

# Particle filtering for Gumbel-distributed daily maxima of methane and nitrous oxide

Gwladys Toulemonde<sup>a\*</sup>, Armelle Guillou<sup>b</sup> and Philippe Naveau<sup>c</sup>

In atmospheric chemistry, daily maxima concentrations capture information about the variability among peak values. Statistically, they can often be modeled by a Gumbel distribution. This is the case for two very important greenhouse gases methane and nitrous oxide maxima when they are measured at our site of interest, Gif-sur-Yvette, a city south west of Paris. In practice, those two daily concentrations are not always recorded during the same period, and it would be of interest to reconstruct one from the other one. Such a type of inference can be handled within a state space modeling framework, but state space models are not tailored to represent the dynamics among Gumbel-distributed maxima. By building on our previous work, which made a link between linear autoregressive time series and Gumbel-distributed maxima, we propose and study such a state space model. It has the advantages of being linear and of preserving the Gumbel characteristic in both the state and observational equations. Concerning the inference of the hidden maxima at the state equation level, we derive the optimal weights of the auxiliary particle filtering approach of Pitt and Shephard. A simulation study indicates that our approach offers a gain over the Kalman filter, the bootstrap filter, and the nonmodified version of the Pitt and Shephard auxiliary filter. Copyright © 2012 John Wiley & Sons, Ltd.

**Keywords:** atmospheric chemistry; data assimilation; extreme value theory; hidden Markov models; particle filters; state space models

## 1. INTRODUCTION

Since at least 2002, the RAMCES network (Atmospheric Network for the Measurement of Greenhouse Compounds) has been recording daily maxima of nitrous oxide ( $N_2O$ ) and of methane ( $CH_4$ ) at Gif-sur-Yvette, a small city about 30 km south west of Paris, France. The two time series of daily maxima concentrations in parts per billion by volume displayed in Figure 1 for the period 2002–2007 represent complex temporal variations with peaks and missing data during a few time lags. Nitrous oxide is a greenhouse gas with tremendous global warming potential (has about more than 300 times of capability to trap heat in the atmosphere than carbon dioxide). Methane is also an important greenhouse gas with a global warming potential of 25 compared with carbon dioxide. In this context, the monitoring of daily maxima of  $CH_4$  and  $N_2O$  over long periods is paramount, for example, for health reasons. As the recording of the  $CH_4$  and  $N_2O$  appears to be intermittent, one may wonder if daily maxima concentrations of nitrous oxide could be inferred (reconstructed) from the methane observations (the inverse question can also be asked).

Moreover, since 2007, the RAMCES network has proceeded to record  $CH_4$  but has stopped the regular recordings of  $N_2O$ . That is why we would like to be able to reconstruct daily maxima of  $N_2O$  step by step in an online reasoning.

This inquiry can be explored if these two random variables are dependent. The scatter plot shown in Figure 2 clearly indicates a real but complex relationship between the two concentrations. The cloud shape in Figure 2 can be, in part, explained by the distributional nature of the random variables at hand. Coming back to Figure 1, the marginal distributions of those two atmospheric concentrations do not appear to be symmetrical around a mean value. Large concentrations tend to be more disperse than smaller ones. This is a clear indication that a symmetrical distribution such as a Gaussian one will not provide a reasonable fit, especially for the upper tail behavior of these two greenhouse gases. This discrepancy with the Gaussian paradigm is typical of random variables that are defined as daily maxima.

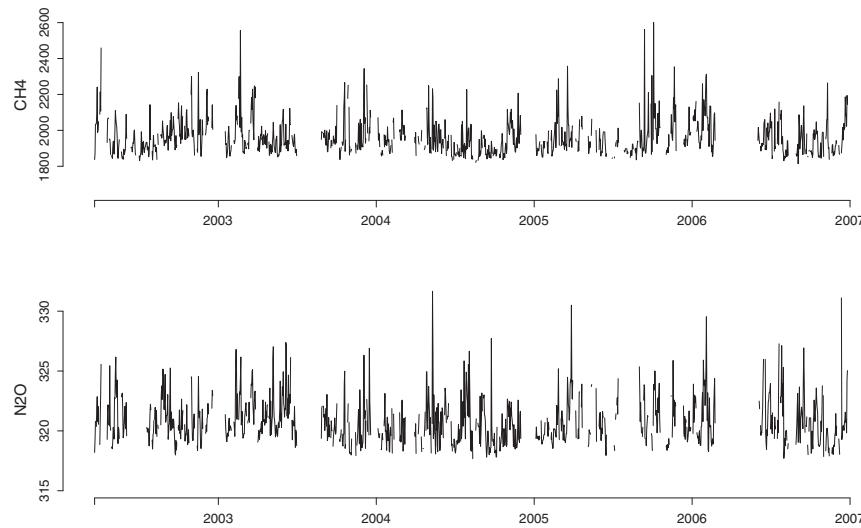
Since the seminal work of Fisher and Tippett (1928), extreme value theory (e.g., Embrechts *et al.*, 1997; Coles, 2001; Beirlant *et al.*, 2004; de Haan and Ferreira, 2006) has been specially developed to model distributions of maxima. Under suitable assumptions, correct normalized maxima should follow a generalized extreme value (GEV) distribution that merges three different tail behaviors: light (Gumbel type), heavy (Fréchet type), and bounded (Weibull type). Concerning daily maxima concentrations of  $CH_4$  and  $N_2O$ , our previous work (Toulemonde *et al.*, 2010) pointed out that a Gumbel distribution offers a reasonable fit for such data. The Gumbel cumulative distribution function is given by

\* Correspondence to: Gwladys Toulemonde, Institut de Mathématiques et de Modélisation de Montpellier, UMR CNRS 5149, Université Montpellier II, 34095 Montpellier, France. E-mail: gwladys.toulemonde@univ-montp2.fr

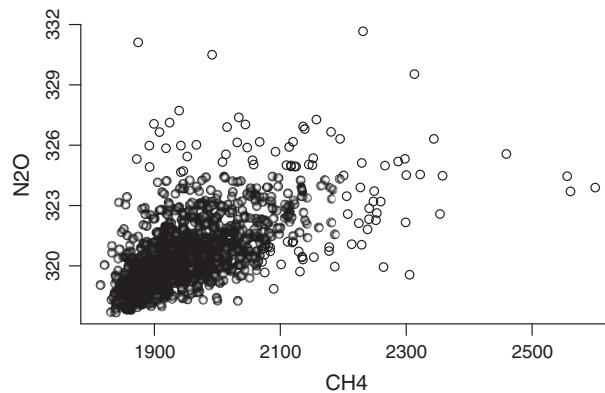
a Institut de Mathématiques et de Modélisation de Montpellier, UMR CNRS 5149, Université Montpellier II, 34095 Montpellier, France

b Institut de Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 67084 Strasbourg Cedex, France

c Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL-CNRS, 91191 Gif-sur-Yvette, France



**Figure 1.** Daily maxima of CH<sub>4</sub> and N<sub>2</sub>O during the period 2002–2007. Measurements in parts per billion by volume were made at LSCE, a laboratory located at Gif-sur-Yvette, a city south west of Paris, France. Data are missing during a few time lags, and daily maxima are computed over a block size of 24 h.



**Figure 2.** Scatter plot between daily maxima concentrations of CH<sub>4</sub> (*x*-axis) and N<sub>2</sub>O (*y*-axis)—see Figure 1.

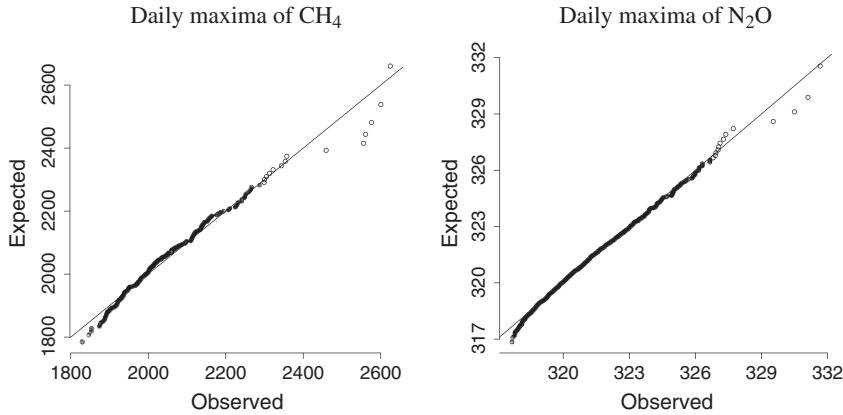
$$H_{\mu, \sigma}(x) = \exp \left\{ -\exp \left( -\frac{x - \mu}{\sigma} \right) \right\}, \text{ with } -\infty < x < +\infty \quad (1)$$

where  $\mu$  and  $\sigma$  correspond to the so-called location and scale parameters, respectively. As a visual check, the quantile–quantile plot displayed in the left panel of Figure 3 compares the ranked observed CH<sub>4</sub> maxima (*x*-axis) with their expected values obtained from a fitted Gumbel model (*y*-axis). The inference was made by implementing the method-of-moment technique studied in Proposition 4 in Toulemonde *et al.* (2010) in a temporal dependence context. The diagonal line indicates a perfect fit. Overall, the assumption that both daily CH<sub>4</sub> and N<sub>2</sub>O maxima are marginally Gumbel distributed appears to be reasonable.

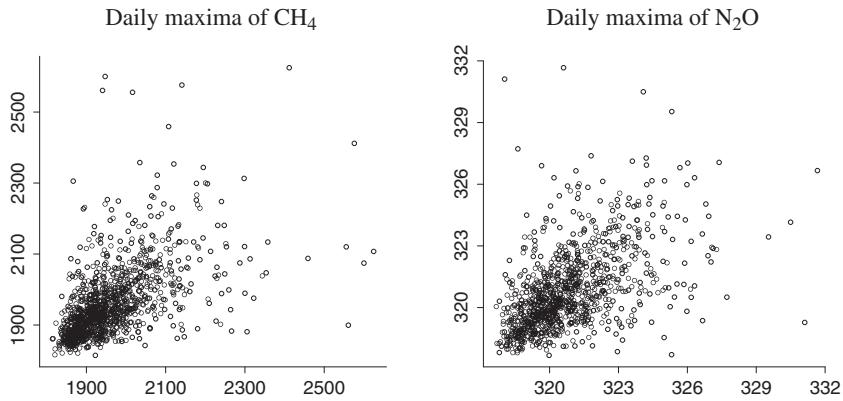
Consequently, this Gumbel hypothesis for the margins will be assumed throughout this paper. Concerning the temporal dependence, the scatter plot of consecutive maxima of CH<sub>4</sub> from day  $t$  (*x*-axis) to day  $t + 1$  (*y*-axis) shown in the left panel of Figure 4 indicates a short-term dependence. The same can be said from N<sub>2</sub>O maxima; see the right panel of Figure 4.

The inference of such dependences between bivariate maxima can be undertaken via different methods. For example, one could follow a copula approach (Joe, 1997) or estimate a bivariate extremal dependence function (e.g., Naveau *et al.*, 2009). Here, we opt for another modeling road because our goal is to infer unobserved daily maxima of N<sub>2</sub>O from measured daily maxima of methane. In addition, we would like to reproduce the temporal structure plotted in the left panel of Figure 4. The classical state space models formalism (e.g., Cappé *et al.*, 2005) represents an appealing solution to address these issues. The main difficulty here resides in ensuring that the marginals of daily maxima (see Figure 3) follow a Gumbel distribution within a state space modeling structure.

In a state space model context, the measurement  $Y_t$  recorded at time  $t$ , that is, the daily maxima of methane in our case, is described by two equations: the observational and the state ones. The first one depicts the link between  $Y_t$  and  $Z_t$ , where the latter represents an unobserved state variable, that is, the daily maxima of nitrous oxide in our application. The second equation, also called the system equation, represents a temporal dynamical structure, like the one shown in the right panel of Figure 4. Typically, most state space models have the following form (e.g., Doucet *et al.*, 2001)



**Figure 3.** Quantile-quantile plots of daily maxima of CH<sub>4</sub> and N<sub>2</sub>O obtained after fitting a Gumbel distribution via a method-of-moment technique proposed in Toulemonde *et al.* (2010). In (1), we obtain the following estimates with 95% confidence intervals:  $\hat{\sigma} = 79.8 \in [73.3; 86.4]$  and  $\hat{\mu} = 1915.9 \in [1904.4; 1927.4]$ , and for nitrous oxide,  $\hat{\sigma} = 1.52 \in [1.39; 1.64]$  and  $\hat{\mu} = 320.0 \in [319.7; 320.2]$ .



**Figure 4.** Scatter plots of consecutive maxima of CH<sub>4</sub> and N<sub>2</sub>O. The  $x$ -axis corresponds to day  $t$  and the  $y$ -axis to day  $t + 1$ . The empirical estimate of the lag 1 autocorrelation is equal to 0.55 for the CH<sub>4</sub> and to 0.52 for the N<sub>2</sub>O.

$$\begin{cases} Y_t = F_t(Z_t, \varepsilon_t) & \text{(observational equation)} \\ Z_t = G_t(Z_{t-1}, \eta_t) & \text{(state equation)} \end{cases}$$

Factorizing the joint probability density of states and observations denoted by  $p(Z_{0:T}, Y_{1:T})$ , we remark that such a state space model is determined by its initial law  $p(Z_0)$ , its transition density  $p(Z_t|Z_{t-1})$ , and its conditional probability density  $p(Y_t|Z_t)$  via

$$p(Z_{0:T}, Y_{1:T}) = p(Z_0) \prod_{t=1}^T p(Z_t|Z_{t-1})p(Y_t|Z_t)$$

where the vector  $Z_{0:T}$  denotes  $(Z_0, Z_1, \dots, Z_T)'$  and  $Y_{1:T} = (Y_1, \dots, Y_T)'$ . The main issue is to estimate  $Z_t$  from an a priori knowledge of  $p(Z_0)$  and from the observations vector  $Y_{1:T}$ . In other words, the conditional density  $p(Z_t|Y_{1:t})$ , called the filtering density, is of primary interest in state space modeling. To compute this density, two assumptions are often made in geosciences. The noises  $\varepsilon_t$  and  $\eta_t$  are assumed to be Gaussian, and the transfer functions  $F_t$  and  $G_t$  are assumed to be linear. This leads to the well-known Kalman filter equations (Kalman, 1960; Kalman and Bucy, 1961), which describe a sequential scheme for computing the conditional mean and the covariance matrix of the underlined system.

Going back to our daily maxima of concentrations, a Gaussian linear state space model cannot reproduce the behaviors shown in our four Figures 1–4. This remark leaves us with three options: to remove the Gaussian hypothesis, to discard the linearity assumption, or to do both. The Gumbel nature of the margins (see Figure 3) implies that the Gaussian hypothesis cannot be kept. For practical reasons, an additive linear structure offers many advantages. It is simple to explain and interpret; for example, the units such as parts per billion volume can be preserved. In contrast, introducing nonlinearity makes the inference and the error propagation more difficult to handle than in a linear framework. Hence, our strategy in this paper is to propose and study a linear state space model that preserves the Gumbel feature of our daily maxima. To implement this strategy, we build on our previous work (Toulemonde *et al.*, 2010) that focused on linear Gumbel autoregressive (AR) models. This extension to a state space context will be explained in Section 2. In particular, abandoning the Gaussian world for the

Gumbel one implies that the filtering law will not be explicit anymore and particle filtering techniques will be tailored to take advantage of the Gumbel constraint (see Section 3). Our introductory example will be treated in Section 4, and a few conclusions will be proposed in Section 5.

## 2. A GUMBEL STATE SPACE MODEL

In 2010, we developed a linear Gumbel AR model (Toulemonde *et al.*, 2010). The key ingredient of this AR model was that adding a Gumbel-distributed variable to the logarithm of a positive  $\alpha$ -stable variable can still be Gumbel distributed (see e.g., Fougères *et al.*, 2009). More precisely, if  $X$  is Gumbel distributed with parameters  $\mu$  and  $\sigma$  and independent of  $S_\alpha$ , a positive  $\alpha$ -stable variable, then the sum  $X + \sigma \log S_\alpha$  is also Gumbel distributed with parameters  $\mu$  and  $\sigma/\alpha$ . We recall here that a positive  $\alpha$ -stable variable (Zolotarev, 1986; Nolan, 2012) with  $\alpha \in (0, 1)$  is defined by its Laplace transform

$$\mathbb{E}(\exp(-uS)) = \exp(-u^\alpha), \text{ for all } u \geq 0 \quad (2)$$

In terms of notations, it is convenient to call the random variable  $\mu + \sigma \log S_\alpha$ , where  $S_\alpha$  is defined by (2), an exponential-stable variable with parameters  $\alpha \in (0, 1)$ ,  $\mu, \sigma > 0$ , denoted by  $\text{ExpS}(\alpha, \mu, \sigma)$ . This allows us to write down our state space model.

**PROPOSED MODEL.** Let  $\{Z_t, t \in \mathbb{Z}\}$  and  $\{Y_t, t \in \mathbb{Z}\}$  be two stochastic processes defined as follows:

$$\begin{cases} Y_t = v_t + H_t Z_t + \eta_{t,\alpha_2} & \text{(observational equation)} \\ Z_t = \alpha_1 Z_{t-1} + \varepsilon_{t,\alpha_1} & \text{(state equation)} \end{cases} \quad (3)$$

where  $H_t > 0$ ,  $\alpha_1 \in (0, 1)$ ,  $\alpha_2 \in (0, 1)$  and the sequences  $\{\varepsilon_{t,\alpha_1}\}_t$  and  $\{\eta_{t,\alpha_2}\}_t$  correspond to two independent samples of exponential-stable variables,  $\text{ExpS}(\alpha_1, -\sigma\gamma(1-\alpha_1), \alpha_1\sigma)$  and  $\text{ExpS}(\alpha_2, -H_t\sigma\gamma(1/\alpha_2-1), H_t\sigma)$ , respectively. The variable  $\varepsilon_{t,\alpha_1}$  is independent of  $\{Z_{t'}\}_{t' \leq t-1}$ , and the variable  $\eta_{t,\alpha_2}$  is independent of  $\{Z_{t'}\}_{t' \leq t}$ . The scalar  $\gamma$  is the Euler's constant.

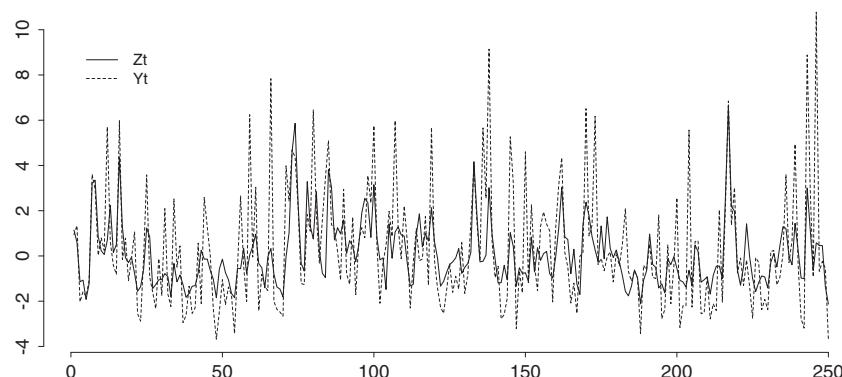
The state equation in (3) has a unique strictly stationary solution, which we will consider and where  $Z_t$  is Gumbel distributed. The observational equation corresponds to adding an exponential-stable noise to the Gumbel-distributed variable  $Z_t$ . This ensures that the margins of  $\{Y_t\}$  are also Gumbel distributed. More precisely, the variables  $Z_t$  and  $Y_t$  are Gumbel distributed with parameters  $(-\gamma\sigma, \sigma)$  and  $(v_t - \frac{H_t\gamma\sigma}{\alpha_2}, H_t \frac{\sigma}{\alpha_2})$ , respectively (see Appendix for details). Having a linear model makes the computation of covariances and correlation between  $\{Z_t\}$  and  $\{Y_t\}$  simple, and we obtain

$$\text{Cov}(Z_t, Z_{t-h}) = \alpha_1^{|h|} \text{Var}(Z_t) \quad (4)$$

$$\text{Cov}(Y_t, Z_t) = H_t \text{Var}(Z_t) \quad (5)$$

$$\text{Cor}(Y_t, Z_t) = \alpha_2 \quad (6)$$

These formulae allow for a straightforward interpretation of the main parameters  $\alpha_1$ ,  $\alpha_2$ , and  $H_t$  in terms of the dependences within  $\{Z_t\}$  (temporally) and between  $\{Z_t\}$  and  $\{Y_t\}$ . As an illustration of pathwise trajectories that model (3) can produce, Figure 5 presents a simulated realization of  $\{(Y_t, Z_t)\}_{t=1, \dots, T}$ , with  $T = 250$ ,  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.6$ ,  $H_t = \sigma = 1$ , and  $v_t = 0$ , and Figure 6 represents the scatter plot of successive values of  $Z_t$  (left panel) and the scatter plot of  $Y_t$  against  $Z_t$  (right panel). As expected from formulae (5) and (6), the path of  $\{Y_t\}$  mimics well the one of  $\{Z_t\}$  ( $\alpha_2 = 0.6$ ), and the associated scatter plot shows correspondences between both series. However, the variability within  $\{Y_t\}$  is larger than the one observed in  $\{Z_t\}$  ( $H_t = 1$  and  $\alpha_2 = 0.6$ ).



**Figure 5.** Simulated paths for  $\{Z_t\}$  (solid line) and  $\{Y_t\}$  (dotted line) from the system defined by (3) with  $T = 250$ ,  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.6$ ,  $H_t = \sigma = 1$ , and  $v_t = 0$ .

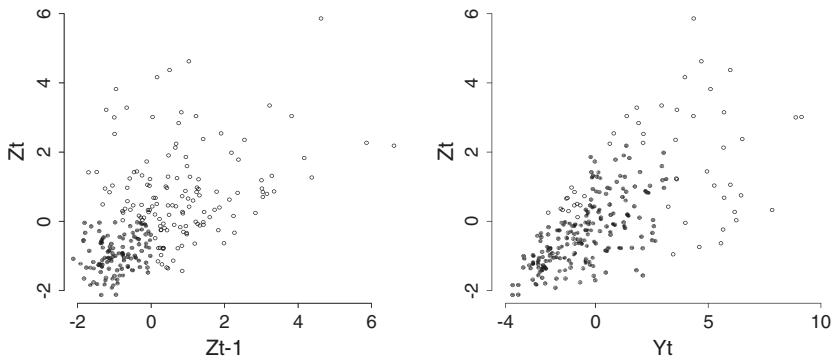


Figure 6. Scatter plot of consecutive values in the simulated hidden series  $Z_t$  and scatter plot between the simulated observed series  $Y_t$  and  $Z_t$ .

### 3. ESTIMATION OF THE FILTERING DENSITY

Under the assumption that the parameters of the state space model defined by (3) are known, we would like to approximate the filtering density  $p(Z_t|Y_{1:t})$  and the prediction density  $p(Z_t|Y_{1:t-1})$ . This leads to the classical recursive system in two steps (Doucet *et al.*, 2001; Cappé *et al.*, 2005) symbolized by arrows just as follows:

$$p(Z_{t-1}|Y_{1:t-1}) \xrightarrow{\text{prediction}} p(Z_t|Z_{t-1}) \xrightarrow{\text{correction}} p(Z_t|Y_{1:t})$$

with

$$p(Z_t|Y_{1:t-1}) = \int p(Z_t|Z_{t-1}) p(Z_{t-1}|Y_{1:t-1}) dZ_{t-1} \quad (\text{Prediction step})$$

$$p(Z_t|Y_{1:t}) = \frac{p(Y_t|Z_t) p(Z_t|Y_{1:t-1})}{\int p(Y_t|Z_t) p(Z_t|Y_{1:t-1}) dZ_t} \quad (\text{Correction step})$$

The prediction step at time  $t$  is based on the available information at step  $t-1$  about the hidden state  $Z_{t-1}$ . The prediction density is deduced by the filtering density at time  $t-1$  and by the transition density. For the correction step at time  $t$ , a new observation  $Y_t$  is available, and the current state information can be updated. According to the proposed model (3),  $Y_t$  and  $Z_t$  are Gumbel distributed and consequently have a skewed density. Thus, the mean and the variance do not entirely capture the full distribution. Consequently, applying the classical Kalman filter equations to estimate the hidden states is not optimal in this Gumbel setting.

#### 3.1. Presentation of particle filters

During the last decade, many authors (see, for example, Doucet *et al.*, 2000; Cappé *et al.*, 2007) worked on sequential Monte Carlo techniques to estimate the necessary integrals to obtain the filtering density. Classically, the set of weighted particles  $\{\xi_{t-1}^i, w_{t-1}^i\}_{i=1}^N$  approximates the filtering density  $p(Z_{t-1}|Y_{1:t-1})$ . The empirical distribution corresponding to this approximation is

$$p^N(Z_{t-1}|Y_{1:t-1}) = \sum_{i=1}^N w_{t-1}^i \delta_{\xi_{t-1}^i}(Z_{t-1})$$

In the sequel, for sake of simplicity, we note  $p(\cdot|\xi_{t-1}^i) = p(\cdot|Z_{t-1} = \xi_{t-1}^i)$ ,  $p(\cdot|\xi_t^i) = p(\cdot|Z_t = \xi_t^i)$  and  $q(\cdot|\xi_{t-1}^i, Y_t) = q(\cdot|Z_{t-1} = \xi_{t-1}^i, Y_t)$ . An approximation of the prediction density and the filtering density follows:

$$p^N(Z_t|Y_{1:t-1}) = \sum_{i=1}^N w_{t-1}^i p(Z_t|\xi_{t-1}^i)$$

$$p^N(Z_t|Y_{1:t}) \propto \sum_{i=1}^N w_{t-1}^i p(Y_t|Z_t) p(Z_t|\xi_{t-1}^i)$$

## The Bootstrap Filter (BF)

At time  $t = 0$

$$\xi_0^{1:N} \stackrel{iid}{\sim} p(Z_0)$$

At time  $0 < t \leq T$ ,

1) *Propagation*

$$\xi_t^i \sim p(Z_t | \xi_{t-1}^i) \text{ for } i = 1, \dots, N$$

2) *Computation of the weights for  $i = 1, \dots, N$*

$$w_t^i \leftarrow p(Y_t | \xi_t^i)$$

$$w_t^i \leftarrow \frac{w_t^i}{\sum_{j=1}^N w_t^j}$$

3) *Selection step*

simulate  $\phi_t^{1:N}$  from  $(w_t^{1:N}, \xi_t^{1:N})$  (Multinomial resampling)

$$\xi_t^{1:N} \leftarrow \phi_t^{1:N}.$$

According to the importance sampling principle, to obtain a set of weighted particles  $\{\xi_t^i, w_t^i\}_{i=1}^N$  that approximate  $p(Z_t | Y_{1:t})$ , one can sample particles  $\xi_t^i$  according to a proposal density  $q(\cdot | \xi_{t-1}^i, Y_t)$  and in computing the associated importance weights with the recursive relation

$$w_t^i \propto w_{t-1}^i \frac{p(Y_t | \xi_t^i) p(\xi_t^i | \xi_{t-1}^i)}{q(\xi_t^i | \xi_{t-1}^i, Y_t)}$$

A suitable choice for the proposal density  $q(\cdot | \xi_{t-1}^i, Y_t)$  has to be done; for instance, we can use the transition density  $p(\xi_t^i | \xi_{t-1}^i)$ . We consider this choice for which the update of the weights has the simpler expression  $w_t^i \propto w_{t-1}^i p(Y_t | \xi_t^i)$ , and this leads to a particular case of the sampling importance sampling algorithm. In a few iterations, this algorithm presents a problem of weights degeneracy in the sense that almost any weight is equal to zero. Gordon *et al.* (1993) introduced a resampling step in the algorithm, leading us to the sampling importance resampling algorithm, also called the bootstrap filter for this particular choice of the transition density as the proposed one. This algorithm is decomposed into three steps. The first step is an initialization step bringing into play an initial distribution. According to Toulemonde *et al.* (2010), the state equation in (3) has a unique strictly stationary solution, which is Gumbel distributed and which justifies the Gumbel distribution for the initial one in the algorithm. After a propagation step, the last one is a selection (resampling) step, where the particles are eliminated or multiplied according to their weights.

The resampling step improves the sampling importance sampling algorithm, but it leads to a progressive loss of diversity in the particle set. The particles are either eliminated or multiplied according to their adequacy with the new observation. In some cases, only one or two particles are selected. So they are multiplied and constitute the set of particles for the next step. Ideally, it would be better to force the propagation mechanism to directly propose particles in areas associated to high likelihood. To this aim, Pitt and Shephard (1999) proposed the auxiliary particle filter (APF) introducing first-stage weights called auxiliary variables denoted here by  $\beta$ . These new weights concentrate the computation effort on some promising particles. A description of this APF, which generalizes the sampling importance resampling filter, is given in the boxed-text on the next page.

The optimal auxiliary weights depend on  $p(Y_t | \xi_{t-1}^i)$ , which cannot always be computed. So the weights need to be estimated. Pitt and Shephard (1999) proposed to consider  $\hat{p}(Y_t | \xi_{t-1}^i) = p(Y_t | Z_t = \mu_t^i)$ , with  $\mu_t^i = \mathbb{E}(Z_t | \xi_{t-1}^i)$ . In the sequel, this filter is denoted by APF-PS $_N$ , with  $N$  corresponding to the number of particles.

For our proposed Gumbel state space model defined by (3), it is possible to compute optimal weights as a density (derivative with respect to  $y_t$ ):

$$\begin{aligned} p(y_t | \xi_{t-1}^i) &= \frac{\partial}{\partial y_t} \mathbb{P}(C + H_t \sigma(\alpha_1 \log S_{t,\alpha_1} + \log S_{t,\alpha_2}) \leq y_t) \\ &= \frac{\partial}{\partial y_t} \mathbb{P}\left(\alpha_1 \log S_{t,\alpha_1} + \log S_{t,\alpha_2} \leq \frac{y_t - C}{H_t \sigma}\right) \end{aligned}$$

$$\begin{aligned}
&= \frac{\partial}{\partial y_t} \int_0^\infty \mathbb{P} \left( \log S_{t,\alpha_2} \leq \frac{y_t - C}{H_t \sigma} - \alpha_1 \log S_{t,\alpha_1} \mid S_{t,\alpha_1} = s \right) f_{S_{t,\alpha_1}}(s) ds \\
&= \frac{\partial}{\partial y_t} \int_0^\infty \mathbb{P} \left( S_{t,\alpha_2} \leq \exp \left( \frac{y_t - C}{H_t \sigma} - \alpha_1 \log s \right) \mid S_{t,\alpha_1} = s \right) f_{S_{t,\alpha_1}}(s) ds \\
&= \frac{\partial}{\partial y_t} \int_0^\infty \mathbb{P} \left( S_{t,\alpha_2} \leq \exp \left( \frac{y_t - C}{H_t \sigma} - \alpha_1 \log s \right) \right) f_{S_{t,\alpha_1}}(s) ds \\
&= \int_0^\infty \frac{\partial}{\partial y_t} \mathbb{P} \left( S_{t,\alpha_2} \leq \exp \left( \frac{y_t - C}{H_t \sigma} - \alpha_1 \log s \right) \right) f_{S_{t,\alpha_1}}(s) ds \\
&= \int_0^\infty \frac{1}{H_t \sigma} \exp \left( \frac{y_t - C}{H_t \sigma} - \alpha_1 \log s \right) f_{S_{t,\alpha_2}} \left( \exp \left( \frac{y_t - C}{H_t \sigma} - \alpha_1 \log s \right) \right) f_{S_{t,\alpha_1}}(s) ds
\end{aligned} \tag{7}$$

where  $y_t$  is the observation at time  $t$ ,  $C = v_t - \frac{H_t \gamma \sigma}{\alpha_2} + H_t \alpha_1 \gamma \sigma + H_t \alpha_1 \xi_{t-1}^i$ ,  $S_{t,\alpha_i}$  ( $i = 1, 2$ ) are two independent  $\alpha_i$ -stable variables as defined in (2), and  $f_{S_{t,\alpha_i}}$  is the density of an  $\alpha_i$ -stable variable.

Such an integral is numerically time consuming because it involves two exponential-stable densities that have no explicit expression. To compute these optimal weights without increasing computation time, we approximate these weights by using a kernel estimator. According to (7), we have

$$p(y_t | \xi_{t-1}^i) = \frac{\partial}{\partial y_t} \mathbb{P} \left( U_{t,\alpha_1,\alpha_2} \leq \frac{y_t - C}{H_t \sigma} \right)$$

with  $U_{t,\alpha_1,\alpha_2} = \alpha_1 \log S_{t,\alpha_1} + \log S_{t,\alpha_2}$ . Denoting by  $f_{U_{t,\alpha_1,\alpha_2}}$  the associated density to  $U_{t,\alpha_1,\alpha_2}$ , we obtain

$$p(y_t | \xi_{t-1}^i) = \frac{1}{H_t \sigma} f_{U_{t,\alpha_1,\alpha_2}} \left( \frac{y_t - C}{H_t \sigma} \right) \tag{8}$$

There is no integral in this expression of  $p(y_t | \xi_{t-1}^i)$ . We just need to accurately estimate the light-tailed density  $f_{U_{t,\alpha_1,\alpha_2}}(\cdot)$  for any fixed  $\alpha_1$  and  $\alpha_2$ . This approximation of  $f_{U_{t,\alpha_1,\alpha_2}}$  has to be carried out only once in a preprocessing step (ie., before our filtering algorithm). This can be implemented by a kernel method using Gaussian kernel associated to a bandwidth chosen by a rule of thumb (Silverman, 1986). In practice, 1000 samples of size 10,000 are simulated as  $U_{t,\alpha_1,\alpha_2} = \alpha_1 \log S_{t,\alpha_1} + \log S_{t,\alpha_2}$ .

By this way, we propose an adapted APF to our model denoted in the sequel as the APF-Opt $_N$  filter with  $N$  corresponding to the number of particles.

### Auxiliary Particle Filter (APF)

At time  $t = 0$

$$\xi_0^{1:N} \stackrel{iid}{\sim} p(Z_0)$$

$$w_0^{1:N} \leftarrow \frac{1}{N}$$

At time  $0 < t \leq T$ ,

#### 1) Selection step

$$\beta_t^i \leftarrow w_{t-1}^i \hat{p}(Y_t | \xi_{t-1}^i)$$

simulate  $j^{1:N}$  from  $(\beta_t^{1:N}, 1 : N)$  (Multinomial resampling).

#### 2) Propagation

$$\xi_t^i \sim p(Z_t | \xi_{t-1}^{j^i}) \text{ for } i = 1, \dots, N$$

#### 3) Computation of the weights for $i = 1, \dots, N$

$$w_t^i \leftarrow \frac{p(Y_t | \xi_t^i)}{\hat{p}(Y_t | \xi_{t-1}^{j^i})}$$

$$w_t^i \leftarrow \frac{w_t^i}{\sum_{j=1}^N w_t^j}$$

### 3.2. Comparison of four different filters

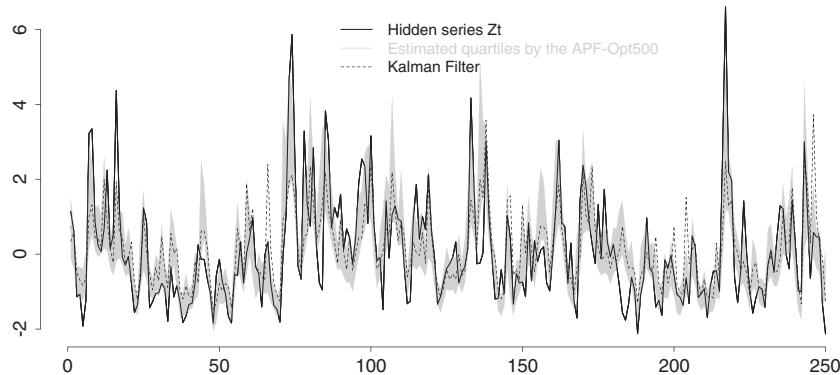
At this stage, we have four different filters at our disposal: Kalman filter (KF),  $BF_N$ , and the two versions of APF. The notations APF- $PS_N$  and APF- $Opt_N$  denote the classical Pitt and Shephard one and our version for which the weights are computed to respect the Gumbel structure of our model, respectively. For practical reasons, we set the number of particles to  $N = 500$  and the time length of the time series to  $T = 250$ . To compare the different filters, we have simulated 100 pairs of series according to the model defined in (3) with  $\alpha_1 \in \{0.1, 0.5, 0.9\}$ ,  $\alpha_2 \in \{0.4, 0.6\}$ ,  $H_t = \sigma = 1$ , and  $v_t = 0$ . Other choices for  $H_t$ ,  $\sigma$ , and  $v_t$  do not affect the conclusion of the simulation results. We aim to compare the estimated  $[Z_t|Y_{1:t}]$  with the unobserved  $Z_t$ .

As an example, in the case  $\alpha_1 = 0.5$  and  $\alpha_2 = 0.6$ , Figure 7 overlays the true hidden signal  $Z_t$  (solid line) with the empirical quartiles (gray) from our APF- $Opt_{500}$  method obtained from the  $Y_t$  displayed in Figure 5. The dotted line corresponds to the classical Kalman filter estimate. Unsurprisingly, the APF- $Opt_{500}$  outperforms the Kalman filter in this example. By construction, the Kalman filter cannot reproduce skewed distribution behaviors. This is not the case for the  $BF_N$  and the APF- $PS_N$  filters. To compare the different approaches, we compute for each series the mean square error (MSE):

$$MSE = \frac{1}{T} \sum_{t=1}^T (\hat{Z}_t - Z_t)^2 \quad (9)$$

where  $\hat{Z}_t$  represents the particle set average at each time  $t$ . Table 1 compares the mean of these MSEs based on 100 replica for the four filters in the six cases. It is important to notice that the computation times of the different particle filters are comparable. Our main tuning choice concerns the number of particles  $N$ . Table 2 compares the different filtering methods according to three different  $N$  (each row) in the case where the parameter choices ( $\alpha_1 = 0.5$  and  $\alpha_2 = 0.6$ ) correspond to our application at hand (see Section 4). It is sufficient to set  $N = 100$  for our method to outperform (in terms of MSE) the other approaches, even if the others have  $N = 1000$ . Hence, our choice of  $N$  should be rather a question of computation time. As the latter classically increases linearly with  $N$  for particle filtering methods, the number  $N = 500$  particles has been chosen as a compromise between performance and computing cost.

Coming back to Table 1, as expected from formulae (5) and (6), the gain in MSE is stronger when  $\alpha_2$  increases. This is also true when dependence (characterized by  $\alpha_1$ ) into the hidden series increases. Kalman filter gives generally the worst results and this is exacerbated when  $\alpha_1$  becomes high. In some cases (in particular with  $\alpha_1 = 0.9$ ), APF- $PS_{500}$  and APF- $Opt_{500}$  are comparable. Overall, the APF- $Opt_{500}$  outperforms the APF- $PS_{500}$  and the other filtering approaches.



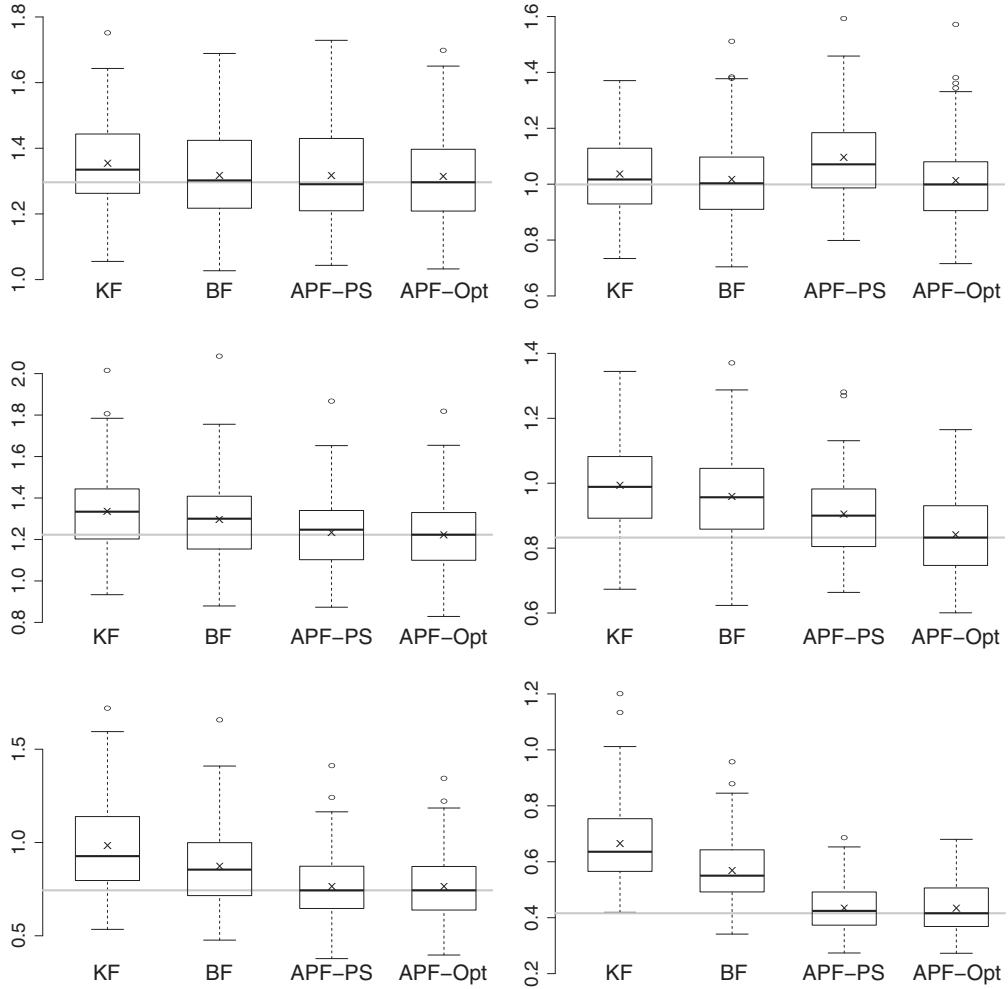
**Figure 7.** Comparison of filtered states with the Kalman filter (dotted line) and with the APF- $Opt_{500}$  (quartiles in gray area).

**Table 1.** Mean of the mean square errors based on 100 replica.

$\alpha_1$	$\alpha_2$	KF	$BF_{500}$	APF- $PS_{500}$	APF- $Opt_{500}$
0.1	0.4	1.354	1.317	1.317	<b>1.314</b>
0.1	0.6	1.036	1.017	1.096	<b>1.013</b>
0.5	0.4	1.336	1.296	1.233	<b>1.222</b>
0.5	0.6	0.994	0.959	0.905	<b>0.841</b>
0.9	0.4	0.984	0.873	<b>0.764</b>	<b>0.764</b>
0.9	0.6	0.665	0.569	<b>0.434</b>	<b>0.434</b>

**Table 2.** Mean of the mean square errors based on 100 replica for  $\alpha_1 = 0.5$  and  $\alpha_2 = 0.6$ .

$N$	KF	$BF_N$	$APF-PS_N$	$APF-Opt_N$
100	0.992	1.004	0.939	<b>0.876</b>
500	0.994	0.959	0.905	<b>0.841</b>
1000	0.989	0.948	0.897	<b>0.829</b>



**Figure 8.** Box plots of the mean square errors (MSEs) of the KF,  $BF_{500}$ ,  $APF-PS_{500}$ , and  $APF-Opt_{500}$ ,  $H_t = \sigma = 1$ ,  $\nu_t = 0$ . Results with  $\alpha_2 = 0.4$  (respectively  $\alpha_2 = 0.6$ ) are on the left (respectively on the right). On the top, the box plots correspond to  $\alpha_1 = 0.1$ ; on the middle,  $\alpha_1 = 0.5$ ; and on the bottom,  $\alpha_1 = 0.9$ . The crosses represent the means of the MSEs for each case. The whiskers extend to the most extreme data point, which is no more than 1.5 times the interquartile range from the box. The MSEs whose values are out of these whiskers are represented by circles.

Figure 8 summarizes the dispersion in the MSEs. Overall, the optimal filter provides the best results of the 100 computed MSEs in each case. In a general way, the auxiliary filters are better than the two others. For example, in the case  $\alpha_1 = 0.9$  and  $\alpha_2 = 0.6$ , the mean of the MSEs corresponding to the auxiliary filters is equal to the best MSE we have obtained on 100 series with the Kalman one.

We propose another simulation exercise highlighting the practical impact of our approach (inference of high quantiles) and exploring how return levels can be improved by using a filtering method and especially the APF-Opt filter. By simulating three months of stationary bivariate daily maxima  $\{(Z_t, Y_t)\}$  according to the proposed model (3), we can assume that, for the two last weeks, we only observe  $\{Y_t\}$  but not  $\{Z_t\}$  (considered as our missing data). By construction,  $Z_t$  is Gumbel distributed. To compute a return level for  $Z_t$ , we can either use the observed  $\{Z_t\}$  (no data during the two last weeks of the third month) or extend this time series by reconstructing the  $\{Z_t\}$  from the  $\{Y_t\}$ . After this procedure was repeated 1000 times, Table 3 presents MSEs for return levels computed from the Gumbel distribution associated with four return periods computed for four cases. The first case corresponds to the situation where only available observations have been

**Table 3.** Mean square errors for return levels based on 1000 replica for  $\alpha_1 = 0.5$  and  $\alpha_2 = 0.6$  with 500 particles.

Return period (year/s)	Incomplete data	APF-PS <sub>N</sub>	APF-Opt <sub>N</sub>	Whole data	True value
1	0.68	0.60	0.59	0.57	5.9
5	1.03	0.92	0.91	0.87	7.5
10	1.21	1.08	1.07	1.02	8.2
50	1.67	1.51	1.50	1.41	9.8

used. In other words, the return levels have been computed with *incomplete data*. In the cases denoted by *APF-PS<sub>N</sub>* and *APF-Opt<sub>N</sub>*, we have reconstructed the two last weeks with the auxiliary filters (the Pitt and Shephard one and the optimal one). Then, we compute the return levels with these extended data. The column *whole data* corresponds to return levels based on complete data. Obviously, results are better when the return levels are computed from the whole dataset than from incomplete data. The question is, *what could be gained by using the APF-Opt filter?* With these simulation results, we show that the MSEs with reconstructed data from the APF-Opt<sub>N</sub> are rather comparable with the MSEs with the whole data and consequently the benefit of using reconstruction by filtering approach is important.

As return levels represent a practical and theoretical summary of marginal extreme behaviors, this simulation study suggests that inferences on the Gumbel parameters distribution or other characteristics of extremal behavior could benefit from such a reconstruction of missing data within a dynamical system.

#### 4. ANALYSIS OF MAXIMA OF N<sub>2</sub>O FROM MAXIMA OF CH<sub>4</sub>

We divide our data sets into two periods (January 2002–June 2006 and July 2006–December 2006). The first period is used to calibrate our model parameters ( $\sigma$ ,  $\alpha_1$ ,  $\alpha_2$ ,  $\nu_t = \nu$ , and  $H_t = H$ ). The second period focuses on the problem of reconstruction of daily maxima of N<sub>2</sub>O ( $Z_{1:T}$  will denote the corresponding centered series) from the daily maxima of CH<sub>4</sub> ( $Y_{1:T}$ ).

For the calibration period, the parameters  $\sigma$  and  $\alpha_1$  are estimated as in Toulemonde *et al.* (2010) and are consequently almost surely consistent and asymptotically zero-mean Gaussian. The parameters  $\sigma$  and  $\alpha_1$  are respectively related to the variance of  $Z_t$  and to the covariance in (4). Natural estimators of these parameters have the following form:

$$\hat{\sigma} = \pi^{-1} \sqrt{6} s_Z$$

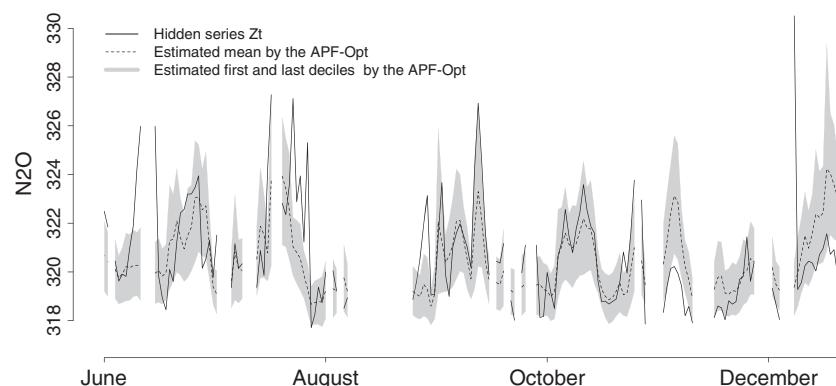
$$\hat{\alpha}_1 = \frac{1}{s_Z^2 T} \sum_{t=1}^{T-1} Z_t Z_{t-1}$$

where  $s_Z^2 = \sum_{t=1}^T Z_t^2 / T$ . As  $\nu = \mathbb{E}(Y_t)$ , a simple estimator  $\hat{\nu}$  is  $\bar{Y} = \sum_{t=1}^T Y_t / T$ . Concerning  $H$  and  $\alpha_2$ , we use respectively Equations (5) and (6) replacing moments by their empirical counterparts. This leads us to consider

$$\hat{H} = \frac{s_{YZ}}{s_Z^2}$$

$$\hat{\alpha}_2 = \frac{s_{YZ}}{s_Y s_Z}$$

where  $s_Y^2 = \sum_{t=1}^T (Y_t - \bar{Y})^2 / T$  and  $s_{YZ} = \sum_{t=1}^T Z_t (Y_t - \bar{Y}) / T$ .



**Figure 9.** Mean values of particles from APF-Opt500 (dotted line) and punctual empirical IC<sub>80%</sub> for the series of N<sub>2</sub>O daily maxima in Gif-sur-Yvette from June to December 2006.

We obtain the following estimations:  $\hat{\sigma} = 1.48$ ,  $\hat{\alpha}_1 = 0.53$ ,  $\hat{\alpha}_2 = 0.58$ ,  $\hat{\nu} = 1968$ , and  $\hat{H} = 30.8$ .

To reconstruct daily maxima of  $\text{N}_2\text{O}$  that we have hidden for the period from July to December 2006, we apply our filter (APF-Opt $N$ ). In Figure 9, the black line corresponds to the series of nitrous oxide and the dotted line corresponds to the filtered states (means of particle sets) using the previous parameters estimation and the auxiliary filter with optimal weights and with 500 particles. We also compare our filter with the auxiliary filter of Pitt and Shephard with 500 particles. If we compute the MSE as defined in Equation (9), we obtain a lower MSE with the optimal weights ( $N = 500$ ,  $\text{MSE} = 3.95$ ) than with the Pitt and Shephard weights ( $N = 500$ ,  $\text{MSE} = 4.19$ ).

## 5. CONCLUSION

We have proposed a new state space model in order to reconstruct a series of hidden light-tailed maxima. In order to estimate the hidden states, we compute the optimal weights of the auxiliary particle filter. In view of our simulation results and data analysis, this approach outperforms the other filters such as the auxiliary filter of Pitt and Shephard. In this paper, we have focused on light-tailed maxima, which is reasonable for many variables in atmospheric sciences. A possible extension concerns maxima stemming from heavy-tailed or bounded-tailed distribution. In other words, this corresponds to an extension from Gumbel marginals to GEV marginals. Considering  $X$  as a random variable from a GEV distribution with parameter  $\mu$ ,  $\sigma$ , and  $\xi$ , the quantity  $\log(X - \mu + \sigma/\xi)$  follows a Gumbel( $\log(\sigma/\xi)$ ,  $\xi$ ) distribution if  $\xi > 0$  and the quantity  $-\log(-X + \mu - \sigma/\xi)$  follows a Gumbel( $\log(-\xi/\sigma)$ ,  $-\xi$ ) distribution if  $\xi < 0$ . This leads us to a nonlinear state space model also depending on  $\alpha$ -stable random variables allowing  $Z_t$  and  $Y_t$  to be GEV distributed. Unfortunately, such a model is not additive anymore, and its complexity restricts its interest from a practitioner's point of view.

In an offline issue, estimating smoothing density would also be an interesting perspective. It corresponds to infer the distribution of the current state at a current time given past, present, and some future observations. Even if particle filtering techniques can be used in this case, it does not perform very well (Doucet and Johansen, 2011), and several particle smoothing methods for this problem mainly relying on the forward filtering-backward smoothing formula could be adapted to our problem.

## APPENDIX A. JUSTIFICATION OF THE DISTRIBUTIONS OF $Y_T$ AND $Z_T$

The distribution of  $Z_t$  is directly deduced from the paper of Toulemonde *et al.* (2010).

Using the fact that  $Z_t$  is Gumbel distributed with parameters  $-\gamma\sigma$  and  $\sigma$ , and using the key relationship between the Gumbel distribution and the positive  $\alpha$ -stable distribution, it is straightforward to show that  $Y_t$  is also Gumbel distributed with parameters  $\nu_t - \frac{H_t\gamma\sigma}{\alpha_2}$  and  $H_t\frac{\sigma}{\alpha_2}$ . A rigorous proof based on the computation of the characteristic function is detailed as follows.

$$\begin{aligned}\mathbb{E}(e^{iuY_t}) &= \mathbb{E}(e^{iuv_t} e^{iuH_t Z_t} e^{iu\eta_{t,\alpha_2}}) \\ &= e^{iuv_t} \mathbb{E}(e^{iuH_t Z_t}) \mathbb{E}(e^{iu\eta_{t,\alpha_2}})\end{aligned}$$

Because  $Z_t$  is Gumbel distributed with parameters  $-\gamma\sigma$  and  $\sigma$  and  $\eta_{t,\alpha_2} = -\gamma\sigma H_t(1/\alpha_2 - 1) + H_t\sigma \log S_{\alpha_2}$ , with  $S_{\alpha_2}$  defined by (2), we obtain

$$\begin{aligned}\mathbb{E}(e^{iuY_t}) &= e^{iuv_t} e^{-iuy\sigma H_t} \Gamma(1 - iuH_t\sigma) e^{-iuy\sigma H_t(1/\alpha_2 - 1)} \frac{\Gamma\left(1 - \frac{iuH_t\sigma}{\alpha_2}\right)}{\Gamma(1 - iuH_t\sigma)} \\ &= e^{iu[\nu_t - (\gamma\sigma H_t)/\alpha_2]} \Gamma\left(1 - \frac{iuH_t\sigma}{\alpha_2}\right)\end{aligned}$$

which corresponds to the characteristic function of a Gumbel random variable with parameters  $\nu_t - (\gamma\sigma H_t)/\alpha_2$  and  $(H_t\sigma)/\alpha_2$ .

## Acknowledgements

We would like to thank Frédéric Chevallier of LSCE (France) for his insightful comments about  $\text{CH}_4$  and  $\text{N}_2\text{O}$  data and the RAMCES team for providing those data. Part of this work has been supported by the EU-FP7 ACQWA Project ([www.acqwa.ch](http://www.acqwa.ch)) under Contract No. 212250, by the PEPER-GIS project, by the ANR-MOPERA project, by the ANR-McSim project, and by the MIRACCLE-GICC project. Finally, we would like to thank the editor and the two anonymous reviewers whose insightful suggestions have significantly improve the scope and the clarity of this work.

## REFERENCES

- Beirlant J, Goegebeur Y, Segers J, Teugels J. 2004. *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics: Chichester.
- Cappé O, Godsill SJ, Moulines E. 2007. An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE* **95**: 899–924.
- Cappé O, Moulines E, Rydén T. 2005. *Inference in Hidden Markov Models*. Springer: New York.
- Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics. Springer-Verlag: London.
- Doucet A, Dodsill S, Andrieu C. 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* **10**: 197–208.
- Doucet A, de Freitas N, Gordon N. 2001. *Sequential Monte Carlo Methods in Practice*, Statistics for engineering and Information Science. Springer-Verlag: New York.

- Doucet A, Johansen AM. 2011. A tutorial on particle filtering and smoothing: fifteen years later. In *The Oxford Handbook of Nonlinear Filtering*, Crisan D, Rozovsky B (eds). Oxford University Press: Oxford; 656–704.
- Embrechts P, Klüppelberg C, Mikosch T. 1997. *Modelling Extremal Events for Insurance and Finance*, Applications of Mathematics, Vol. 33. Springer-Verlag: Berlin.
- Fisher RA, Tippett LHC. 1928. Limiting forms of the frequency distribution in the largest particle size and smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* **24**: 180–190.
- Fougères AL, Nolan JP, Rootzén H. 2009. Models for dependent extremes using stable mixtures. *Scandinavian Journal of Statistics* **36**: 42–59.
- Gordon NJ, Salmond DJ, Smith AFM. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F on Radar and Signal Processing* **140**: 107–113.
- de Haan L, Ferreira A. 2006. *Extreme Value Theory: An Introduction*, Springer Series in Operations Research. Springer: New York.
- Joe H. 1997. *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability, Vol. 73. Chapman & Hall: London.
- Kalman RE. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions of the ASME, Series D* **82**: 35–45.
- Kalman RE, Bucy R. 1961. New results in linear filtering and prediction theory. *Journal of Basic Engineering, Transactions of the ASME, Series D* **83**: 95–108.
- Naveau P, Guillou A, Cooley D, Diebolt J. 2009. Modeling pairwise dependence of maxima in space. *Biometrika* **96**: 1–17.
- Nolan JP. 2012. *Stable Distributions—Models for Heavy Tailed Data*. Birkhäuser: Boston. In progress, Chapter 1 online at academic2.american.edu/~jpnnolan.
- Pitt MK, Shephard N. 1999. Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association* **94**: 590–599.
- Silverman BW. 1986. *Density Estimation*. Chapman and Hall: London.
- Toulemonde G, Guillou A, Naveau P, Vrac M, Chevallier F. 2010. Autoregressive models for maxima and their applications to CH<sub>4</sub> and N<sub>2</sub>O. *Environmetrics* **21**: 189–207.
- Zolotarev VM. 1986. *One-Dimensional Stable Distributions*, American Mathematical Society Translations of Mathematical Monographs, Vol. 65. American Mathematical Society: Providence, RI. Translation of the original 1983 (in Russian).