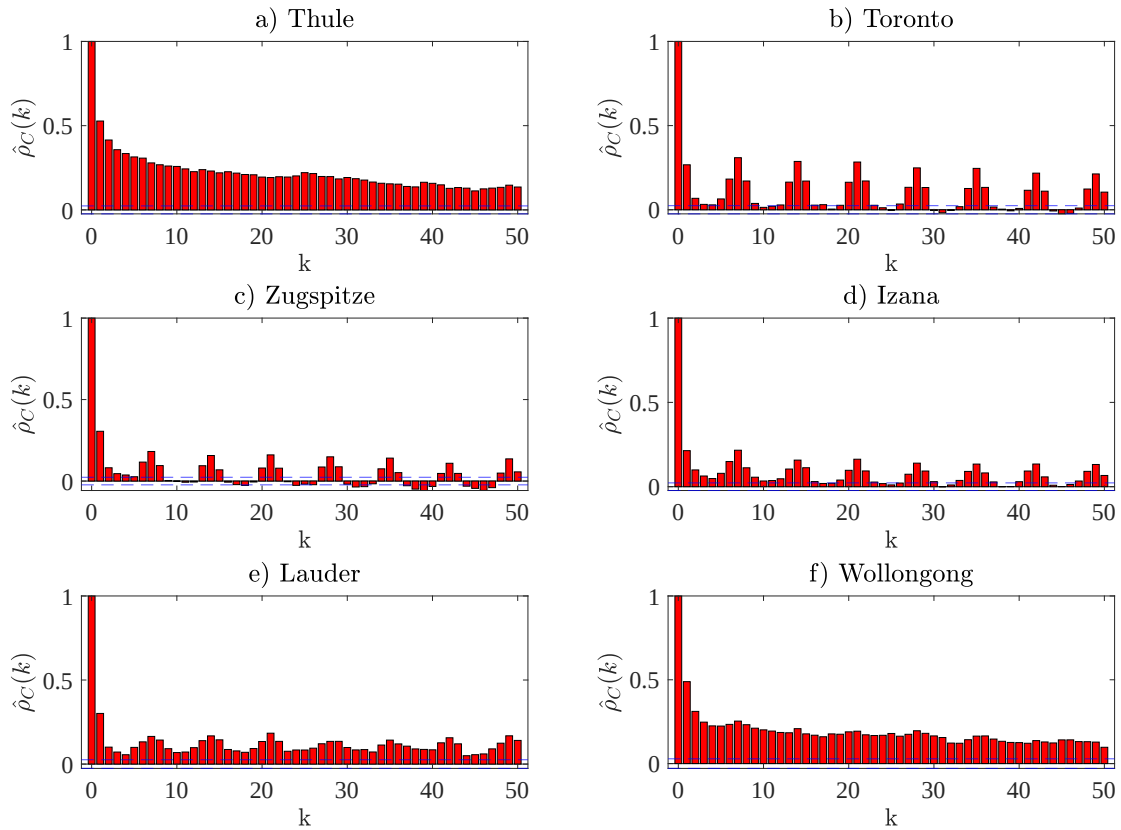
  respectively, in the case the t-th observation is missing.

Handling Missing Data

- At the moment, no assumption has been made concerning the missing data generating process C_t.
- Investigating on the dependence structure of C_t may be of some interest, at least for simulation purposes. Since C_t is directly observable in real world time series, we know that missing observations do not appear randomly in time, but often follow some serial correlation pattern. This because the observational process depends by the interaction of several variables, which can be both completely random or not.

Handling Missing Data

Figure: **Empirical Example:** Sample autocorrelation function of the missing data generating process C_t for the Thule, Toronto, Zugspitze, Izana, Lauder and Wollongong ethane series.



Handling Missing Data

- Notice that a sort of seasonality at the weekly frequency characterises the binary process C_t in some cases while in the others the observational process seem mostly driven by a geometrically decrease of the sample autocorrelation.
- To model such serial correlation patterns, we propose the following logistic process for the conditional probability of C_t:

$$\begin{aligned} \mathrm{P}(C_t = 1 | C_{t-1}) &= \left(1 + \exp(-\kappa_t)\right)^{-1} \\ \kappa_t &= \delta + \beta \kappa_{t-s} + \varrho \xi_{t-1} \\ \xi_t &= C_t - \mathrm{P}(C_t | C_{t-1}) \end{aligned} \tag{15}$$

such that the κ_t component follows a standard seasonal AR(1) process, where ξ_t is a martingale difference sequence with $E(\kappa_t) = \delta/(1-\beta)$ and the process displays seasonality at the frequency $2\pi/s$.

Simulation Exercise

- Let us implement a simulation exercise by generating the partially observed process $y_t = x_t C_t$, where C_t is generated according to (15).
- Assume $\lambda = \pi/60$, $M = 2$, $\mu_0 = 3.4$, $(d_1, d_2) = (0.45, 0.33)$, $(\sigma_{1,\eta}, \sigma_{2,\eta}) = (0.83, 0.55)$ and $\sigma_\zeta = 0.08$.
- Notice that we have set the variance of the trend disturbance as the lowest, such that the component will result almost unobservable from the time series plot. The detection of the hidden trend is the main purpose of this exercise.
- Finally, we consider two possible scenarios for generating C_t according to the logistic seasonal AR(1) model. In **Scenario I**, we set $s = 1$ (C_t follows a logistic AR(1)). In **Scenario II**, $s = 7$ (C_t exhibits a seasonality at frequency $2\pi/7$). The true values of the other parameters are $\delta = 0$, $\beta = 0.95$, $\varrho = 1.2$.

Simulation Exercise

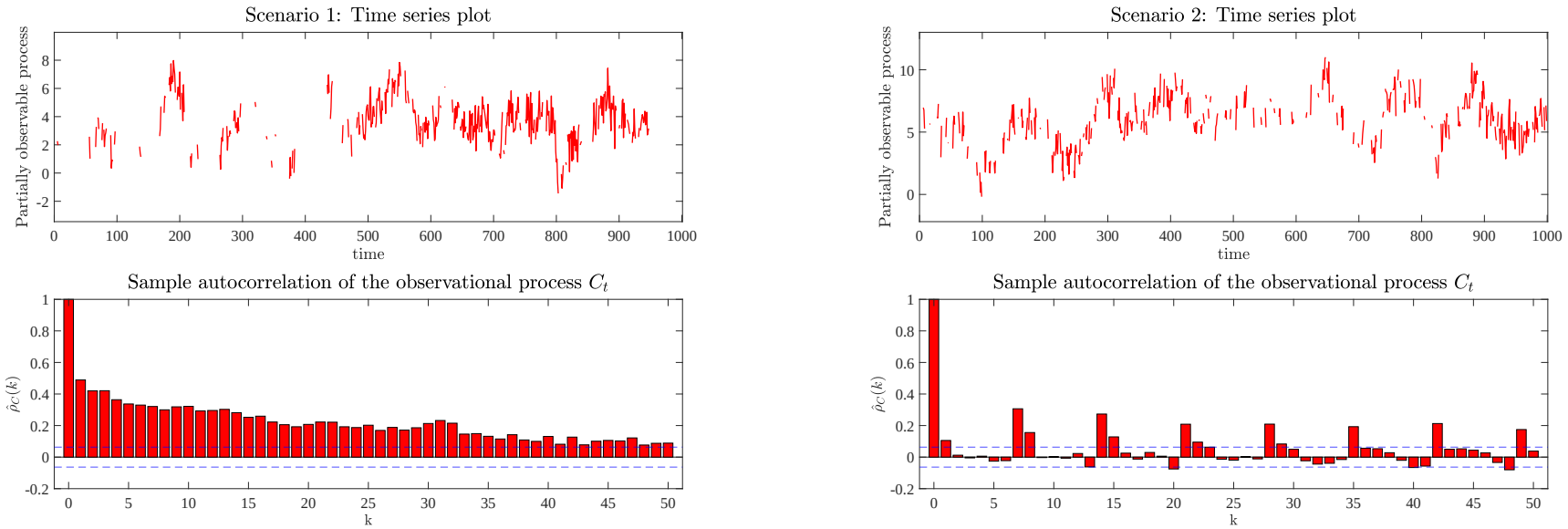
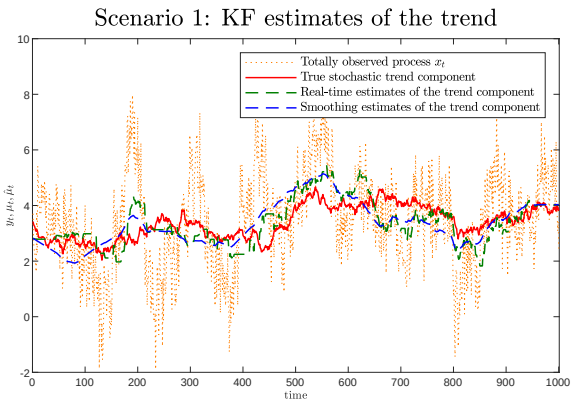
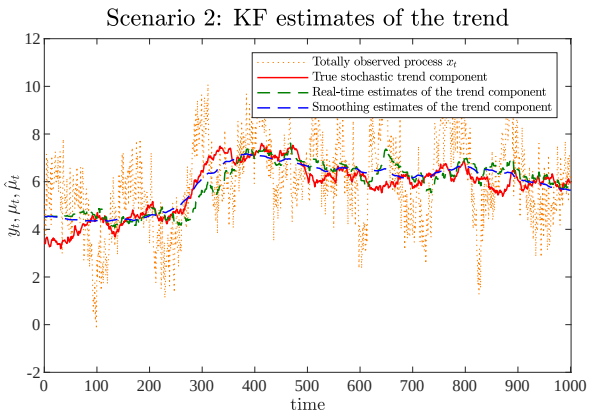


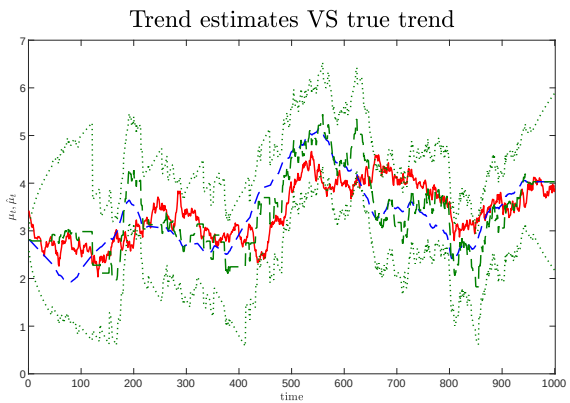
Figure: Partially observed processes generated according to the two scenarios and correlograms of the corresponding missing data generating processes. Missing observations are about 50% in both cases, such that the features of the hidden trend component are not easily identifiable from a visual inspection of the time series plot.

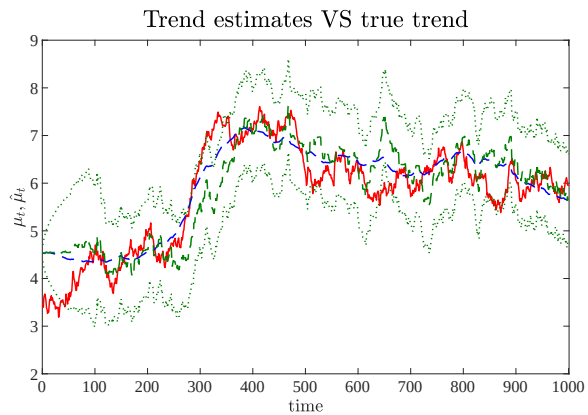
Simulation Exercise

Figure: Real-time (dashed green lines) and smoothed (dashed blue lines) KF estimates of the trend, against the true trend (red lines). The dotted green lines are the 95% confidence bands $\pm 1.96\hat{\text{Var}}(\mu_t|\mathcal{F}_t)^{1/2}$ computed via the KF.



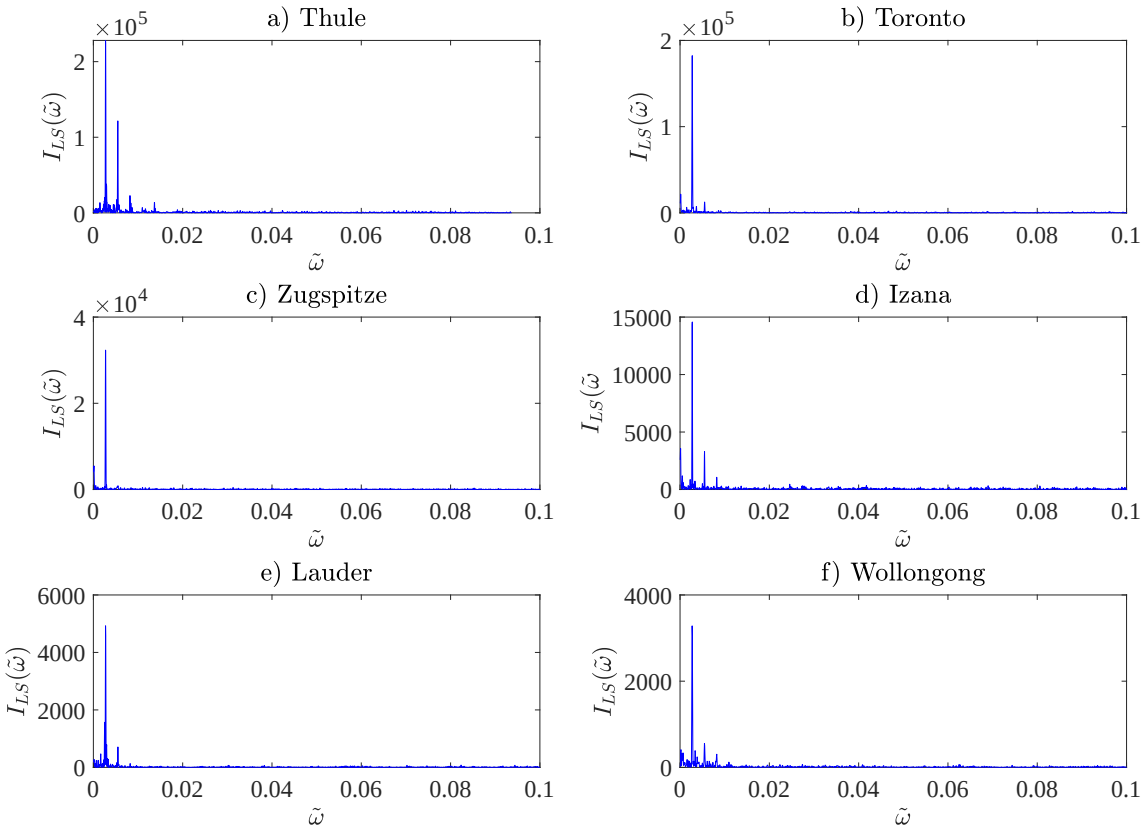






Main analysis and results

Figure: Let us present the main analysis and estimation results, when the structural model in (1) is fitted to the log transforms of the six ethane time series, we start via the inspection of the Lomb-Scargle periodograms.



Model Specification

- Set $\lambda = 2\pi/365.25$, which is the fundamental of the annual cycle, the other frequency parameters are set as its successive harmonics $lambda_1 j$ for $j = 2, \cdots, M$.
- The number of harmonics is set $M = 12$ and the length of the AR representation of the cyclical component is $p = 200$. We found that allowing for a higher value of M and p does not change results remarkably.

Main analysis and results

Table: **Model Estimation**: Exact log likelihood estimates of the parameters.

Thule

Trend	$\hat{\sigma}_\zeta$	$\hat{\mu}_0$			
	0.0022	2.9377			
Cycle	j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\zeta j}$	$\hat{\sigma}^2_{\psi j}$
	1	0.0172	0.4962	0.0805	0.2728
	2	0.0344	0.2022	0.0055	0.0000
	3	0.0516	-0.1041	0.0074	0.0001

Zugspitze

Trend	$\hat{\sigma}_\zeta$	$\hat{\mu}_0$			
	0.0030	2.3173			
Cycle	j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\zeta j}$	$\hat{\sigma}^2_{\psi j}$
	1	0.0172	0.4944	0.0230	0.0153
	2	0.0344	0.1514	0.0183	0.0004
	3	0.0516	0.1372	0.0279	0.0008

Lauder

Trend	$\hat{\sigma}_\zeta$	$\hat{\mu}_0$			
	0.0035	1.4696			
Cycle	j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\zeta j}$	$\hat{\sigma}^2_{\psi j}$
	1	0.0172	0.4635	0.1126	0.0610
	2	0.0344	0.2135	0.0143	0.0002
	3	0.0516	0.1506	0.0711	0.0053

Toronto

Trend	$\hat{\sigma}_\zeta$	$\hat{\mu}_0$			
	0.0032	3.1041			
Cycle	j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\zeta j}$	$\hat{\sigma}^2_{\psi j}$
	1	0.0172	0.4924	0.0450	0.0434
	2	0.0344	0.2646	0.0173	0.0004
	3	0.0516	0.1647	0.1520	0.0245

Izana

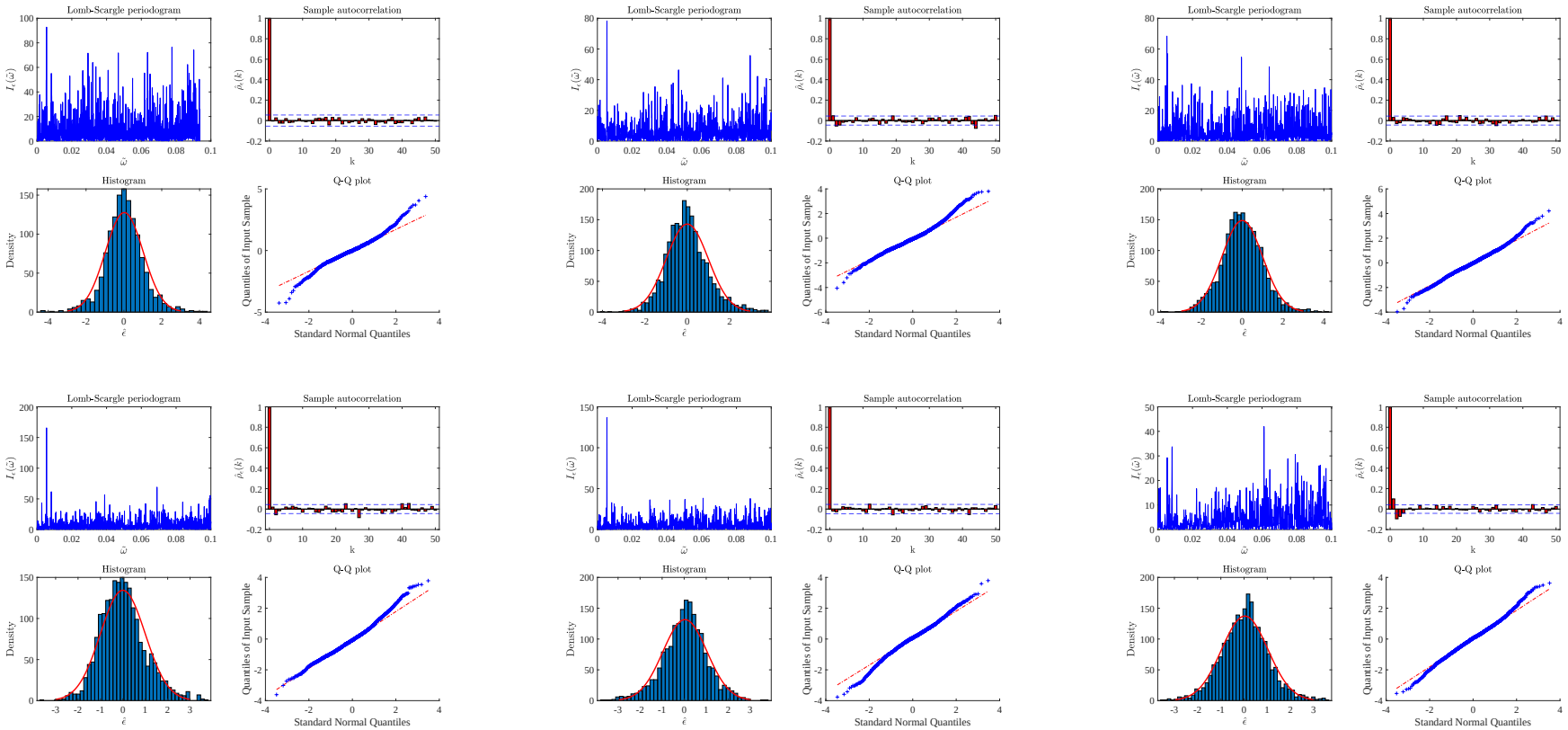
Trend	$\hat{\sigma}_\zeta$	$\hat{\mu}_0$			
	0.0018	2.1860			
Cycle	j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\zeta j}$	$\hat{\sigma}^2_{\psi j}$
	1	0.0172	0.4328	0.1198	0.0407
	2	0.0344	0.2945	0.0154	0.0003
	3	0.0516	0.3936	0.0357	0.0025

Wollongong

Trend	$\hat{\sigma}_\zeta$	$\hat{\mu}_0$			
	0.0021	1.6010			
Cycle	j	λ_j	$\hat{d}_j$	$\hat{\sigma}_{\zeta j}$	$\hat{\sigma}^2_{\psi j}$
	1	0.0172	0.4679	0.1335	0.0966
	2	0.0344	0.4011	0.0409	0.0035
	3	0.0516	0.3412	0.0018	0.0000

Main analysis and results

Figure: **Residual Diagnostics:** periodograms, correlograms, histograms and q-q plots. Thule (top left), Toronto (top mid), Zugspitze (top right), Izana (bottom left), Lauder (bottom mid) and Wollongong (bottom right).



Main analysis and results

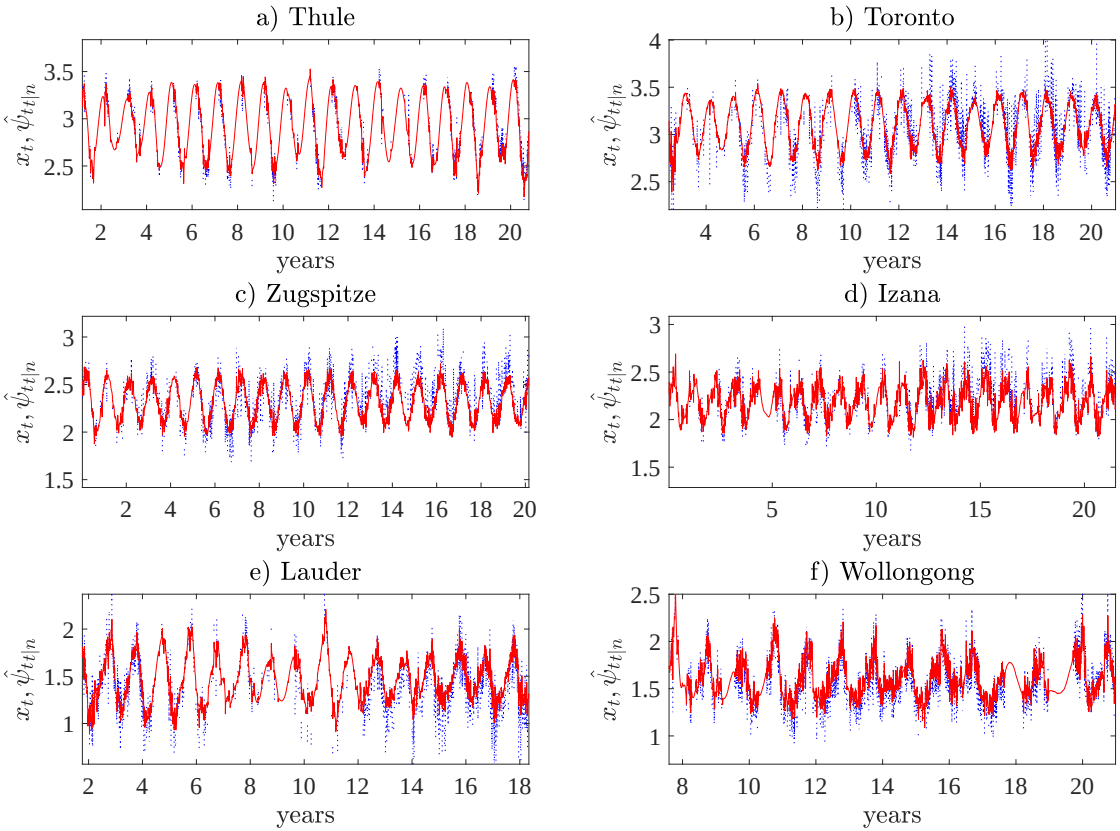


Figure: Smoothed estimates of the cycle component (red lines) for the Thule, Toronto, Zugspitze, Izana, Lauder and Wollongong series. The dotted blue lines in the background are the log transforms of the ethane time series.

Main analysis and results

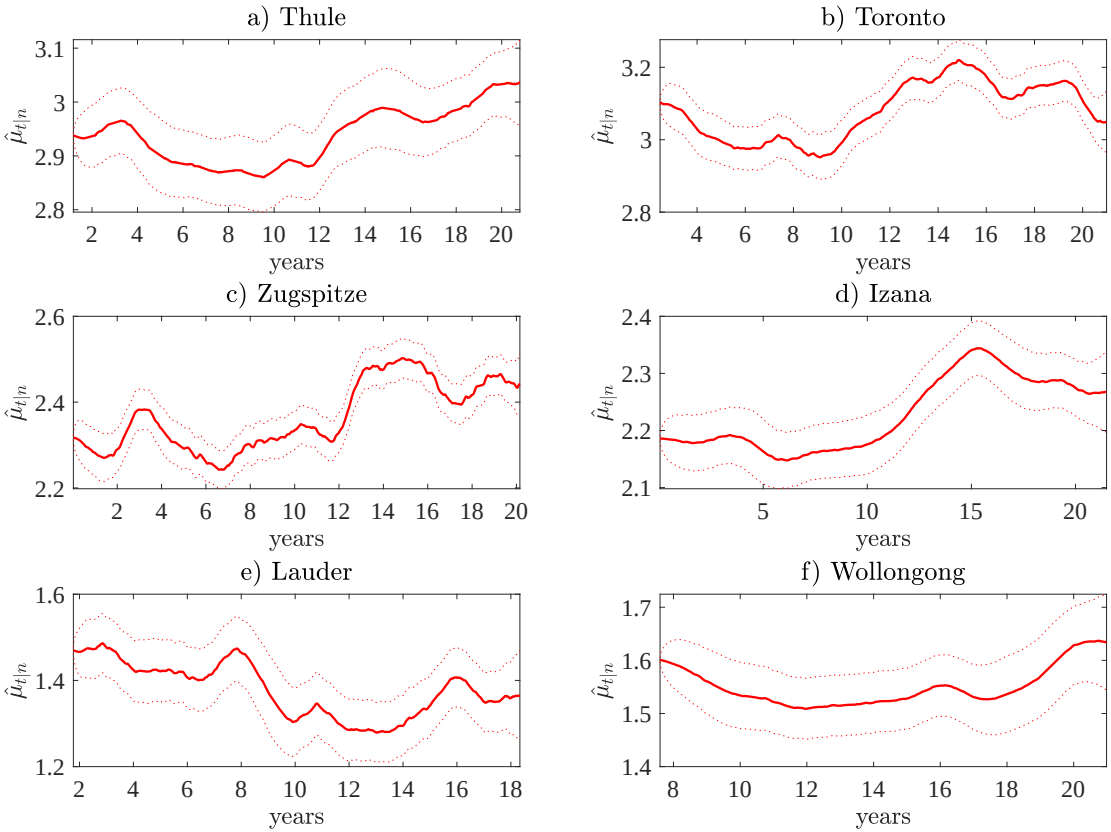


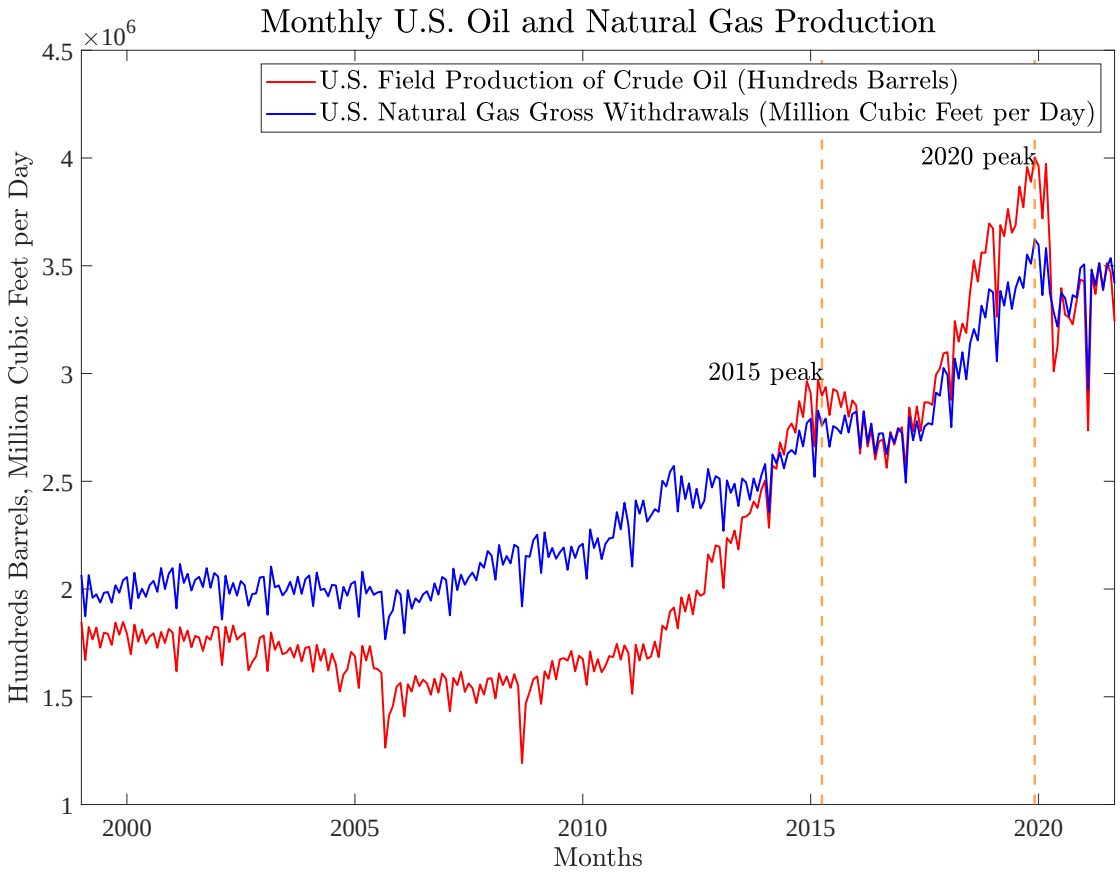
Figure: Smoothed estimates of the trend components (red lines) for the Thule, Toronto, Zugspitze, Izana, Lauder and Wollongong series. The dotted red lines are the 97.5% confidence bands obtained via the KF.

Summary

1. The inference on the parameters suggests that the ethane series may be driven by stochastic cycles.
2. The variability of the cyclical components is mostly depended by the first harmonic.
3. The Zugspitze trend 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

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 are mostly likely driven by biomass burning emissions, as the Australian bushfire season.
3. The common peak between the two trends estimates of Lauder and Wollongong around 2016 may be probably attributed to the fire taking place in the Victoria and Tasmania states during such period.
4. Finally, the huge peak in 2020 in the Wollongong trend can be associated to the last devastating bushfire season.
5. 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!