European Geosciences Union General Assembly 2012
Vienna, Austria, 22-27 April 2011
Session HS7.5: Hydroclimatic Stochastics

A Bayesian approach to hydroclimatic prognosis using the Hurst-Kolmogorov stochastic process

H. Tyralis and D. Koutsoyiannis
Department of Water Resources and Environmental Engineering
National Technical University of Athens

(Presentation available online: itia.ntua.gr/1202)

1. Abstract

It has now been well recognized that hydrological processes exhibit a scaling behaviour, also known as the Hurst phenomenon. An appropriate way to model this behaviour is to use the Hurst-Kolmogorov stochastic process. This process is associated with large scale fluctuations and also enhanced uncertainty in the parameter estimation. When we have to make a prognosis for the future evolution of the process, the total uncertainty must be evaluated. The proper technique to this is provided by Bayesian methods. We develop a Bayesian framework with Monte Carlo implementation for the uncertainty estimation of future prognoses assuming a Hurst-Kolmogorov stochastic process with a non-informative prior distribution of parameters. We derive the posterior distribution of the parameters and use it to make inference for future hydroclimatic variables.

2. Examined cases – data sets

Boeoticos Kephisos river basin

Berlin

Vienna

The cases examined here are:

- Temperature, rainfall and runoff at the Boeoticos Kephisos river basin which is part of the water supply system of Athens. Its climate is Mediterranean.
- Temperature at Berlin which has a humid continental climate.
- Temperature at Vienna which lies within a transition of oceanic climate and humid continental climate.

Examined data sets

	Kephisos Boeoticos river basin		
	Runoff (mm)	Rainfall (mm)	Temperature (°C)
Start year	1908	1908	1898
End year	2003	2003	2003
Size, n	96	96	106

	Berlin	Vienna
	Temperature (°C)	Temperature (°C)
Start year	1756	1775
End year	2009	2009
Size, n	254	235

3. Definitions

We assume that $\{\underline{x}_t\}$ is a stationary Gaussian stochastic process with mean μ, standard deviation σ and autocorrelation matrix $\boldsymbol{R}_n$ with elements $r_{ij} = \rho_{|i-j|}$, $i,j = 1,2, \ldots,n$, where $\rho_{|i-j|}$, the autocorrelation function (ACF), is a function of a parameter $\boldsymbol{\varphi}$ and $\boldsymbol{\theta} = (\mu, \sigma^2, \boldsymbol{\varphi})$ the parameter of the process. The distribution of the variable $\underline{\boldsymbol{x}}_n = (\underline{x}_1 \ldots \underline{x}_n)$ is given by

$$f(\boldsymbol{x}_n|\boldsymbol{\theta}) = (2\pi)^{-n/2} \, |\sigma^2 \boldsymbol{R}_n|^{-1/2} \exp[(-1/2\sigma^2) \, (\boldsymbol{x}_n - \mu \, \boldsymbol{e}_n)^{\mathrm{T}} \boldsymbol{R}_n^{-1} (\boldsymbol{x}_n - \mu \, \boldsymbol{e}_n)] \tag{1}$$

where $\boldsymbol{e}_n = (1 \; 1 \; \ldots \; 1)^{\mathrm{T}}$ is a vector with n elements.

For white noise (WN), the ACF is given by

$$\rho_0 = 1, \rho_k = 0, k = 1,2,\ldots, \tag{2}$$

For a first-order autoregressive (AR(1)) stochastic process, the ACF is given by

$$\rho_k = \varphi_1^k, k = 0, 1,\ldots, |\varphi_1| < 1 \tag{3}$$

For a Hurst-Kolmogorov (HK) stochastic process, the ACF is given by

$$\rho_k = |k + 1|^{2H} / 2 + |k - 1|^{2H} / 2 - |k|^{2H}, k = 0, 1,\ldots, 0 < H < 1 \tag{4}$$

4. Posterior distributions of the parameters

We assume that the non-informative distribution of $\underline{\boldsymbol{\theta}}$ is

$$\pi(\boldsymbol{\theta}) \propto 1/\sigma^2 \tag{5}$$

The posterior distribution of the parameters does not have a closed form. It is easily shown (see also Falconer and Fernadez, 2007[1] for some results and Tyralis and Koutsoyiannis, 2012[3] for more detailed results) that

$$\underline{\mu}|\sigma^2, \boldsymbol{\varphi}, \boldsymbol{x}_n \sim N[(\boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{e}_n)/(\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{e}_n), \sigma^2/(\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{e}_n)] \tag{6}$$

$$\underline{\sigma}^2|\boldsymbol{\varphi}, \boldsymbol{x}_n \sim \mathit{Inv\text{-}gamma}\{(n-1)/2, [\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{e}_n \boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{x}_n - (\boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{e}_n)^2]/(2\, \boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{e}_n)\} \tag{7}$$

$$\pi(\boldsymbol{\varphi}|\boldsymbol{x}_n) \propto |\boldsymbol{R}_n|^{-1/2} [\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{e}_n \boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{x}_n - (\boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{e}_n)^2]^{-(n-1)/2} (\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{e}_n)^{n/2-1} \tag{8}$$

We can obtain a simulated sample from this mixture (see for definition of mixture Gamerman and Lopes, 2006,[2]) simulating from $\pi(\boldsymbol{\varphi}|\boldsymbol{x}_n)$ using a MCMC algorithm and later from the known normal and inverse gamma distributions.

[1]Falconer, K., and Fernadez, C. (2007). "Inference on fractal processes using multiresolution approximation", Biometrica, 94 (2), 313-334.
[2]Gamerman, D., and Lopes, H. (2006). "Markov Chain Monte Carlo Stochastic Simulation for Bayesian inference", second edition, Chapman & Hall, London.
[3]See details concerning our method in Tyralis, H., and Koutsoyiannis, D. (2012). "A Bayesian statistical model for posterior prediction of hydroclimatic variables", (in preparation).

5. Posterior predictive distributions

We define $\underline{\boldsymbol{x}}_{n+1,n+m} := (\underline{x}_{n+1},\ldots,\underline{x}_{n+m})$. The posterior predictive distribution of $\underline{\boldsymbol{x}}_{n+1,n+m}$ given $\boldsymbol{\theta}$ and $\boldsymbol{x}_n$ is

$$f(\boldsymbol{x}_{n+1,n+m}|\boldsymbol{\theta},\boldsymbol{x}_n) = (2\pi\sigma^2)^{-m/2}\,|\boldsymbol{R}_{m|n}|^{-1/2}\,\exp[(-1/2\sigma^2)\,(\boldsymbol{x}_{n+1,n+m} - \boldsymbol{\mu}_{m|n})^{\mathrm{T}}\,\boldsymbol{R}_{m|n}^{-1}\,(\boldsymbol{x}_{n+1,n+m} - \boldsymbol{\mu}_{m|n})] \quad (9)$$

where $\boldsymbol{\mu}_{m|n}$ and $\boldsymbol{R}_{m|n}$ are given by:

$$\boldsymbol{\mu}_{m|n} = \mu\boldsymbol{e}_m + \boldsymbol{R}^{\mathrm{T}}_{[(n+1):(n+m)]\,[1:n]}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,(\boldsymbol{x}_n - \mu\boldsymbol{e}_n) \quad (10)$$

$$\boldsymbol{R}_{m|n} = \boldsymbol{R}_{[(n+1):(n+m)]\,[(n+1):(n+m)]} - \boldsymbol{R}^{\mathrm{T}}_{[1:n]\,[(n+1):(n+m)]}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{R}_{[1:n]\,[(n+1):(n+m)]} \quad (11)$$

The posterior predictive distribution of $\underline{\boldsymbol{x}}_{n+m+1,n+m+l} := (\underline{x}_{n+m+1},\ldots,\underline{x}_{n+m+l})$, given $\boldsymbol{x}_n$ and $\boldsymbol{\theta}$ as $m \to \infty$ is:

$$f(\boldsymbol{x}_{n+m+1,n+m+l}|\boldsymbol{\theta},\boldsymbol{x}_n)=(2\pi\sigma^2)^{-l/2}|\boldsymbol{R}_{l|n}|^{-1/2}\exp[(-1/2\sigma^2)(\boldsymbol{x}_{n+m+1,n+m+l}-\boldsymbol{\mu}_{l|n})\boldsymbol{R}_{l|n}^{-1}(\boldsymbol{x}_{n+m+1,n+m+l}-\boldsymbol{\mu}_{l|n})^{\mathrm{T}}] \quad (12)$$

where $\boldsymbol{\mu}_{l|n} = \mu\boldsymbol{e}_l$ and $\boldsymbol{R}_{l|n} = \boldsymbol{R}_{[1:l]\,[1:l]}$.

6. Climatic variable of interest

Following the framework by Koutsoyiannis et al. (2007[4]) we define the climatic variable of interest to be the 30-year moving average as follows:

$$\underline{x}_t^{30} := (1/30)\left(\sum_{l=t-29}^{n} x_l + \sum_{l=n+1}^{t} \underline{x}_l\right),\ t = n+1, \ldots, n+29 \text{ and } \underline{x}_t^{30} := (1/30)\sum_{l=t-29}^{t} \underline{x}_l,\ t = n+30, n+31, \ldots \quad (13)$$

To simulate from the distribution of this variable, we first simulate from (6),(7),(8) and then use the posterior samples (μ,σ,H) to simulate from (9) or (12). We examine the following cases.

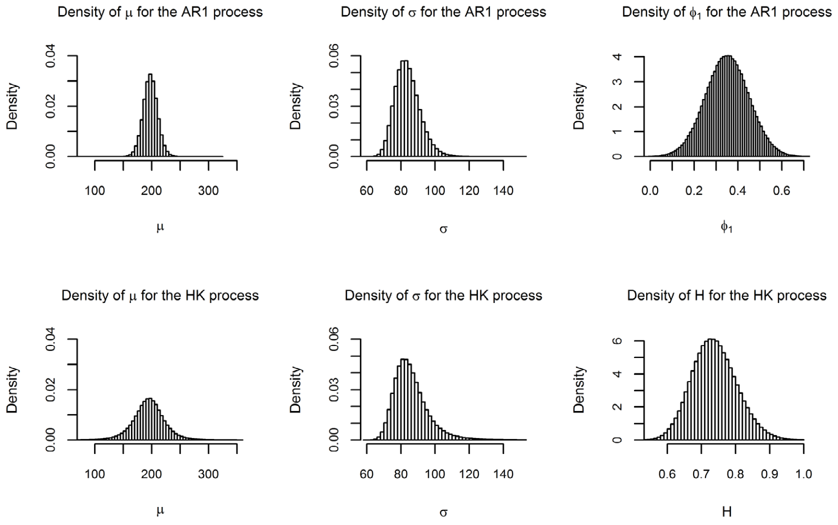
- White Noise.
- AR(1).
- Asymptotic behaviour of AR(1) ($m \to \infty$).
- HK, where we consider that H is known and equal to its maximum likelihood estimate (see Tyralis and Koutsoyiannis, 2011[5]).
- HK, where we consider that H is unknown.
- Asymptotic behaviour of HK ($m \to \infty$, H unknown).

(4) Koutsoyiannis, D., Efsratiadis, A., and Georgakakos, K.P. (2007). "Uncertainty assessment of Future Hydrovlimatic Predictions: A Comparison of Probabilistic and Scenario-Based Approaches", Journal of Hydrometeorology", 8 (3), 261-281.

(5)Tyralis, H., and Koutsoyiannis, D. (2011). "Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process", Stochastic Environmental Research & Risk Assessment", 25 (1), 21-33.

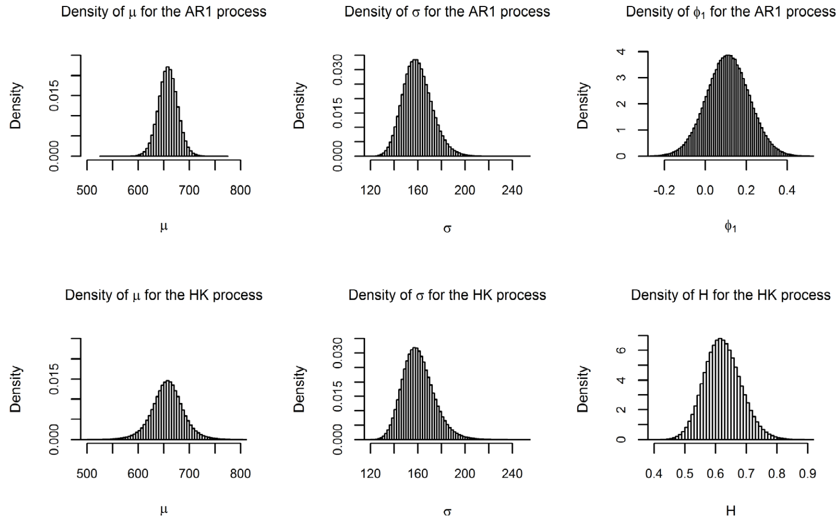
7. Posterior probability distributions for the AR(1) and HK parameters for Boeoticos Kephisos river basin

Posterior distributions for the parameters of Boeoticos runoff

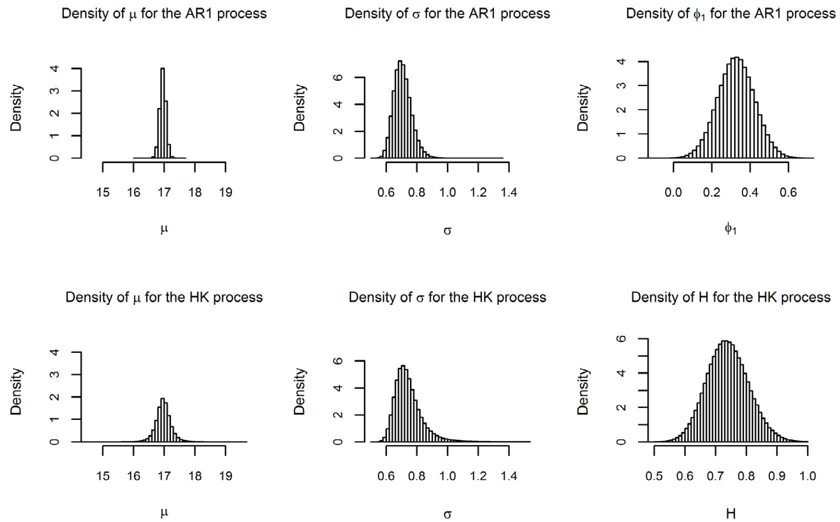


Case	Mean	2.5%	25%	50%	75%	97.5%
		Quantiles				
Boeoticos runoff						
μ	197.7	172.6	189.4	197.6	205.9	222.9
σ	83.930593	71.48	78.77	83.22	88.29	100.45
φ_1	0.3513	0.1572	0.2847	0.3511	0.4176	0.5462
μ	194.7672	131.8	178.1	195	211.6	256.4
σ	86.67381	71.21	79.17	84.43	91.09	114.49
H	0.7386	0.6177	0.6930	0.7357	0.7812	0.8759
Aliartos rainfall						
μ	658.19571	621.6	646	658.2	670.4	694.8
σ	159.9	138.3	151.3	159	167.5	186.2
φ_1	0.1106	−0.09254	0.04089	0.11049	0.18012	0.31521
μ	657.11322	592.6	638.4	657.3	676	720.6
σ	160.6	138	151.4	159.5	168.5	190.1
H	0.6235605	0.5139	0.5823	0.6209	0.662	0.7475
Aliartos temperature						
μ	16.96	16.76	16.89	16.96	17.02	17.15
σ	0.7072	0.6085	0.6668	0.7022	0.7419	0.8351
φ_1	0.3267542	0.1388	0.2624	0.3266	0.3909	0.5154
μ	16.97	16.44	16.83	16.97	17.11	17.52
σ	0.749	0.6174	0.6844	0.7293	0.7867	0.9921
H	0.7397	0.6119	0.6929	0.7377	0.7844	0.8792

Posterior distributions for the parameters of Aliartos rainfall

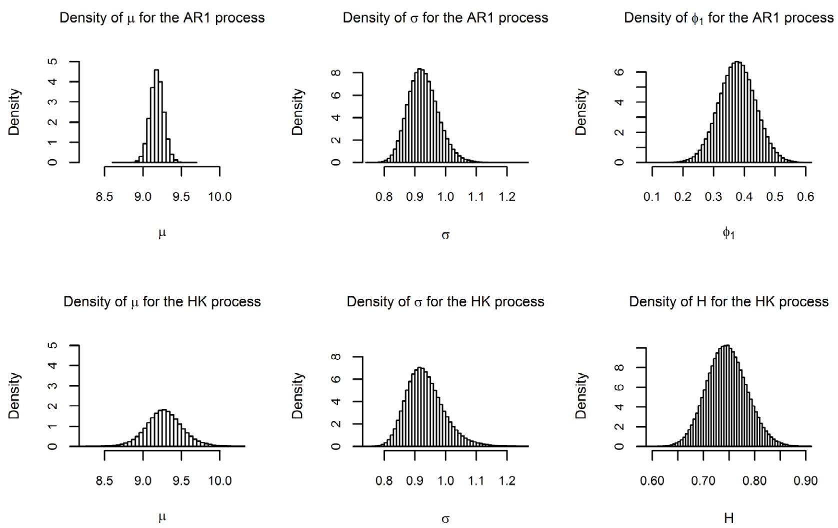


Posterior distributions for the parameters of Aliartos temperature

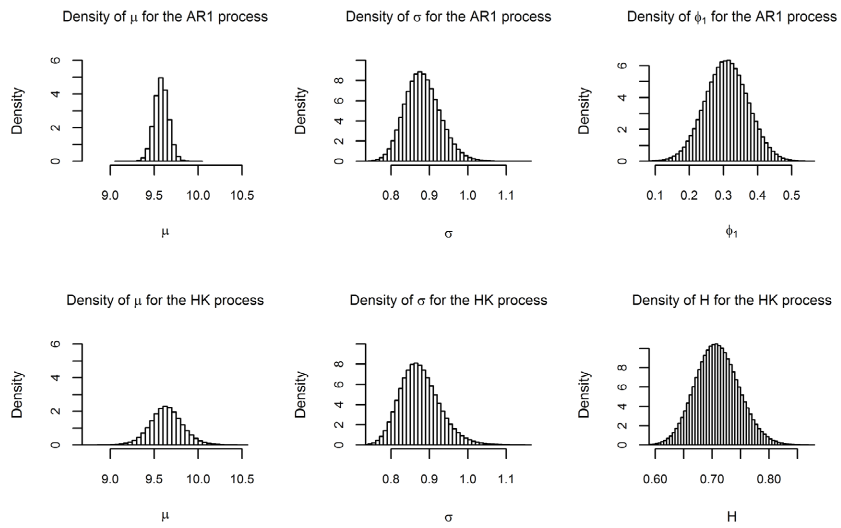


8. Posterior probability distributions for the AR(1) and HK parameters for the temperature at Berlin and Vienna

Posterior distributions for the parameters of Berlin temperature



Posterior distributions for the parameters of Vienna temperature



Case	Mean	Quantiles 2.5%	25%	50%	75%	97.5%
Berlin temperature						
μ	9.181	9.009	9.122	9.18	9.238	9.353
σ	0.9275	0.84	0.8934	0.9245	0.9583	1.0319
φ_1	0.3767	0.2601	0.3365	0.3766	0.4167	0.4937
μ	9.2790997	8.803	9.125	9.274	9.427	9.785
σ	0.935	0.8322	0.8914	0.9280	0.9703	1.0784
H	0.7459	0.6722	0.7192	0.7450	0.7717	0.8251
Vienna temperature						
μ	9.581	9.423	9.527	9.581	9.634	9.738
σ	0.8825	0.7995	0.8505	0.88	0.9117	0.9799
φ_1	0.309	0.1858	0.2666	0.3090	0.3514	0.4328
μ	9.6400404	9.269	9.52	9.637	9.757	10.028
σ	0.876	0.7863	0.8396	0.8716	0.9072	0.9918
H	0.7093	0.6374	0.6830	0.7082	0.7345	0.7876

RESULTS

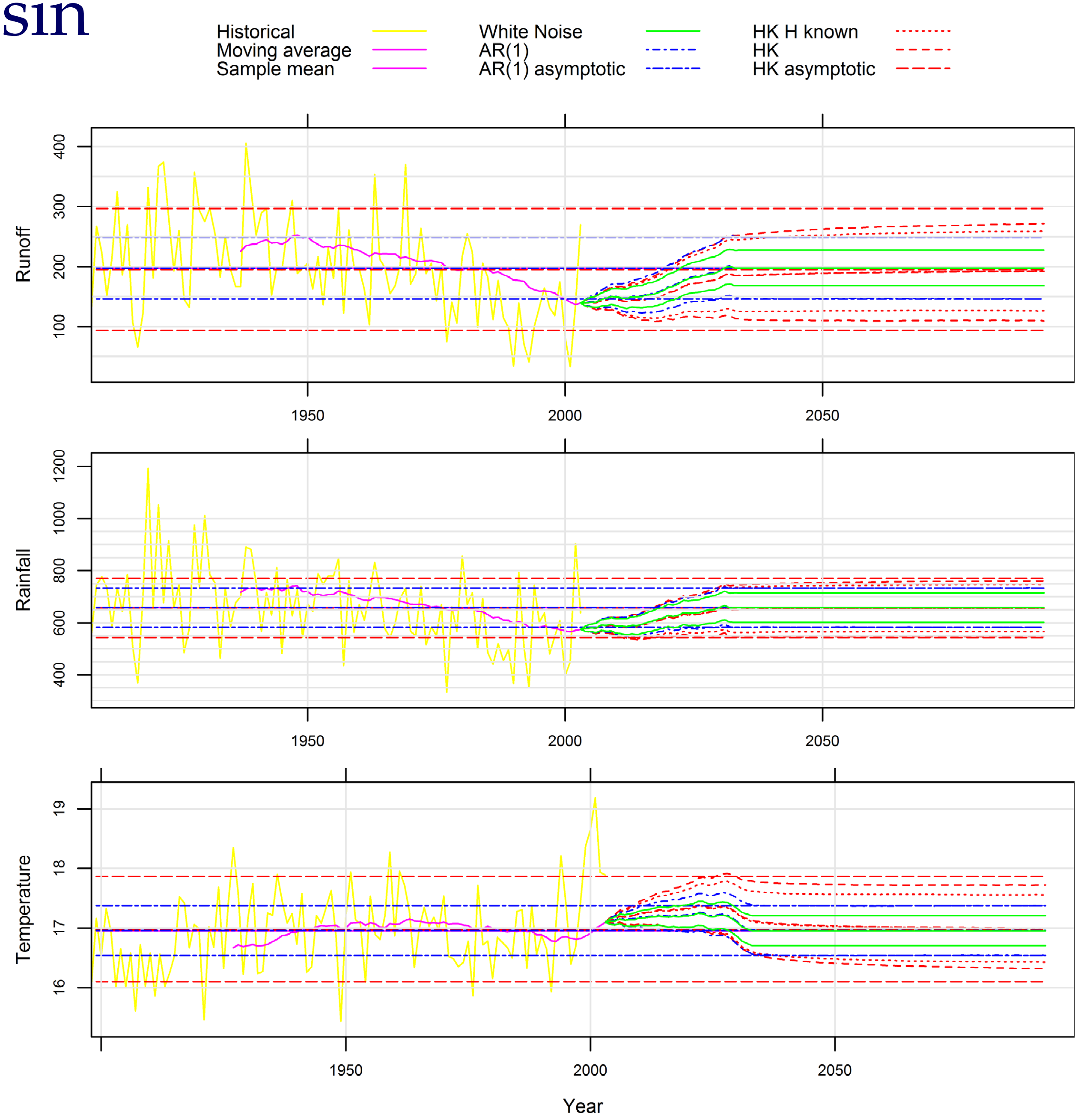
- Medians (50% quantiles) of μ are almost equal, irrespective of the model.
- Posterior medians of μ can be used as estimates of μ. (Robert, 2006[6])
- Distribution and confidence regions of μ are wider in the case of HK.
- Medians (50% quantiles) of σ are almost equal, irrespective of the model.
- Confidence regions of σ are wider in the case of HK, because the distribution is skewed to the right. As a result quantiles smaller than 50% are almost equal.
- Distributions of μ, H, φ_1 are almost symmetrical.

[6]Robert, C. (2007). “The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation”, Springer, New York.

9. Hydroclimatic prognosis for the Boeoticos Kephisos river basin

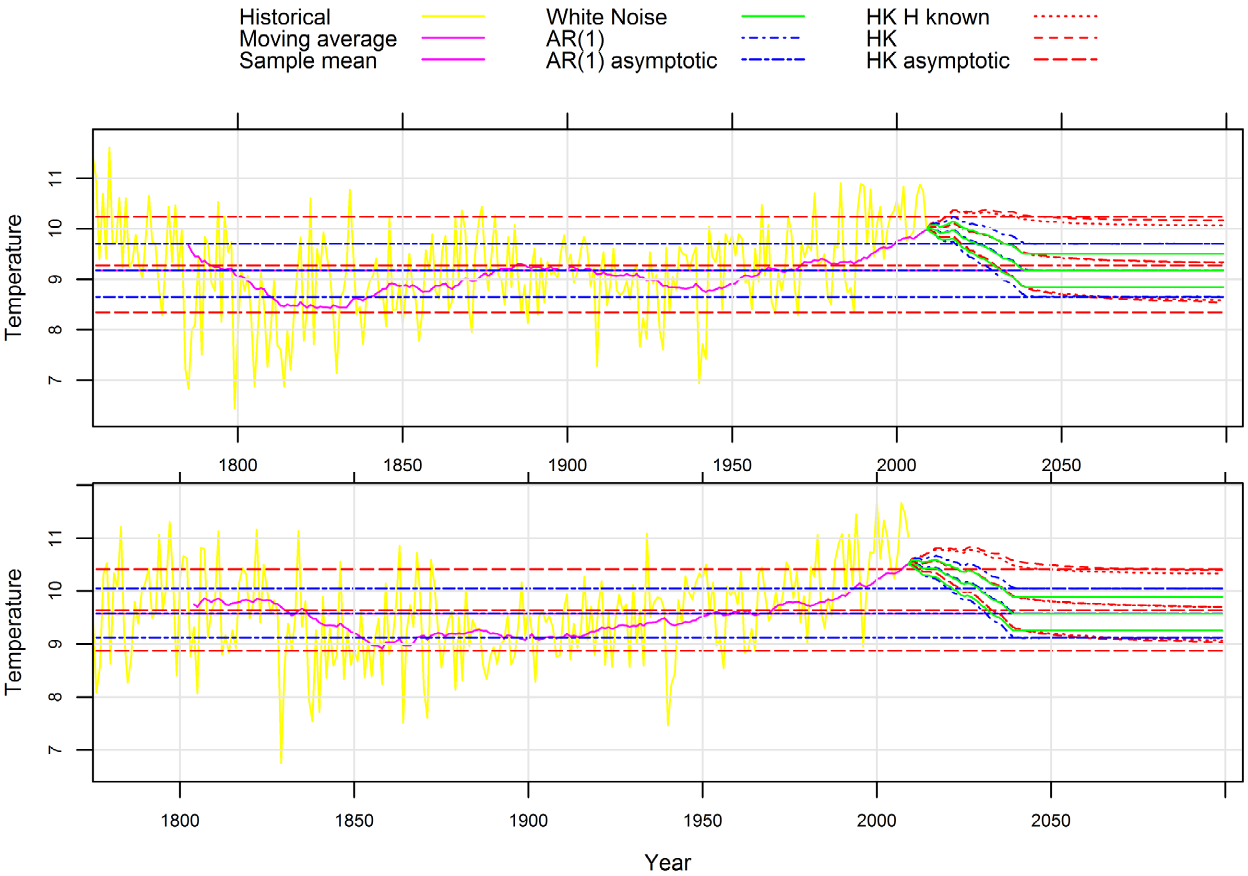
95% confidence regions for the 30-year moving average

The widest confidence regions correspond to the asymptotic HK case, due to the largest variability of the posterior distributions of μ and σ



10. Hydroclimatic prognosis for temperature at Berlin and Vienna

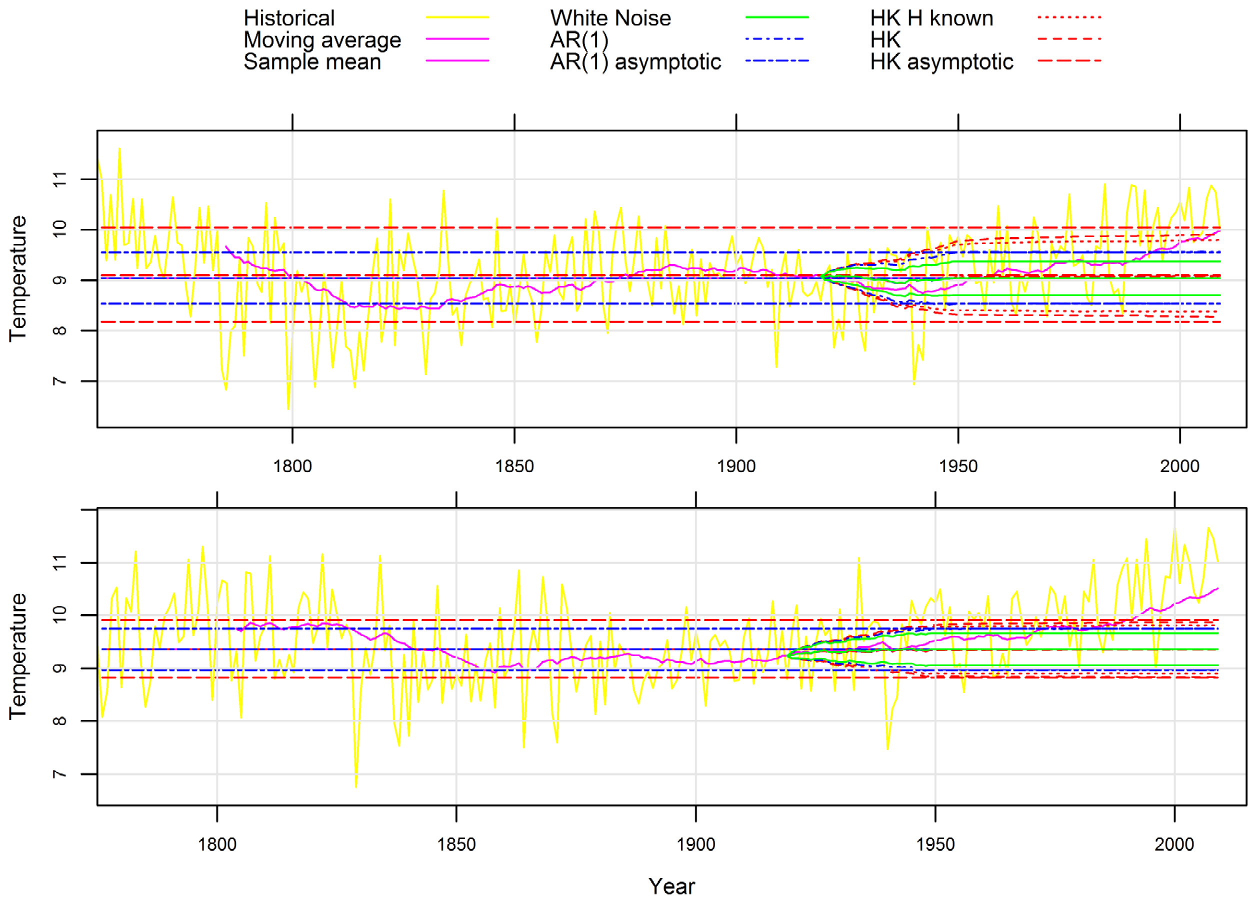
95% confidence regions for the 30-year moving average



11. Hydroclimatic prognosis for Berlin and Vienna, excluding historical data from last 90 years

95% confidence regions for the 30-year moving average

- Here these regions were calculated excluding the last 90 years from the data sets. These years were used for validation.
- In the case of Berlin it seems that when examining the asymptotic behavior of the HK, the model seems to behave well
- Additionally the widest confidence regions are almost equal to that derived from the full data set.
- In contrast, in the case of Vienna none of the models was able to catch the temperature increase that appeared in the last years. The smaller data set examined here gives smaller confidence regions. Notice that in frame 10 the confidence regions are wider.
- It is important to take account of the variability of the parameters. For example the case where *H* is considered known results in narrower confidence regions, compared to the case where *H* is considered to be unknown.



12. Conclusions

- Here we developed a Bayesian statistical methodology to make hydroclimatic prognosis in terms of estimating future confidence regions on the basis of a stationary stochastic process.
- We applied this methodology to five cases, namely the runoff, the rainfall and the temperature at Boeoticos Kephisos river basin, as well as the temperature at Berlin and the temperature at Vienna.
- We derived the posterior distributions of the parameters of the models. It turned out that when we took into account the Hurst-Kolmogorov behaviour of the examined process, the confidence regions of the parameters became wider.
- This resulted in a wider confidence region for the 30-year moving average, which represents a climatic variable.
- In all cases the HK model seemed to work well. WN and AR(1) did not seem to capture the variability.
- In one case, when we excluded the last 90 years of the data set of the Vienna temperature, it seemed that due to the increase of temperature in last decades, the model did not work well. But when we examined the full data set, the behaviour in last 90 years did not appear extraordinary.