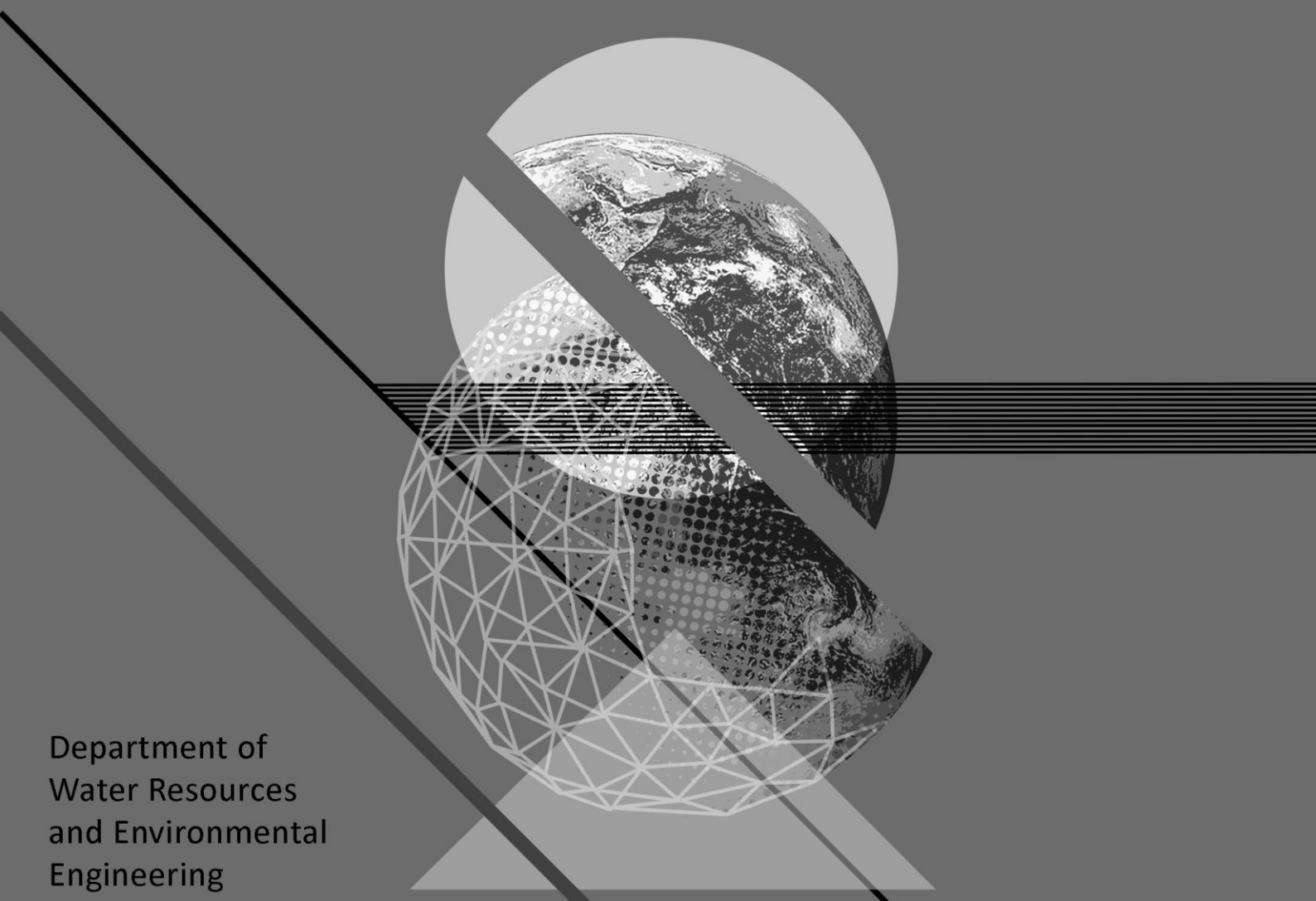
NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF CIVIL ENGINEERING

Use of Bayesian techniques in hydroclimatic prognosis

Hristos Tyralis



Department of
Water Resources
and Environmental
Engineering

Use of Bayesian techniques in hydroclimatic prognosis

Διατριβή

για την απόκτηση του τίτλου του διδάκτορος
από το Εθνικό Μετσόβιο Πολυτεχνείο,
κατά την θητεία των Πρυτάνεων Καθηγητή Σίμου Ε. Σιμόπουλου και Καθηγητή Ιωάννη Γκόλια
μετά από την δημόσια υπεράσπιση

την Παρασκευή, 6 Νοεμβρίου 2015 ώρα 11:00

από

τον Χρήστο Τύραλη

πολιτικό μηχανικό
γεννηθέντα στις Σέρρες

This thesis has been approved by the advisory committe

Professor Demetris Koutsoyiannis (Supervisor)	National Technical University of Athens
Reader Christian Onof	Imperial College London
Assistant Professor Nikolaos D. Lagaros	National Technical University of Athens

Examination committee composition

Professor Demetris Koutsoyiannis (Supervisor)	National Technical University of Athens
Reader Christian Onof	Imperial College London
Assistant Professor Nikolaos D. Lagaros	National Technical University of Athens
Professor Apostolos Burnetas	National and Kapodistrian University of Athens
Professor Alberto Montanari	University of Bologna
Associate Professor Harry Pavlopoulos	Athens University of Economics and Business
Assistant Professor Andreas Langousis	University of Patras

This research was performed at the Department of Water Resources and Environmental Engineering, School of Civil Engineering of National Technical University of Athens

Keywords: Bayesian statistics, general circulation models, Hurst-Kolmogorov, hydroclimatic prognosis, point estimation, posterior predictive distribution

This thesis was written using Microsoft Office
Printed in Athens

See itia.ntua.gr/en/docinfo/1581/ for more background information and updates.

To my parents

Rather than love, than money, than fame, give me truth.

- In Walden; or, Life in the Woods, by Henry David Thoreau (1854)

Preface

During the last decades numerous studies on the prediction of climatic variables have been published. The prediction problem is usually solved using state-of-the-art deterministic models which predict the climate for decades or centuries ahead. However the results of these deterministic models were not verified in the last decade.

A shift of paradigm from deterministic to stochastic approach is proposed. The stochastic prediction of geophysical variables is well established in hydrology, but rarely implemented by climate scientists. The initial aim of this research was to use well established stochastic techniques of hydrology in climate science, as well as to develop new statistical methods for the problem at hand. However the lack of knowledge on this direction, forced us to focus on the development of these tools and minimize the amount of numerical results.

The stochastic modelling of phenomena was based on well established notions of physics. It is assumed that processes exhibiting Hurst-Kolmogorov behaviour are appropriate to model geophysical variables. The author hopes that the models used in this study are correct, but an a priori verification is out of the scope of this thesis.

The simple Bayesian methodology (model choice → prior distribution → collection of data → posterior distribution → posterior predictive distribution) is mathematically strict and can quantify the uncertainty of the predictions, not just providing point estimates. Therefore Bayesian tools for the prediction of hydroclimatic processes assuming that they exhibit Hurst-Kolmogorov behaviour were developed here. Also statistical tools used in previous frequentist studies have also been verified.

Summary

Climatic prognosis is performed, using the deterministic General Circulation Models. These models whose use started half a century ago, are based on the Navier–Stokes equations and are numerical representations of the climate system based on the physical, chemical and biological properties of its components, their interactions and feedback processes. Recently it was proved that their older versions failed to provide adequate predictions, while newer versions are still not able to reproduce the climate of the past.

Thus a shift of paradigm to stochastic models has been proposed. Toy models have shown that stochastic models are more capable of predicting for long horizons, and additionally they can quantify the uncertainty on their predictions. We prefer to follow the path less travelled and model geophysical processes stochastically. Whereas a usual approach to stochastic modelling is the ad hoc choice of the appropriate stochastic model for the time series at hand, we again prefer to use results obtained from the implementation of general notions of physics, such as the maximization of entropy, albeit a satisfactory answer to the question, which is an appropriate stochastic model for climate has not yet been given. Maximization of entropy under certain constraints results in models exhibiting Hurst-Kolmogorov behaviour. In this thesis the Hurst-Kolmogorov stochastic process will be used to model this behaviour.

The overall aim of this thesis is the development of tools for climate prediction. Attempts to achieve this aim in a typical statistical context have been proved successful so far, but they do not offer much space for further improvements. A Bayesian approach offers more flexibility at the cost an additional assumption, i.e. the introduction of the prior distribution for the parameters of the models.

The main questions that are addressed in this thesis are:

- How can the uncertainty in the estimation of the parameters of the model be integrated in the uncertainty of the prediction?

- How can the data be used for the prediction?

- Which is an appropriate framework to gain information from the available deterministic models?

The main components of the framework that will be developed in this thesis are the stochastic model and the data. The development of tools should contribute in quantifying the uncertainty in the prediction of climate. Uncertainty quantification contrary to point estimation may explain climate variability.

To answer these questions, a previous typical statistical approach of the problem is investigated and is justified analytically. The general algorithm for the estimation of confidence intervals of parameters of interest that was used in this study is compared to other general algorithms and it is found that it performs satisfactorily. Properties of the algorithm are discovered within an analytical framework strengthening the arguments in favour of its use in this earlier study. However this approach is not adequate to solve the problem, owing to its indirect but encouraging results. Thus to continue the research, another path is chosen, namely the Bayesian approach.

To strengthen the Bayesian choice some results on the estimation of the parameters of the Hurst-Kolmogorov process using a maximum likelihood estimator are presented. A novel estimator is proposed as well and its properties are examined analytically. It is shown that handling all parameters of the process simultaneously is critical to obtain valid results. The posterior predictive distributions of the climate variables for the Hurst-Kolmogorov process are calculated conditional on past observations within a Bayesian stochastic framework. The examined variables are assumed to be normal or truncated normal. The results are compared with cases where some of the parameters are considered known and the effect of the uncertainty in their estimation is shown. Uncertainties not previously given attention are revealed.

We conclude trying to use information from deterministic model outputs to improve stochastic prediction. To this end properties of the maximum likelihood estimator of the bivariate Hurst-Kolmogorov process are analysed. A stochastic framework including both data and model outputs is developed.

On a more practical level the stochastic framework is applied to temperature, rainfall and runoff data from Greece and Europe and it is shown that it is able to explain climatic variability within a stationary context. The latter framework is applied to historical global temperature and over land precipitation data. General Circulation Models are used as deterministic models. It is shown that the information added by the General Circulation Models to that contained in the historical datasets is not substantial. This means that the output of the General Circulation Models has almost null effect on the stochastic predictions.

Περίληψη

Εισαγωγή

Στο άρθρο "A random walk on water" (Koutsoyiannis 2010) το οποίο εκπονήθηκε για την τελετή απονομής του βραβείου Henry Darcy Medal τίθεται το ερώτημα εάν η ντετερμινιστική πρόβλεψη του κλίματος είναι δυνατή και η απάντηση είναι κατηγορηματικά «όχι». Ο συγγραφέας επιχειρηματολογεί ότι «*είναι δυνατόν να σχηματοποιήσουμε μια συνεπή στοχαστική αναπαράσταση γεωφυσικών διεργασιών, στην οποία η δυνατότητα πρόβλεψης (η οποία παρέχεται από ντετερμινιστικούς κανόνες) και η αδυναμία (τυχαιότητα) συνυπάρχουν και δεν αποτελούν ξεχωριστά ή πρόσθετα στοιχεία το ένα του άλλου. Η απόφαση ποιό από τα δύο κυριαρχεί εξαρτάται απλά από τον χρονικό ορίζοντα και της κλίμακας της πρόβλεψης. Μακρινοί ορίζοντες πρόβλεψης αναπόφευκτα συσχετίζονται με υψηλή αβεβαιότητα, της οποίας η ποσοτικοποίηση εξαρτάται από τις ασυμπτωτικές στοχαστικές ιδιότητες των διεργασιών*».

Οι εργασίες σχετικά με την επιλογή ενός κατάλληλου στοχαστικού μοντέλου για την κλιματική πρόβλεψη είναι λίγες (Keenan 2014). Σε αυτή την εργασία επιλέχθηκε η προσέγγιση του Koutsoyiannis (2011) με την χρήση συγκεκριμένα της στοχαστικής ανέλιξης Hurst-Kolmogorov (HKp), όπως ονομάζεται από τον Koutsoyiannis (2010) και η οποία παρουσιάζει συμπεριφορά Hurst-Kolmogorov (HK). Στοχαστικά μοντέλα τα οποία παρουσιάζουν συμπεριφορά HK προκύπτουν από την εφαρμογή του δεύτερου νόμου της θερμοδυναμικής, δηλαδή την μεγιστοποίηση της εντροπίας, υπό ορισμένους περιορισμούς (Koutsoyiannis 2011). Η HKp είναι στάσιμη στοχαστική ανέλιξη, και η συμπεριφορά HK χαρακτηρίζεται από την τιμή της παραμέτρου Hurst H. Υποθέτουμε ότι η $\{\underline{x}_t\}$, $t = 1, 2, \ldots$ είναι μια στάσιμη στοχαστική ανέλιξη με μέση τιμή μ

$$\mu := \mathrm{E}[\underline{x}_t] \tag{1}$$

τυπική απόκλιση σ

$$\sigma := \sqrt{\mathrm{Var}[\underline{x}_t]} \tag{2}$$

συνάρτηση συνδιασποράς γ_k

$$\gamma_k := \mathrm{Cov}[\underline{x}_t, \underline{x}_{t+k}],\ k = 0, \pm1, \pm2, \ldots \tag{3}$$

και συνάρτηση αυτοσυσχέτισης (autocorrelation function, ACF) ρ_k

$$\rho_k := \gamma_k / \sigma,\ k = 0, \pm1, \pm2, \ldots \tag{4}$$

Τότε η συμπεριφορά HK μπορεί να μοντελοποιηθεί από την $\{\underline{x}_t\}$, εάν (Beran 1994 p.42)

$$\lim_{k\to\infty} \rho_k / (c\, k^{-a}) = 1,\ 0 < a < 1,\ 0 < c \tag{5}$$

Η παράμετρος H ορίζεται από την

$$H := 1 - a / 2 \tag{6}$$

Επιπλέον υποθέτουμε ότι η $\{\underline{x}_t\}$, $t = 1, 2, \ldots$ είναι κανονική. Έστω κ ένας θετικός ακέραιος ο οποίος αναπαριστά μια χρονική κλίμακα μεγαλύτερη από 1, δηλαδή την χρονική κλίμακα της ανέλιξης $\{\underline{x}_t\}$. Η μέση στοχαστική ανέλιξη σε αυτήν την κλίμακα δηλώνεται ως

$$\underline{x}_t^{(\kappa)} := (1/\kappa) \sum_{l=(t-1)\,\kappa+1}^{t\,\kappa} \underline{x}_l \tag{7}$$

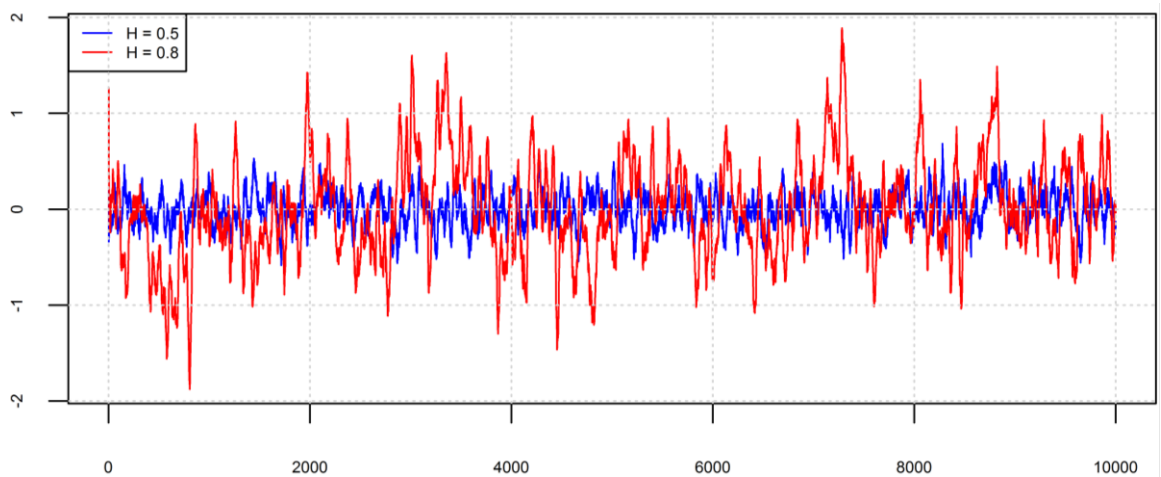
Η ακόλουθη εξίσωση ορίζει την HKp (Koutsoyiannis 2003).

$$(\underline{x}_i^{(\kappa)} - \mu) \overset{\mathrm{d}}{=} \left(\frac{\kappa}{\lambda}\right)^{H-1} (\underline{x}_j^{(\lambda)} - \mu),\ 0 < H < 1,\ \ i, j = 1, 2, \ldots \text{ and } \kappa, \lambda = 1, 2, \ldots \tag{8}$$

Η ACF της HKp είναι (Koutsoyiannis 2003)

$$\rho_k = |k+1|^{2H} / 2 + |k-1|^{2H} / 2 - |k|^{2H},\ k = 0, 1, \ldots \tag{9}$$

Μεγάλες τιμές τις παραμέτρου H είναι κατάλληλες για την μοντελοποίηση διεργασιών οι οποίες παρουσιάζουν μεγάλες διακυμάνσεις όπως στο Σχήμα 1 και μπορούν να χρησιμοποιηθούν αντί μη στάσιμων στοχαστικών μοντέλων των οποίων η χρήση πρέπει να αποφεύγεται, όπως προτείνεται από τον Koutsoyiannis (2006b).



Σχήμα 1. Κινούμενος μέσος για 30 σημεία για μια προσομοιωμένη HKp με $\mu = 0$, $\sigma = 1$ και (α) $H = 0.5$ και (β) $H = 0.8$.

Υποθέτοντας ότι το στοχαστικό μοντέλο είναι κατάλληλο το πρόβλημα της πρόβλεψης, δεδομένων των παρατηρήσεων, μπορεί να επιλυθεί σε ένα Μπεϋζιανό στατιστικό πλαίσιο. Σε

αυτό το πλαίσιο είναι δυνατή η εκτίμηση των παραμέτρων του μοντέλου, η εύρεση της κατανομής τους και η εύρεση της κατανομής μεταβλητών, όπως αυτών που χαρακτηρίζουν την μελλοντική εξέλιξη της στοχαστικής ανέλιξης. Ορίζουμε την τυχαία μεταβλητή $\underline{\boldsymbol{x}}_{1:n} := (\underline{x}_1 \ldots \underline{x}_n)^T$, όπου:

$$\underline{\boldsymbol{x}}_{1:n} \sim f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta}) \tag{10}$$

Το Μπεϋζιανό στατιστικό μοντέλο αποτελείται από τις $\boldsymbol{x}_{1:n}$, $\underline{\boldsymbol{x}}_{1:n}$ και π (Robert 2007 p.9). Η αβεβαιότητα της παραμέτρου $\boldsymbol{\theta}$ μοντελοποιείται από την κατανομή πιθανότητας π, η οποία ονομάζεται εκ των προτέρων κατανομή και ορίζεται ως εξής

$$\underline{\boldsymbol{\theta}} \sim \pi(\boldsymbol{\theta}) \tag{11}$$

Τότε η κατανομή της $\underline{\boldsymbol{\theta}}$ είναι η εξής:

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:n}) = \frac{f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \tag{12}$$

και η κατανομή μιας οποιασδήποτε (που μπορεί να είναι και μελλοντική) μεταβλητής $\underline{\boldsymbol{y}}$ είναι η εξής:

$$g(\boldsymbol{y}|\boldsymbol{x}_{1:n}) = \int g(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x}_{1:n})\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})\mathrm{d}\boldsymbol{\theta} \tag{13}$$

Η εργασία λοιπόν εστιάζει στην πρόβλεψη υδροκλιματικών μεταβλητών, όπως η θερμοκρασία και η βροχόπτωση, εντός ενός στοχαστικού πλαισίου. Η επίλυση αυτού του προβλήματος σε αυτό το πλαίσιο μπορεί να μην είναι η τελευταία λέξη της τεχνολογίας, ωστόσο είναι η μόνη βιώσιμη επιλογή για μια επαρκή απάντηση σε αυτό το πρόβλημα (Koutsoyiannis 2010). Μια βιώσιμη απάντηση στο ερώτημα, ποιό είναι το πλέον κατάλληλο στοχαστικό μοντέλο για την μελέτη του κλίματος δεν έχει δοθεί (Keenan 2014). Σε αυτήν την εργασία γίνεται η παραδοχή ότι οι γεωφυσικές διεργασίες παρουσιάζουν συμπεριφορά ΗΚ και μοντελοποιούνται αντίστοιχα. Όλη η πληροφορία της κλιματικής μεταβλητής περιλαμβάνεται στην κατανομή της. Είναι κρίσιμο λοιπόν να βρεθεί αυτή η κατανομή. Η πλέον πρακτική παραμετρική προσέγγιση του προβλήματος είναι η Μπεϋζιανή, η οποία μειώνει την πολυπλοκότητα του προβλήματος σε τεχνικά ζητήματα, με το κόστος μιας επιπλέον υπόθεσης, δηλαδή την παραδοχή μιας εκ των προτέρων κατανομής για την $\underline{\boldsymbol{\theta}}$.

Ένας αλγόριθμος για την κατασκευή Μόντε Κάρλο διαστημάτων εμπιστοσύνης για συναρτήσεις παραμέτρων πιθανοτικών κατανομών

Μια πρώτη προσέγγιση του προβλήματος από τους Koutsoyiannis et al. (2007) σε ένα στοχαστικό πλαίσιο έγινε με τυπικές στατιστικές μεθόδους. Αναπτύχθηκε διαισθητικά ένας

γενικός αλγόριθμος (Monte Carlo Confidence Interval, MCCI) χωρίς αναλυτική τεκμηρίωσή του και χρησιμοποιήθηκε για την εύρεση διαστημάτων εμπιστοσύνης για τις παραμέτρους και συναρτήσεις των παραμέτρων της HKp. Στη συνέχεια υπολογίστηκαν τα διαστήματα εμπιστοσύνης ποσοστημορίων της HKp. Τα διαστήματα εμπιστοσύνης χρησιμοποιήθηκαν στην εκτίμηση της αβεβαιότητας μελλοντικών υδροκλιματικών μεταβλητών. Το μέγεθος της αβεβαιότητας ηταν πολύ μεγαλύτερο, συγκρινόμενο με μοντέλα στα οποία γίνεται η παραδοχή ανεξαρτησίας ή Μαρκοβιανής σχέσης μεταξύ των μεταβλητών. Σε συγκεκριμένες περιπτώσεις αποδείχθηκε ότι το μοντέλο ήταν ικανό να αναπαραστήσει τις μεταβολές στις παρατηρημένες τιμές των γεωφυσικών μεταβλητών.

Στην παρούσα εργασία μελετήθηκε και επιβεβαιώθηκε αναλυτικά και αριθμητικά η αξία του αλγορίθμου, ο οποίος αποτελεί μια γενίκευση της μεθόδου του Ripley (1987 p.176-178). Ο αλγόριθμος εκτιμά ένα προσεγγιστικό $1 - \alpha$ διάστημα εμπιστοσύνης για την παράμετρο θ μιας μονοπαραμετρικής κατανομής $f(x|\theta)$ που, όπως προκύπτει και από το Σχήμα 2, δίνεται από την επόμενη σχέση:

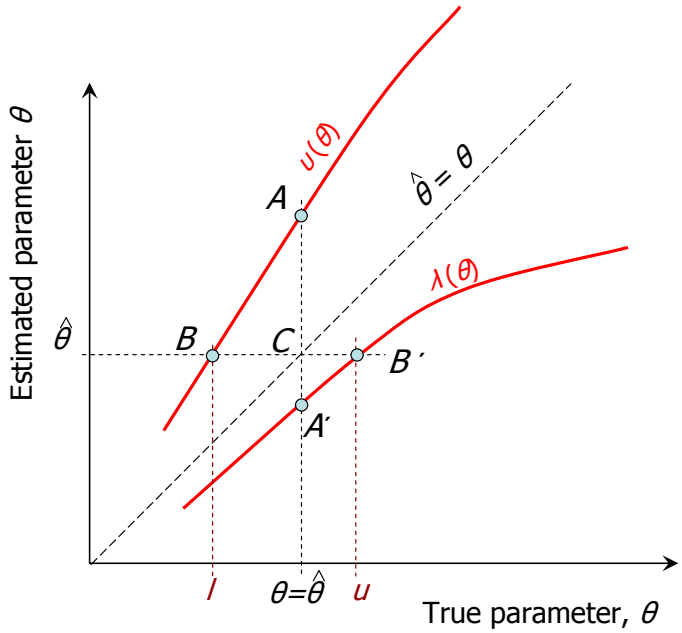
$$[l(\underline{\boldsymbol{x}}),u(\underline{\boldsymbol{x}})] = [b(\underline{\boldsymbol{x}}) + \frac{b(\underline{\boldsymbol{x}}) - \upsilon(b(\underline{\boldsymbol{x}}))}{(\mathrm{d}\upsilon/\mathrm{d}\theta)_{|\theta = b(\underline{\boldsymbol{x}})}}, b(\underline{\boldsymbol{x}}) + \frac{b(\underline{\boldsymbol{x}}) - \lambda(b(\underline{\boldsymbol{x}}))}{(\mathrm{d}\lambda/\mathrm{d}\theta)_{|\theta = b(\underline{\boldsymbol{x}})}}] \tag{14}$$

όπου $\underline{b} := b(\underline{\boldsymbol{x}}_{1:n})$ είναι η εκτιμήτρια μέγιστης πιθανοφάνειας (ΕΜΠ) της παραμέτρου θ και $\hat{\theta} = b(\boldsymbol{x}_{1:n})$ η εκτίμηση της παραμέτρου. Υποθέτοντας ότι η κατανομή του στατιστικού $b(\underline{\boldsymbol{x}}_{1:n})$ είναι $g(b|\theta)$, τότε οι συναρτήσεις $\lambda(\theta)$, $\upsilon(\theta)$ ορίζονται ως εξής:

$$\lambda(\theta) = G^{-1}(\alpha/2|\theta) \text{ and } \upsilon(\theta) = G^{-1}(1 - \alpha/2|\theta) \tag{15}$$

Η $G^{-1}(\cdot|\theta)$ δηλώνει την αντίστροφη της συνάρτηση της συνάρτησης κατανομής G. Στην σχέση (14) οι άγνωστοι $(\mathrm{d}\upsilon/\mathrm{d}\theta)_{|\theta = b(\underline{\boldsymbol{x}})}$ και $(\mathrm{d}\lambda/\mathrm{d}\theta)_{|\theta = b(\underline{\boldsymbol{x}})}$ υπολογίζονται μετά από προσομοιώσεις Μόντε Κάρλο. Αποδείχθηκε αναλυτικά ότι αυτός ο αλγόριθμος είναι ακριβής γα οικογένειες κατανομών θέσης και κλίμακας. Επιπλέον το διάστημα εμπιστοσύνης της σχέσης (14) αποδείχθηκε ότι είναι ασυμπτωτικά ισοδύναμο με ένα διάστημα τύπου Wald για οποιαδήποτε συνάρτηση του θ και συνεπώς και για το ίδιο το θ.

Αποδείχτηκε αναλυτικά ότι μια διαισθητική τροποποίηση του αλγορίθμου για πολυπαραμετρικές κατανομές αποδίδει επίσης διαστήματα εμπιστοσύνης ασυμπτωτικά ισοδύναμα με διαστήματα τύπου Wald. Ο αλγόριθμος εφαρμόστηκε στην εκθετική κατανομή, την κανονική κατανομή, την κατανομή Gamma και την κατανομή Weibull και συγκρίθηκε με άλλους γενικούς αλγορίθμους. Από τα αποτελέσματα του Πίνακα 1, φαίνεται ότι υπερείχε.



Σχήμα 2. Σκίτσο που εξηγεί τον προσδιορισμό των ορίων εμπιστοσύνης *l* και *u* από την αντιστροφή του έλεγχου υπόθεσης.

Πίνακας 1. Μόντε Κάρλο πιθανότητες κάλυψης και κατάταξη κάθε μεθόδου για τον υπολογισμό διαστημάτων εμπιστοσύνης 0.975 μετά από 10 000 επαναλήψεις (η κατάταξη 1 αποδίδεται στην μέθοδο με την καλύτερη απόδοση.

							Πιθανότητες κάλυψης (με κατατάξεις σε παρενθέσεις) για όλες τις μεθόδους				
Περίπτωση	Κατανομή	Παράμετρος	Μέγεθος δείγματος	Τιμή παραμέτρου	Τιμή παραμέτρου	Προσεγγιστικό	Ripley θέσης	Ripley κλίμακας	Τύπου Wald	Bootstrap	MCCI
1	Εκθετική	Κλίμακα	10	$\sigma = 2$			0.889 (5)	0.977 (2)	0.975 (1)	0.916 (4)	0.966 (3)
2	Κανονική	Θέση	10	$\mu = 0$	$\sigma = 1$		0.946 (3)	0.946 (3)	0.947 (2)	0.931 (5)	0.968 (1)
3	Κανονική	Ποσοστημόριο	10	$\mu = 0$	$\sigma = 1$		0.919 (4)	0.929 (2)	0.929 (2)	0.867 (5)	0.973 (1)
4	Gamma	Κλίμακα	50	$\alpha = 2$	$\sigma = 3$	0.753	0.923 (5)	0.976 (1)	0.940 (4)	0.957 (3)	0.974 (1)
5	Gamma	Μορφή	50	$\alpha = 2$	$\sigma = 3$	0.976	0.948 (5)	0.972 (2)	0.978 (2)	0.956 (4)	0.974 (1)
6	Weibull	Κλίμακα	50	$a = 2$	$b = 3$	0.971	0.969 (3)	0.970 (2)	0.966 (4)	0.965 (5)	0.973 (1)
7	Weibull	Ποσοστημόριο	50	$a = 2$	$b = 3$	0.971	0.968 (3)	0.970 (1)		0.961 (4)	0.969 (2)
	μέση κατάταξη						4.000	1.857	2.500	4.286	1.429

Ταυτόχρονη εκτίμηση των παραμέτρων της στοχαστικής ανέλιξης Hurst-Kolmogorov

Αποφασιστική σημασία στην ανάλυση των γεωφυσικών διεργασιών έχει η εκτίμηση της ισχύος της συμπεριφοράς ΗΚ η οποία εκφράζεται από την τιμή της παραμέτρου *H*. Πολλές εκτιμήτριες του *H* έχουν προταθεί. Αυτές οι εκτιμήτριες συνήθως δεν εκτιμούν ταυτόχρονα και τις άλλες παραμέτρους της ΗΚp. Εδώ αποδεικνύεται ότι η εκτίμηση του *H* επηρεάζει την εκτίμηση της τυπικής απόκλισης, ένα γεγονός στο οποίο δεν έχει δοθεί προσοχή στην βιβλιογραφία. Προτείνουμε την βασισμένη στην μέθοδο ελαχίστων τετραγώνων για την διασπορά (Least Squares based on Variance estimator, LSSV) και διερευνούμε αριθμητικά την απόδοσή της, την οποία συγκρίνουμε με την μέθοδο ελαχίστων τετραγώνων βασισμένη στην τυπική απόκλιση (Least Squares based on Standard Deviation, LSSD) και την

εκτιμήτρια της μέγιστης πιθανοφάνειας. Οι τρεις αυτές εκτιμήτριες βασίζονται στην δομή της HKp και εκτιμούν συγχρόνως την παράμετρο H και την τυπική διασπορά. Επιπλέον ελέγχουμε την απόδοση των τριών μεθόδων για ένα εύρος μεγεθών του δείγματος και τιμών του H με μια μελέτη προσομοίωσης και τις συγκρίνουμε με άλλες εκτιμήτριες της βιβλιογραφίας.

Υποθέτουμε ότι η $\{\underline{x}_t\}$, $t = 1, 2, \ldots$ είναι μια HKp. Ορίζουμε επίσης την συναθροισμένη στοχαστική ανέλιξη για κάθε χρονική κλίμακα:

$$\underline{z}_t^{(\kappa)} := \sum_{l=(t-1)\,\kappa+1}^{t\,\kappa} \underline{x}_l = \kappa\, \underline{x}_t^{(\kappa)} \tag{16}$$

Για αυτήν την στοχαστική ανέλιξη ισχύει:

$$\mathrm{E}[\underline{z}_t^{(\kappa)}] = \kappa\,\mu,\ \gamma_0^{(\kappa)} = \mathrm{Var}[\underline{z}_t^{(\kappa)}] = \kappa^{2\cdot H}\,\gamma_0,\ \sigma^{(\kappa)} = (\gamma_0^{(\kappa)})^{1/2} \tag{17}$$

Η συνάρτηση συνδιασποράς για οποιοδήποτε από τα $\underline{x}_t^{(\kappa)}$ και $\underline{z}_t^{(\kappa)}$, και για οποιαδήποτε χρονική κλίμακα συνάθροισης κ, είναι ανεξάρτητη του κ, και δίνεται από

$$\rho_k^{(\kappa)} = \rho_k = |k+1|^{2H}/2 + |k-1|^{2H}/2 - |k|^{2H},\ k = 0, 1, \ldots \tag{18}$$

Για μια παρατήρηση $\boldsymbol{x}_{1:n}$ η πιθανοφάνεια του $\boldsymbol{\theta} := (\mu, \sigma, H)$ παίρνει την μορφή (McLeod and Hippel 1978):

$$l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n}) = \frac{1}{(2\pi)^{n/2}}\ |\sigma^2\, \boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \exp[-1/(2\sigma^2)\ (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)] \tag{19}$$

όπου

$$\boldsymbol{e}_n = (1,1, \ldots,1)^{\mathrm{T}} \tag{20}$$

είναι ένα διάνυσμα στήλη με n στοιχεία, $\boldsymbol{R}_{[1:n]\,[1:n]}$ είναι ο πίνακας αυτοσυσχετίσεων, δηλαδή ένας πίνακας n x n με στοιχεία $r_{ij} = \rho_{|i-j|}$, και $|\cdot|$ δηλώνει την διακρίνουσα του πίνακα.

Η εκτιμήτρια μέγιστης πιθανοφάνειας $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{H})$ αποτελείται από τις ακόλουθες σχέσεις:

$$\hat{\mu} = \frac{\boldsymbol{x}_{1:n}^{\mathrm{T}}\, \hat{\boldsymbol{R}}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n}{\boldsymbol{e}_n^{\mathrm{T}}\, \hat{\boldsymbol{R}}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n}, \tag{21}$$

$$\hat{\sigma} = \sqrt{\frac{(\boldsymbol{x}_{1:n} - \hat{\mu}\, \boldsymbol{e}_n)^{\mathrm{T}}\, \hat{\boldsymbol{R}}^{-1}_{[1:n]\,[1:n]}\, (\boldsymbol{x}_{1:n} - \hat{\mu}\, \boldsymbol{e}_n)}{n}} \tag{22}$$

και το $\hat{H}$ εκτιμάται από την μεγιστοποίηση της συνάρτησης $g_1(H)$, η οποία ορίζεται ως εξής:

$$g_1(H) := -\frac{n}{2}\ln[(\boldsymbol{x}_{1:n} - \frac{\boldsymbol{x}_{1:n}^{\mathrm{T}}\hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\boldsymbol{e}_n}{\boldsymbol{e}_n^{\mathrm{T}}\hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\boldsymbol{e}_n}\boldsymbol{e}_n)^{\mathrm{T}} \cdot$$

$$\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}(\boldsymbol{x}_{1:n} - \frac{\boldsymbol{x}_{1:n}^{\mathrm{T}}\hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\boldsymbol{e}_n}{\boldsymbol{e}_n^{\mathrm{T}}\hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\boldsymbol{e}_n}\boldsymbol{e}_n)] - \frac{1}{2}\ln(|\boldsymbol{R}_{[1:n]\,[1:n]}|) \tag{23}$$

Οι εκτιμήτριες LSSD (Koutsoyiannis 2003) και LSV βασίζονται στην ίδια λογική με την διαφορά ότι η μεταγενέστερη LSV τεκμηριώνεται από αναλυτικές μεθόδους, σε αντίθεση με την LSSD για την τεκμηρίωση της οποίας έγινε μια μελέτη προσομοίωσης, διότι δεν ήταν δυνατή η αναλυτική τεκμηρίωσή της. Αποδεικνύεται ότι (δες επίσης Beran 1994 p.9):

$$\mathrm{E}[\underline{s}_n^{2(\kappa)}] = \frac{(n/\kappa) - (n/\kappa)^{2H-1}}{(n/\kappa) - 1}\,\gamma_0^{(\kappa)} = \frac{(n/\kappa) - (n/\kappa)^{2H-1}}{(n/\kappa) - 1}\,\kappa^{2H}\,\sigma^2 = \mathrm{c}_\kappa(H)\,\kappa^{2H}\,\sigma^2 \tag{24}$$

όπου

$$\underline{s}_n := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\underline{x}_i - \underline{x}_1^{(n)})^2} \tag{25}$$

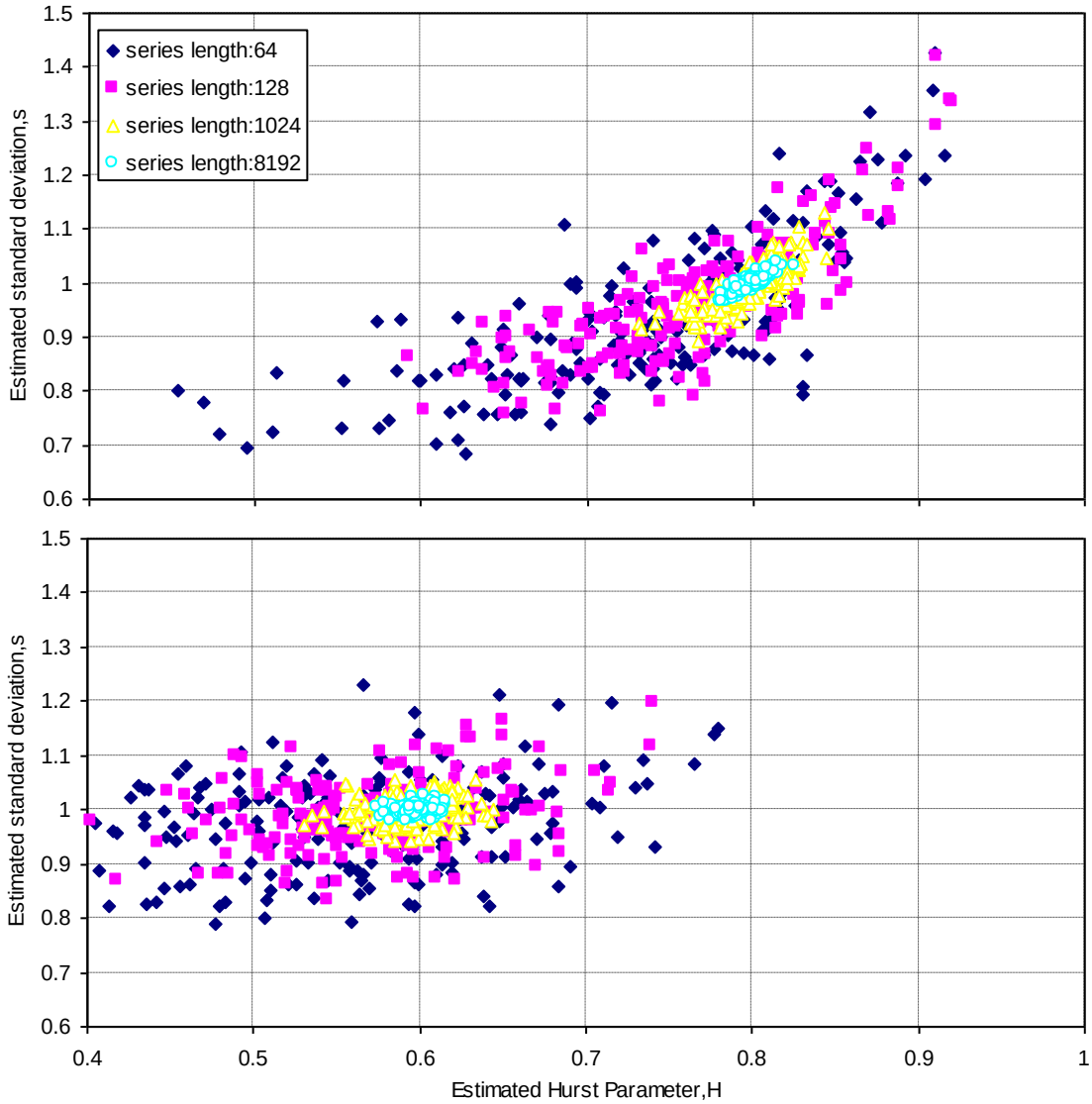
$$\mathrm{c}_\kappa(H) := \frac{(n/\kappa) - (n/\kappa)^{2H-1}}{(n/\kappa) - 1} \text{ and } s_n^{2(\kappa)} = \frac{1}{n/\kappa - 1}\sum_{i=1}^{n/\kappa}(z_i^{(\kappa)} - k\,x_1^{(n)})^2. \tag{26}$$

Η LSV μέθοδος εκτιμά τα H και σ ελαχιστοποιώντας την ακόλουθη συνάρτηση

$$\mathrm{er}^2(\sigma, H) := \sum_{\kappa=1}^{\kappa'}\frac{[\mathrm{E}[\underline{s}_n^{2(\kappa)}] - s_n^{2(\kappa)}]^2}{\kappa^p} = \sum_{\kappa=1}^{\kappa'}\frac{[\mathrm{c}_\kappa(H)\,\kappa^{2H}\,\sigma^2 - s_n^{2(\kappa)}]^2}{\kappa^p},\ \kappa' = [n/10] \tag{27}$$

Αποδεικνύεται μετά από μελέτη προσομοίωσης ότι οι παραπάνω μέθοδοι παρουσιάζουν πολύ καλύτερες ιδιότητες σε σχέση με τις άλλες μεθόδους της βιβλιογραφίας. Επιπλέον οι εκτιμώμενες τιμές του H είναι εντός του πεδίου ορισμού του. Αποδεικνύεται επιπλέον μετά τον υπολογισμό του πίνακα πληροφορίας Fisher ότι τα σ και H δεν είναι ορθογώνια, όπως φαίνεται και στο Σχήμα 3. Οι Cox and Reid (1987) περιγράφουν συνοπτικά έναν αριθμό των στατιστικών συνεπειών της ορθογωνικότητας. Μια μη ταυτόχρονη εκτίμηση των σ και H ίσως να είναι υποδεέστερη σε ότι αφορά την ευστάθεια, συγκρίνοντας με τις μεθόδους ML, LSSD και LSV οι οποίες εκτιμούν ταυτόχρονα τα σ και H. Από πρακτική άποψη, η σημασία της εξάρτησης των εκτιμητριών, μπορεί να κατανοηθεί από τις πολυάριθμες δημοσιεύσεις στις οποίες το σ εκτιμάται από την (25), ενώ την ίδια στιγμή εκτιμούν $H > 0.5$ και μερικές

φορές κοντά στο 1. Προφανώς τέτοιες εκτιμήσεις είναι αρκετά μεροληπτικές.



Σχήμα 3. Εκτιμώμενη παράμετρος Hurst H έναντι εκτιμώμενης τυπικής απόκλισης σ από την μέθοδο μέγιστης πιθανοφάνειας από 200 δείγματα συνθετικών χρονοσειρών με διάφορα μήκη. Το πάνω διάγραμμα αντιστοιχεί σε πραγματικό H = 0.8 και το κάτω διάγραμμα σε πραγματικό H = 0.6.

Η προβλεπτική κατανομή υδροκλιματικών μεταβλητών

Η HKp έχει μεγάλες αυτοσυσχετίσεις ακόμη και για μεγάλες χρονικές αποστάσεις, όπως και μεγάλη μεταβλητότητα σε όλες τις χρονικές κλίμακες. Το πρόβλημα λοιπόν είναι πως θα ενσωματώσουμε τις υδροκλιματικές παρατηρήσεις για να παράγουμε την προβλεπτική κατανομή των υδροκλιματικών διεργασιών σε κλιματικές χρονικές κλίμακες. Με την χρήση Μπεϋζιανών τεχνικών δημιουργούμε ένα πλαίσιο για να λύσουμε αυτό το πρόβλημα.

Υποθέτουμε ότι η $\{\underline{x}_t\}$, t = 1, 2, … είναι μια κανονική στάσιμη στοχαστική ανέλιξη με παραμέτρους που δίνονται από τις (1)-(4). Έστω ότι η $\{\underline{a}_t\}$ είναι ένας λευκός θόρυβος (White

Noise, WN), δηλαδή μια ακολουθία ανεξάρτητων τυχαίων μεταβλητών από μια κανονική κατανομή με μέσο $\mathrm{E}[\underline{a}_t] = 0$ και διασπορά $\mathrm{Var}[\underline{a}_t] = \sigma_a^2$. Από εδώ και στο εξής η $\{\underline{a}_t\}$ θα αναφέρεται ως WN. Η επόμενη εξίσωση ορίζει το μοντέλο αυτοπαλινδρόμησης τάξης 1 (first-order autoregressive process, AR(1)).

$$\underline{x}_t - \mu = \varphi_1(\underline{x}_{t-1} - \mu) + \underline{a}_t,\ |\varphi_1| < 1 \tag{28}$$

Η ACF του AR(1) είναι (Wei 2006 p.34)

$$\rho_k = \varphi_1^k,\ k = 0, 1, \ldots \tag{29}$$

Η κατανομή της μεταβλητής $\underline{\boldsymbol{x}}_{1:n} = (\underline{x}_1, \ldots, \underline{x}_n)^{\mathrm{T}}$ από μια κανονική στάσιμη στοχαστική ανέλιξη είναι:

$$f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta}) = (2\pi)^{-n/2}\, |\sigma^2\, \boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \exp[(-1/2\sigma^2)\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)] \tag{30}$$

όπου $\boldsymbol{R}_{[1:n]\ [1:n]}$ είναι ο πίνακας αυτοσυσχετίσεων με στοιχεία $r_{ij} = \rho_{|i-j|}$, $i,j = 1,2, \ldots,n$. Η αυτοσυσχέτιση $\rho_{|i-j|}$ έστω ότι είναι συνάρτηση της παραμέτρου $\boldsymbol{\varphi}$, τέτοιας ώστε $\boldsymbol{\theta} := (\mu, \sigma^2, \boldsymbol{\varphi})$ να είναι η παράμετρος της ανέλιξης. Επισημαίνεται ότι εάν η $\underline{\boldsymbol{x}}_n$ είναι WN, τότε $\rho_0 = 1$ και $\rho_k = 0$, $k = 1, 2, \ldots$, εάν είναι AR(1) τότε το ρ_k δίνεται από την (29) και αν είναι HKp τότε το ρ_k δίνεται από την (4).

Έστω ότι το $\boldsymbol{\varphi}$ είναι εκ των προτέρων κατανεμημένο ομοιόμορφα. Θέτουμε ως εκ των προτέρων κατανομή για το $\underline{\boldsymbol{\theta}}$ την κατανομή μηδενικής πληροφορίας (δες επίσης Robert 2007 example 3.5.6)

$$\pi(\boldsymbol{\theta}) \propto 1/\sigma^2 \tag{31}$$

Η εκ των υστέρων κατανομή των παραμέτρων δεν έχει αναλυτική μορφή. Ωστόσο μπορεί να υπολογιστεί από ένα μείγμα βασισμένο στις δεσμευμένες κατανομές. Συγκεκριμένα αποδεικνύεται ότι:

$$\underline{\mu}|\sigma^2, \boldsymbol{\varphi}, \boldsymbol{x}_{1:n} \sim N[(\boldsymbol{x}^{\mathrm{T}}_{1:n}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n)/(\boldsymbol{e}^{\mathrm{T}}_n\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n),\ \sigma^2/(\boldsymbol{e}^{\mathrm{T}}_n\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n)] \tag{32}$$

$$\underline{\sigma}^2|\boldsymbol{\varphi}, \boldsymbol{x}_{1:n} \sim \text{Inv-gamma}\{(n-1)/2,\ [\boldsymbol{e}^{\mathrm{T}}_n\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]} \cdot \boldsymbol{e}_n\, \boldsymbol{x}^{\mathrm{T}}_{1:n}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{x}_{1:n} - (\boldsymbol{x}^{\mathrm{T}}_{1:n}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n)^2]/(2\, \boldsymbol{e}^{\mathrm{T}}_n\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n)\} \tag{33}$$

$$\pi(\boldsymbol{\varphi}|\boldsymbol{x}_{1:n}) \propto |\boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2}\, [\boldsymbol{e}^{\mathrm{T}}_n\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n \cdot \boldsymbol{x}^{\mathrm{T}}_{1:n}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{x}_{1:n} - (\boldsymbol{x}^{\mathrm{T}}_{1:n}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n)^2]^{-(n-1)/2}\, (\boldsymbol{e}^{\mathrm{T}}_n\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n)^{n/2-1} \tag{34}$$

Σε πραγματικά προβλήματα επιβάλλονται άνω ή κάτω φράγματα για τις μεταβλητές $\underline{x}_t$. Έστω ότι η κατανομή της $\underline{\boldsymbol{x}}_{1:n}$ είναι φραγμένη και από τις δύο πλευρές με φράγματα a και b,

δηλαδή,

$$f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta}) \propto \exp[(-1/2\sigma^2)\,(\boldsymbol{x}_{1:n} - \mu\,\boldsymbol{e}_n)^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,(\boldsymbol{x}_{1:n} - \mu\,\boldsymbol{e}_n)]\;\mathrm{I}_{[a,b]^n}(x_1, \ldots, x_n) \qquad (35)$$

όπου το I δηλώνει την συνάρτηση δείκτη, τέτοια ώστε $\mathrm{I}_{[a,b]^n}(x_1, \ldots, x_n) = 1$ εάν $\boldsymbol{x}_n \in [a,b]^n$ και 0 αλλού. Έστω ότι το σύνολο που φράσσει το μ είναι $[a,b]$, $a,b \in \mathrm{R}\cup\{-\infty,\infty\}$. Ο ακόλουθος αλγόριθμος Gibbs αποδεικνύεται ότι μας παρέχει εκ των υστέρων δείγμα προσομοίωσης του $\underline{\boldsymbol{\theta}} = (\underline{\mu}, \underline{\sigma^2}, \underline{\boldsymbol{\varphi}})$.

$$\pi(\mu|\sigma^2, \boldsymbol{\varphi}, \boldsymbol{x}_{1:n}) \propto \exp\{-[\mu - (\boldsymbol{x}^{\mathrm{T}}_{1:n}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n)/(\boldsymbol{e}^{\mathrm{T}}_n \cdot \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n)]^2/(2\sigma^2/\boldsymbol{e}^{\mathrm{T}}_n\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n)\}\;\mathrm{I}_{[a,b]}(\mu) \qquad (36)$$

$$\underline{\sigma^2}|\mu, \boldsymbol{\varphi}, \boldsymbol{x}_{1:n} \sim \text{Inv-gamma}\{n/2, (\boldsymbol{x}_{1:n} - \mu\,\boldsymbol{e}_n)^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,(\boldsymbol{x}_{1:n} - \mu\,\boldsymbol{e}_n)/2\} \qquad (37)$$

$$\pi(\boldsymbol{\varphi}|\,\mu, \sigma^2, \boldsymbol{x}_{1:n}) \propto |\boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \exp[-(\boldsymbol{x}_{1:n} - \mu\,\boldsymbol{e}_n)^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,(\boldsymbol{x}_{1:n} - \mu\,\boldsymbol{e}_n)/2\sigma^2] \qquad (38)$$

Η εκ των υστέρων προβλεπτική κατανομή του $\underline{\boldsymbol{x}}_{(n+1):(n+m)}$ δεσμευμένου των $\boldsymbol{\theta}$ και $\boldsymbol{x}_{1:n}$ αποδεικνύεται ότι είναι (Eaton 2007 p.116,117)

$$f(\boldsymbol{x}_{(n+1):(n+m)}|\boldsymbol{\theta},\boldsymbol{x}_{1:n}) = (2\pi\sigma^2)^{-m/2}\,|\boldsymbol{R}_{m|n}|^{-1/2} \cdot \exp[(-1/2\sigma^2)\,(\boldsymbol{x}_{(n+1):(n+m)} - \boldsymbol{\mu}_{m|n})^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{m|n}\,(\boldsymbol{x}_{(n+1):(n+m)} - \boldsymbol{\mu}_{m|n})] \qquad (39)$$

όπου $\boldsymbol{\mu}_{m|n}$ τα $\boldsymbol{R}_{m|n}$ δίνονται από τα:

$$\boldsymbol{\mu}_{m|n} = \mu\boldsymbol{e}_m + \boldsymbol{R}_{[(n+1):(n+m)]\,[1:n]}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,(\boldsymbol{x}_{1:n} - \mu\boldsymbol{e}_n) \qquad (40)$$

$$\boldsymbol{R}_{m|n} = \boldsymbol{R}_{[(n+1):(n+m)]\,[(n+1):(n+m)]} - \boldsymbol{R}^{\mathrm{T}}_{[1:n]\,[(n+1):(n+m)]}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{R}_{[1:n]\,[(n+1):(n+m)]} \qquad (41)$$

όπου το $\boldsymbol{R}_{[k:l]\,[m:n]}$ είναι ένας υποπίνακας του $\boldsymbol{R}$, ο οποίος περιέχει τα στοιχεία r_{ij}, $k \le i \le l$, $m \le j \le n$. τα οποία υπολογίζονται ανάλογα με τις ιδιότητες της στοχαστικής ανέλιξης $\{\underline{x}_t\}$. Οι ανωτέρω σχέσεις μπορούν να χρησιμοποιηθούν για οποιαδήποτε συνάρτηση αυτοσυσχέτισης, οποιασδήποτε στάσιμης κανονικής στοχαστικής ανέλιξης, όπως ο WN, η AR(1) και η HKp. Επιπρόσθετα είναι δυνατή η εύρεση της κατανομής της μεταβλητής $\underline{\boldsymbol{x}}_{(n+m+1):(n+m+l)} := (\underline{x}_{n+m+1},\ldots,\underline{x}_{n+m+l})$ για $m \to \infty$, δεσμευμένου του $\boldsymbol{x}_{1:n}$. Η μεταβλητή αυτή εκφράζει την συμπεριφορά της ανέλιξης, όταν ο χρονικός ορίζοντας τείνει στο άπειρο. Είναι επίσης δυνατή η εύρεση της κατανομής και για την περίπτωση της φραγμένης στοχαστικής ανέλιξης.

Η μέθοδος εφαρμόστηκε σε δεδομένα θερμοκρασίας, βροχόπτωσης και απορροής από τον Βοιωτικό Κηφισό ποταμό (Rozos et al. 2004) και σε δεδομένα θερμοκρασίας από το Βερολίνο (Koutsoyiannis et al. 2007) και την Βιέννη (Koutsoyiannis 2011). Η κλιματική

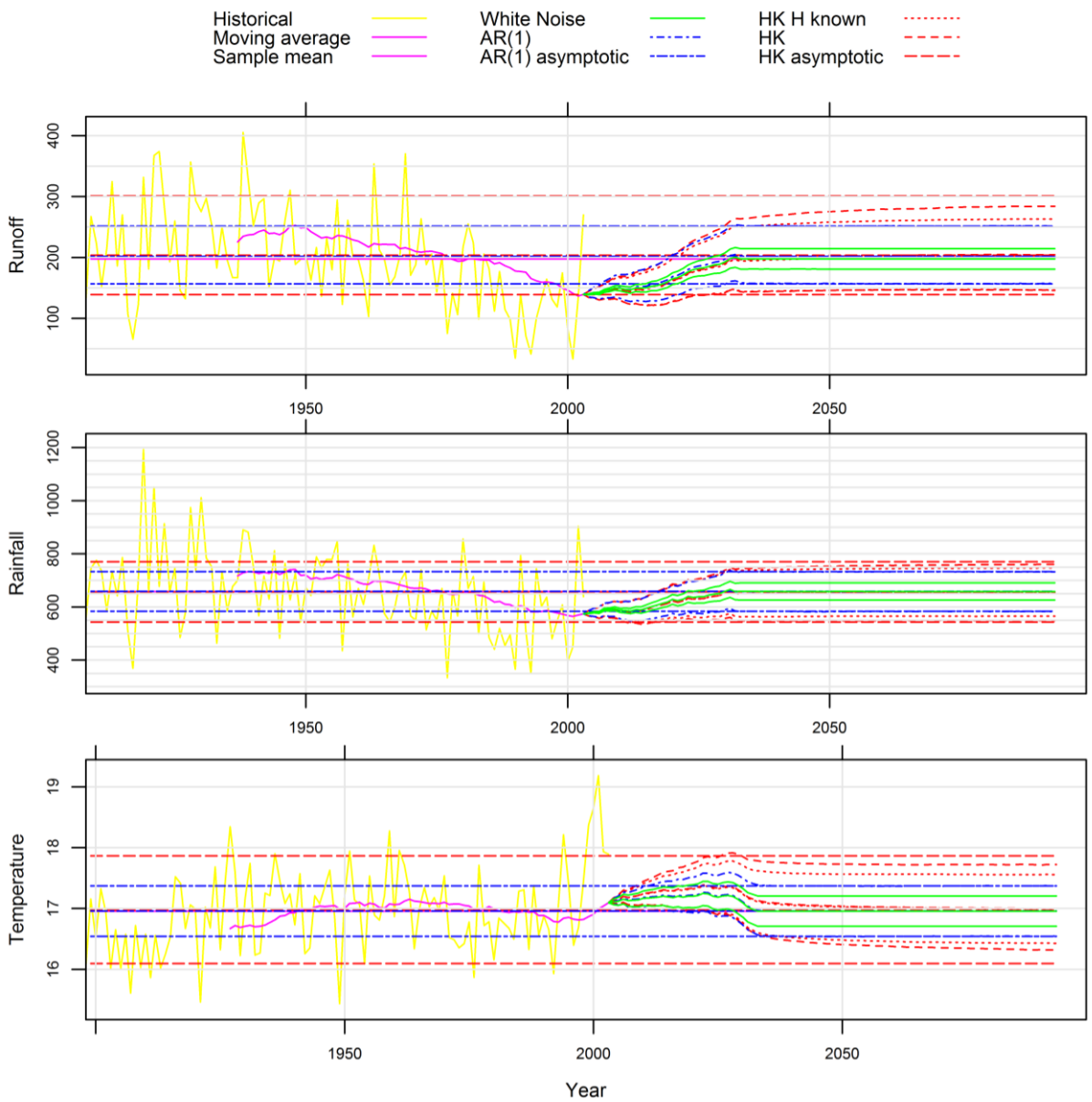
μεταβλητή που εξετάστηκε είναι η $\underline{x}_{t(30)}$ η οποία ορίζεται ως εξής:

$$\underline{x}_{t(30)} := (1/30)\left(\sum_{l=t-29}^{n} x_l + \sum_{l=n+1}^{t} \underline{x}_l\right),\ t = n+1, \ldots, n+29 \text{ και } \underline{x}_{t(30)} := (1/30)\sum_{l=t-29}^{t} \underline{x}_l,\ t = n+30, n+31, \ldots \quad (42)$$

Στα σχήματα 4 και 5 δίνονται οι εκ των υστέρων προβλεπτικές 0.95-περιοχές εμπιστοσύνης. Για την περίπτωση της απορροής του Βοιωτικού Κηφισού στο Σχήμα 4 η περιοχή είναι ασύμμετρη λόγω του κάτω φράγματος και της σχετικά μεγάλης διασποράς, αντίθετα με τις άλλες περιπτώσεις η διασπορά είναι σχετικά μικρή, όπως στην περίπτωση της βροχόπτωσης στην Αλίαρτο στο Σχήμα 4 ή δεν υπάρχουν φράγματα. Για όλες τις περιπτώσεις, οι μεγαλύτερες περιοχές εμπιστοσύνης ήταν αυτές της HKp (λόγω της εμμονής), ακολουθούμενες από την AR(1) και την WN. Φυσικά οι περιοχές όπου το *H* θεωρείται άγνωστο είναι μεγαλύτερες από τις περιπτώσεις που θεωρείται γνωστό και ίσο με την εκτίμηση μέγιστης πιθανοφάνειας. Οι δε ασυμπτωτικές περιοχές εμπιστοσύνης φαίνεται να είναι ικανές να μοντελοποιήσουν τις κλιματικές διακυμάνσεις για την περίπτωση που χρησιμοποιείται η HKp, επιβεβαιώνοντας την επιλογή της για την μελέτη του κλίματος.

Επί της πρόβλεψης έμμονων ανελίξεων χρησιμοποιώντας τα αποτελέσματα ντετερμινιστικών μοντέλων

Ένα πρόβλημα που συναντάται στην τεχνική υδρολογική κοινότητα είναι η πρόβλεψη υδρολογικών μεταβλητών με δεδομένες τις ιστορικές παρατηρήσεις και τις προβλέψεις παρελθοντικών και μελλοντικών γεγονότων από ντετερμινιστικά μοντέλα. Αρκετές μέθοδοι έχουν αναπτυχθεί για να αντιμετωπίσουν αυτό το πρόβλημα υπό την παραδοχή της Μαρκοβιανής σχέσης μεταξύ των μεταβλητών. Σε αυτήν την εργασία γίνεται προσπάθεια επέκτασης του προβλήματος και σε διεργασίες που παρουσιάζουν συμπεριφορά HK.



Σχήμα 4. Ιστορικά δεδομένα και περιοχές εμπιστοσύνης για το μελλοντικό κλίμα (για $1 - a = 0.95$ και κλιματική χρονική κλίμακα 30 έτη) για (πάνω) την απορροή του Βοιωτικού Κηφισού, (μέσο) βροχόπτωση στην Αλίαρτο, και (κάτω) θερμοκρασία στην Αλίαρτο.

Προς αυτήν την κατεύθυνση έστω ότι $\{\underline{x}_{1t}\}$ και $\{\underline{x}_{2t}\}$, $t = 1, 2, \ldots$ είναι δύο HKp's με παραμέτρους (μ_1, σ_1, H_1) και (μ_2, σ_2, H_2) αντίστοιχα. Τότε η διμεταβλητή στοχαστική ανέλιξη $\{\underline{\mathbf{x}}_t = (\underline{x}_{1t},\underline{x}_{2t})\}$, $t = 1, 2, \ldots$ είναι μια ισορροπημένη HKp εάν (Amblard et al. 2012)

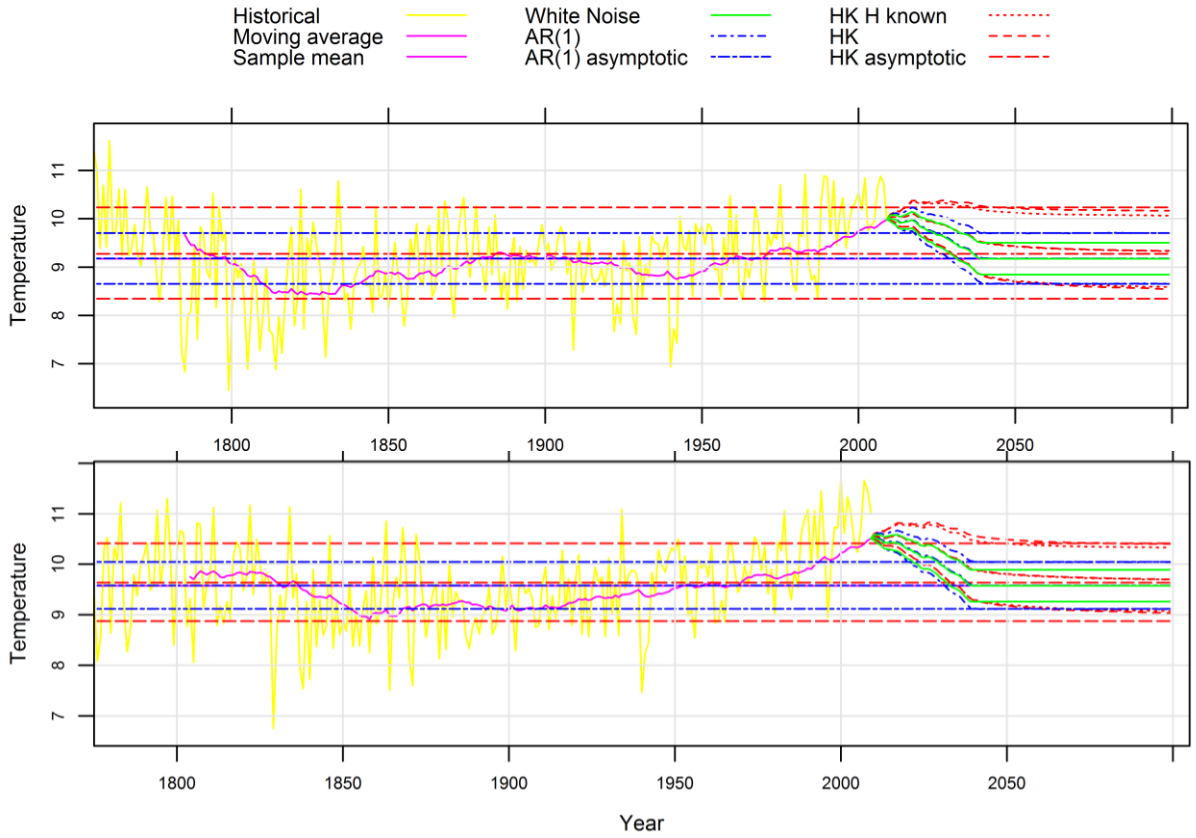
$$w_{ij}(k) := \rho_{i,j}\,|k|^{H_i+H_j},\ \rho_{i,i} = 1,\ \rho_{i,j} = \rho_{j,i} = \rho,\ \{i,j\} \in \{\{1,2\},\{1,2\}\} \tag{43}$$

$$\gamma_{ij}(k) := \mathrm{Cov}[\underline{x}_{it},\ \underline{x}_{j\,t+k}] = (1/2)\ \sigma_i\ \sigma_j\ \ (\ w_{ij}(k-1) - 2\ w_{ij}(k) + w_{ij}(k+1)\ ) \tag{44}$$

υπό τον περιορισμό

$$\rho^2 \le \frac{\Gamma(2H_1+1)\ \Gamma(2H_2+1)\ \sin(\pi H_1)\ \sin(\pi H_2)}{\Gamma^2(H_1+H_2+1)\ \sin^2(\pi(H_1+H_2)/2)} \tag{45}$$

Επισημαίνεται ότι για $i = j$, η (44) είναι ισοδύναμη με την (9). Στην παρούσα εργασία εκτιμήθηκαν οι παράμετροι της διμεταβλητής HKp με την μέθοδο της μέγιστης πιθανοφάνειας.



Σχήμα 5. Ιστορικά δεδομένα και περιοχές εμπιστοσύνης για το μελλοντικό κλίμα (για $1 - a = 0.95$ και κλιματική χρονική κλίμακα 30 έτη) για (πάνω) την θερμοκρασία στο Βερολίνο, και (κάτω) την θερμοκρασία στην Βιέννη.

Έστω ότι $\boldsymbol{x}_{1\ 1:(n+k)}$ είναι το αποτέλεσμα του ντετερμινιστικού μοντέλου και $\boldsymbol{x}_{2\ 1:n}$ είναι τα παρατηρημένα δεδομένα. Επιθυμούμε να βρούμε την κατανομή του $\boldsymbol{x}_{2\ (n+1):(n+m)}$ δεσμευμένων των $\boldsymbol{x}_{1\ 1:(n+m)}$ και $\boldsymbol{x}_{2\ 1:n}$. Ορίζουμε $\underline{\boldsymbol{w}}_{1:n} := (\underline{\boldsymbol{x}}_{1\ 1:n}^{\mathrm{T}}, \underline{\boldsymbol{x}}_{2\ 1:n}^{\mathrm{T}})^{\mathrm{T}}$ με πίνακα συνδιασπορών που δίνεται από την (46) και διαχωρίζεται σύμφωνα με την (50)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix} \tag{46}$$

$$\boldsymbol{\Sigma}_1 := \sigma_1^2 \boldsymbol{R}_1,\ \boldsymbol{R}_1(i,j) = \boldsymbol{R}_1(j,i) := \rho_{1(j-i)} \text{ and } \boldsymbol{\Sigma}_2 := \sigma_2^2 \boldsymbol{R}_2,\ \boldsymbol{R}_2(i,j) = \boldsymbol{R}_2(j,i) := \rho_{2(j-i)} \tag{47}$$

$$\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12} := \rho_{12}\, \sigma_1\, \sigma_2\, \boldsymbol{R}_{21},\ \boldsymbol{R}_{21}(i,j) = \boldsymbol{R}_{21}(j,i) = \boldsymbol{R}_{21}(j-i) := \rho_{21}(j-i) \tag{48}$$

$$\rho_{21}(j-i) := \gamma_{21}(j-i) / (\rho\, \sigma_1\, \sigma_2) = (1/2)\, (\, |j-i-1|^{H_1+H_2} - 2\, |j-i\,|^{H_1+H_2} + |j-i+1|^{H_1+H_2}\,) \tag{49}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{121} & \boldsymbol{\Sigma}_{122} \\ \boldsymbol{\Sigma}_{211} & \boldsymbol{\Sigma}_{2n} & \boldsymbol{\Sigma}_{2nm} \\ \boldsymbol{\Sigma}_{212} & \boldsymbol{\Sigma}_{2mn} & \boldsymbol{\Sigma}_{2m} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}_1 & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_2 \end{bmatrix} \quad (50)$$

όπου ο $\boldsymbol{\Sigma}_{2m}$ είναι ένας m x m πίνακας και

$$\boldsymbol{P}_1 = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{121} \\ \boldsymbol{\Sigma}_{211} & \boldsymbol{\Sigma}_{2n} \end{bmatrix}, \boldsymbol{P}_{21} = \begin{bmatrix} \boldsymbol{\Sigma}_{212} & \boldsymbol{\Sigma}_{2mn} \end{bmatrix}, \boldsymbol{P}_{12} = \begin{bmatrix} \boldsymbol{\Sigma}_{122} \\ \boldsymbol{\Sigma}_{2nm} \end{bmatrix}, \boldsymbol{P}_2 = \boldsymbol{\Sigma}_{2m} \quad (51)$$

Τότε η εκ των υστέρων προβλεπτική κατανομή του $\boldsymbol{x}_{2\,(n+1):(n+m)}$ δεσμευμένων των $\boldsymbol{x}_{1\,1:(n+m)}$, $\boldsymbol{x}_{2\,1:n}$ και $\boldsymbol{\theta}$ είναι

$$f(\boldsymbol{x}_{2\,(n+1):(n+m)} | \boldsymbol{x}_{1\,1:(n+m)}, \boldsymbol{x}_{2\,1:n}, \boldsymbol{\theta}) = (2\pi\sigma^2)^{-m/2} \, |\boldsymbol{R}_{m|n}|^{-1/2} \cdot$$

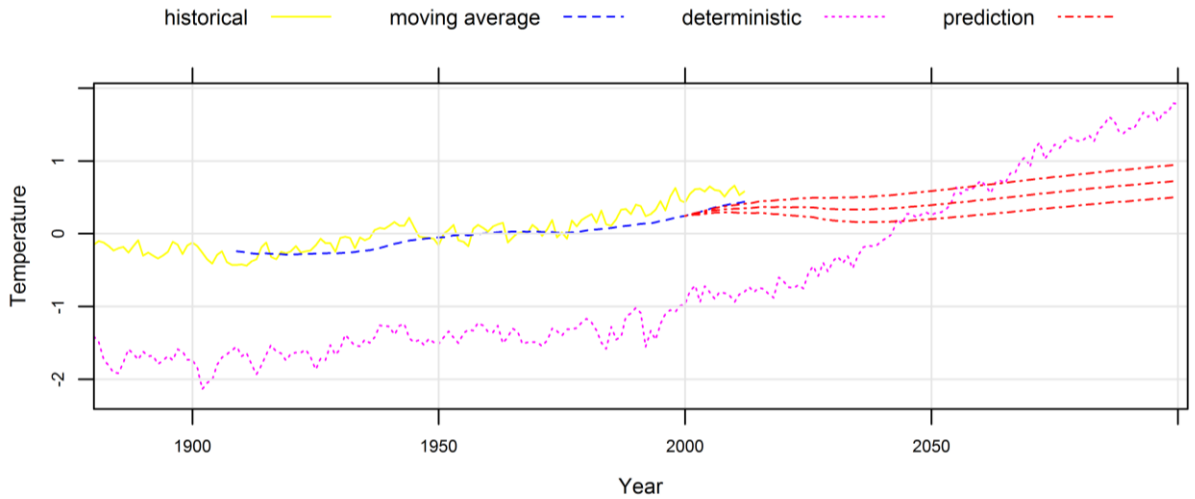
$$\exp[(-1/2\sigma^2) \, (\boldsymbol{x}_{2\,(n+1):(n+m)} - \boldsymbol{\mu}_{m|n})^{\mathrm{T}} \, \boldsymbol{R}_{m|n}^{-1} \, (\boldsymbol{x}_{2\,(n+1):(n+m)} - \boldsymbol{\mu}_{m|n})] \quad (52)$$

όπου τα $\boldsymbol{\mu}_{m|n}$ και $\boldsymbol{R}_{m|n}$ δίνονται από την:

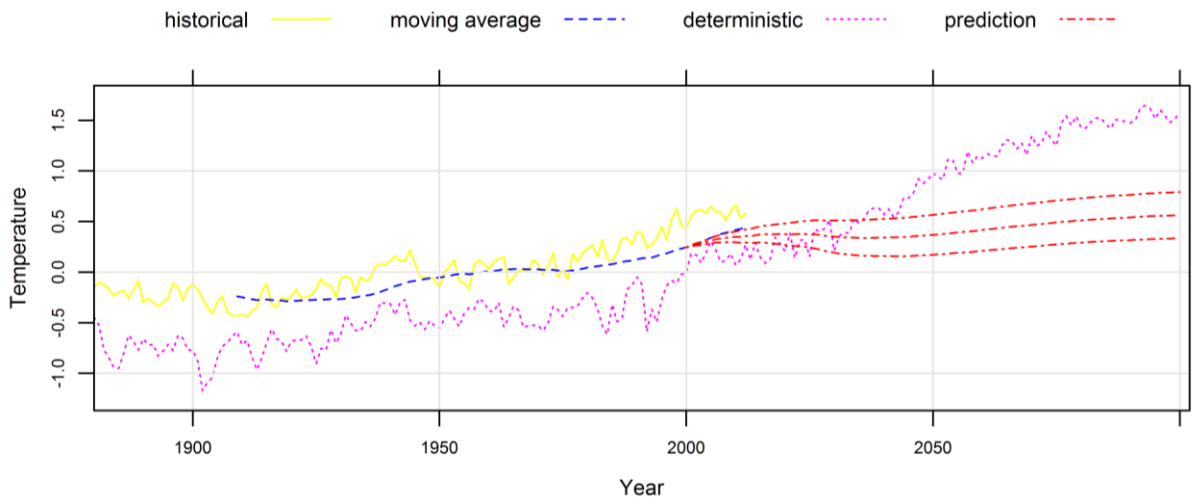
$$\boldsymbol{\mu}_{m|n} = \mu_2 \boldsymbol{e}_m + \boldsymbol{P}_{21} \, \boldsymbol{P}_1^{-1} \, (\, (\boldsymbol{x}_{1\,1:(n+m)}^{\mathrm{T}} \, , \boldsymbol{x}_{2\,1:n}^{\mathrm{T}})^{\mathrm{T}} - (\mu_1 \boldsymbol{e}_{n+m}^{\mathrm{T}} \, , \mu_2 \boldsymbol{e}_n^{\mathrm{T}})^{\mathrm{T}} \,) \quad (53)$$

$$\boldsymbol{R}_{m|n} = \boldsymbol{P}_2 - \boldsymbol{P}_{21} \, \boldsymbol{P}_1^{-1} \, \boldsymbol{P}_{12} \quad (54)$$

Χρησιμοποιώντας ως τιμή του $\boldsymbol{\theta}$ την εκτίμηση μέγιστης πιθανοφάνειας, εξετάστηκαν περιπτώσεις που τα μοντέλα προέβλεπαν θερμοκρασία και βροχόπτωση. Στις περισσότερες περιπτώσεις η εκτιμώμενη παράμετρος ρ ήταν σχεδόν ίση με 0, ώστε τελικά η πληροφορία που προσέθεταν τα ντετερμινιστικά μοντέλα να είναι μηδενική. Παρατίθενται τα Σχήματα 6, 7 που επαληθεύουν αυτόν τον ισχυρισμό.



Σχήμα 6. 95% περιοχή εμπιστοσύνης για την πρόβλεψη του κινούμενου μέσου όρου των 30 ετών της μέση ετήσιας θερμοκρασίας (°C) για το σενάριο A1B του μοντέλου ECHO-G, χρησιμοποιώντας τις διαφορές της μέσης ετήσιας θερμοκρασίας του NOAA.



Σχήμα 7. 95% περιοχή εμπιστοσύνης για την πρόβλεψη του κινούμενου μέσου όρου των 30 ετών της μέση ετήσιας θερμοκρασίας (°C) για το σενάριο B1 του μοντέλου ECHO-G, χρησιμοποιώντας τις διαφορές της μέσης ετήσιας θερμοκρασίας του NOAA.

Επισημαίνεται ότι η ανωτέρω μέθοδος είναι μια επέκταση προηγούμενων μελετών Krzysztofowicz 1999a,b; Wang et al. 2009), στις οποίες όμως εξετάστηκαν περιπτώσεις Μαρκοβιανών μοντέλων.

Συμπεράσματα και προτάσεις

Ο αρχικός σκοπός αυτής της εργασίας ήταν η ανάπτυξη ενός στοχαστικού πλαισίου για την πρόβλεψη υδροκλιματικών μεταβλητών χρησιμοποιώντας Μπεϋζιανές τεχνικές. Για να λυθεί το πρόβλημα αποφασίσαμε να ακολουθήσουμε μια παραμετρική προσέγγιση. Έτσι ένα στοχαστικό μοντέλο επιλέχθηκε. Η επιλογή βασίστηκε σε θεμελιωμένα εκ των προτέρων κριτήρια, και συγκεκριμένα τον δεύτερο νόμο της θερμοδυναμικής, δηλαδή την μεγιστοποίηση της εντροπίας, υπό ορισμένους περιορισμούς η οποία οδηγεί σε μια οικογένεια μεταβλητών οι οποίες παρουσιάζουν συμπεριφορά ΗΚ. Μια Μπεϋζιανή προσέγγιση επιλέχθηκε για την εύρεση της εκ των υστέρων προβλεπτικής κατανομής των μελετώμενων υδροκλιματικών μεταβλητών.

Μια προηγούμενη προσέγγιση η οποία ανέπτυξε ένα στοχαστικό πλαίσιο διερευνήθηκε. Τα αποτελέσματα της μελέτης ήταν ενθαρρυντικά. Ωστόσο αυτή ήταν βασισμένη σε έναν διαισθητικό αλγόριθμο. Σε αυτήν την εργασία αποδείχθηκε αναλυτικά ότι ο αλγόριθμος αυτός έχει ικανοποιητικές ιδιότητες. Εξαιτίας των περιορισμών της πρώτης προσέγγισης αποφασίστηκε η επίλυση του προβλήματος με την χρήση Μπεϋζιανής στατιστικής. Ένα πρώτο βήμα προς αυτήν την κατεύθυνση ήταν η εκτίμηση των εκτιμητριών των παραμέτρων του στοχαστικού μοντέλου. Τα αποτελέσματα ήταν ξανά ενθαρρυντικά. Έτσι σε ένα δεύτερο βήμα λύσαμε το πρόβλημα με Μπεϋζιανή μέθοδο χρησιμοποιώντας μια εκ των προτέρων

κατανομή μηδενικής πληροφορίας για τις παραμέτρους του στοχαστικού μοντέλου. Επιπλέον αποφασίστηκε η χρήση της πληροφορίας ντετερμινιστικών μοντέλων για την βελτίωση των αποτελεσμάτων του στοχαστικού μοντέλου. Η βελτίωση τελικά δεν ήταν σημαντική.

Φυσικά αυτή η εργασία δεν επιλύει πλήρως το πρόβλημα. Τα τελευταίας τεχνολογίας μοντέλα για την κλιματική πρόβλεψη είναι ντετερμινιστικά και η έρευνα επικεντρώθηκε στην ανάπτυξή τους, παρά τις ελλείψεις τους. Ελάχιστα ερευνητικά αποτελέσματα υπάρχουν στον τομέα της κλιματικής πρόβλεψης με στοχαστικές μεθόδους. Ελπίζουμε τα αναλυτικά εργαλεία που αναπτύχθηκαν εδώ να προσφέρουν σε αυτήν την έρευνα.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Long horizons of prediction within a stochastic framework

In the paper "A random walk on water" (Koutsoyiannis 2010), prepared for the Henry Darcy Medal award occasion, the question whether deterministic prediction of climate is possible arises and the answer is emphatically "no". The author argues that "*it is possible to shape a consistent stochastic representation of natural processes, in which predictability (suggested by deterministic laws) and unpredictability (randomness) coexist and are not separable or additive components. Deciding which of the two dominates is simply a matter of specifying the time horizon and scale of the prediction. Long horizons of prediction are inevitably associated with high uncertainty, whose quantification relies on the long-term stochastic properties of the processes*". Hence a deterministic prediction of climate for a long horizon is impossible.

To support his arguments Koutsoyiannis (2010) studies a deterministic toy model of a caricature hydrological system. In this toy model the evolution of the state of the system is observed. The state of the system is a function of time, the parameters of the toy model and the initial conditions of the system. A small change of the initial conditions results in completely different trajectories of the state of the system after a long time horizon. Assuming that a precise observation of the initial conditions is not possible Koutsoyiannis (2010) concludes that deterministic dynamics can produce good predictions only for short time horizons and that a shift of paradigm from determinism to stochastics is needed.

On the other hand the scientific community focuses on General Circulation Models (GCMs) following Phillips' (1956) first published attempt to predict the future climate. GCMs are numerical representations of the climate system based on the physical, chemical and biological properties of its components, their interactions and feedback processes, and accounting for all or some of its known properties. They are applied as a research tool to study and simulate the climate, and for operational purposes, including monthly, seasonal and interannual climate predictions (IPCC 2007 p.946). They are deterministic models based on the Navier–Stokes equations.

There are numerous attempts to predict future climate based on the results of GCMs. For instance, the projection of surface temperature to the year 3000 (!) (IPCC 2007 p.823) is mentioned. However, there have been many criticisms regarding the validity of the results of GCMs. Some of them are listed here:

- They have negligible hindcast properties, i.e. they cannot reproduce the climate of the past, even for relatively small time horizons, e.g. Koutsoyiannis et al. (2008a), Anagnostopoulos et al. (2010), Fyfe et al. (2013).

- They cannot predict the regional climate on seasonal to decadal time scales, even for short time horizons, e.g. Handorf and Dethloff (2012), Scafetta (2013).

- They do not model adequately the climate due to their inherent properties, e.g. Spencer and Braswell (2011), McNider et al. (2012), Stevens and Bony (2013).

1.2 Long-term persistence in predicting the climate

Little work has been done in choosing a suitable stochastic model for climate prediction (Keenan 2014). Here the approach of Koutsoyiannis (2011) will be adopted. Koutsoyiannis (2011) proved that "*under certain constraints, i.e. the preservation of the mean, variance and lag-1 autocovariance and an inequality relationship between conditional and unconditional entropy production, the extremization of entropy production of stochastic representations of natural systems, performed at asymptotic times (zero or infinity) results in Hurst-Kolmogorov processes*". An a priori choice of the statistical model, justified by the implementation of a principle well established in physics, i.e. the second law of thermodynamics, seems appropriate (Keenan 2014).

1.2.1 Hurst-Kolmogorov behaviour of geophysical processes

The Hurst-Kolmogorov (HK) behaviour of hydrological and other geophysical processes as named by Koutsoyiannis (2010) was discovered by Hurst (1951), later became known with several names such as Hurst phenomenon, long-term persistence and long-range dependence, and has subsequently received extensive attention in the literature. Earlier, Kolmogorov (1940), when studying turbulence, had proposed a mathematical model to describe this behaviour, which was further developed by Mandelbrot and van Ness (1968) and has been known as simple scaling stochastic model or fractional Gaussian noise (see Beran 1994; Embrechts and Maejima 2002; Palma 2007; Doukhan et al. 2003; Robinson 2003; and the references therein).

Many studies on this kind of behaviour have been accomplished. Beran (1994) discusses some of them related to diverse scientific fields from climatology to agronomy and from economics to chemistry. The HK behaviour is of special interest for hydrologists, e.g. see Koutsoyiannis (2002, 2003, 2006a), Koutsoyiannis and Montanari (2007), all published in

hydrological journals. Furthermore many studies on the HK behaviour of geophysical processes have been published, e.g. recently Buette et al. (2006) studied the Irish daily wind speeds and Zhang et al. (2009) studied the scaling properties of the hydrological series in the Yellow River basin.

1.2.2 A mathematical definition of Hurst-Kolmogorov behaviour

In the rest of this thesis the Dutch convention for notation, according to which random variables and stochastic processes are underlined (Hemelrijk 1966) will be used. We assume that $\{\underline{x}_t\}$, is a stationary stochastic process in discrete time $t = 1,2,\ldots$ with mean

$$\mu := \mathrm{E}[\underline{x}_t] \tag{1.1}$$

standard deviation

$$\sigma := \sqrt{\mathrm{Var}[\underline{x}_t]} \tag{1.2}$$

autocovariance function

$$\gamma_k := \mathrm{Cov}[\underline{x}_t, \underline{x}_{t+k}],\ k = 0, \pm1, \pm2, \ldots \tag{1.3}$$

and autocorrelation function (ACF)

$$\rho_k := \gamma_k / \sigma,\ k = 0, \pm1, \pm2, \ldots \tag{1.4}$$

Then the HK behaviour can be modelled by $\{\underline{x}_t\}$ if (Beran 1994 p.42)

$$\lim_{k\to\infty} \rho_k / (c\, k^{-a}) = 1,\ 0 < a < 1,\ 0 < c \tag{1.5}$$

The parameter H, defined by

$$H := 1 - a / 2 \tag{1.6}$$

is the Hurst or self-similarity parameter of the process.

Processes exhibiting the HK behaviour include the Fractional ARIMA processes (Beran 1994 p.59) and the Hurst-Kolmogorov stochastic process (HKp), also known as fractional Gaussian noise, (fGn); see also Beran (1994 p.55) and Koutsoyiannis (2010). The typical modelling approach with artificial models such as the Fractional ARIMA processes suffers in many aspects (Koutsoyiannis 2015). On the other hand the HKp does not belong to this class of models, thus we prefer to use it in this manuscript to model geophysical processes.

Furthermore we assume that $\{\underline{x}_t\}$, $t = 1, 2, \ldots$ is normal. Let κ be a positive integer that represents a timescale larger than 1, the original time scale of the process $\{\underline{x}_t\}$. The averaged

stochastic process on that timescale is denoted as

$$\underline{x}_t^{(\kappa)} := (1/\kappa) \sum_{l=(t-1)\,\kappa+1}^{t\,\kappa} \underline{x}_l \tag{1.7}$$

The notation implies that a superscript (1) could be omitted, i.e. $\underline{x}_t^{(1)} \equiv \underline{x}_t$. Now we consider the following equation that defines the HKp (Koutsoyiannis 2003).

$$(\underline{x}_i^{(\kappa)} - \mu) \overset{d}{=} \left(\frac{\kappa}{\lambda}\right)^{H-1} (\underline{x}_j^{(\lambda)} - \mu),\ 0 < H < 1,\ \ i, j = 1, 2, \ldots \text{ and } \kappa, \lambda = 1, 2, \ldots \tag{1.8}$$

The ACF of the HKp is (Koutsoyiannis 2003)

$$\rho_k = |k+1|^{2H}/2 + |k-1|^{2H}/2 - |k|^{2H},\ k = 0, 1, \ldots \tag{1.9}$$

with an asymptotic behaviour same with that of (1.5).

1.2.3 Consequences of the Hurst-Kolmogorov behaviour

A direct consequence of (1.8) is that

$$\mathrm{Var}[\underline{x}_1^{(n)}] := \sigma^2\, n^{2H-2} \tag{1.10}$$

Assuming that the climate variable of interest is modelled by $\{\underline{x}_t\}$ it is obvious that $\mathrm{Var}[\underline{x}_1^{(n)}] > \sigma^2/n$ for $0.5 < H < 1$ and that $\mathrm{Var}[\underline{x}_1^{(n)}] = \sigma^2/n$, when $H = 0.5$, which corresponds to independent $\{\underline{x}_t\}$, $t = 1, 2, \ldots$. Thus a HKp can explain bigger variations of an observed variable compared to independent variables, e.g. see Figure 1.1. This is shown for instance by Koutsoyiannis (2006a), and will be studied thoroughly in Chapter 4.

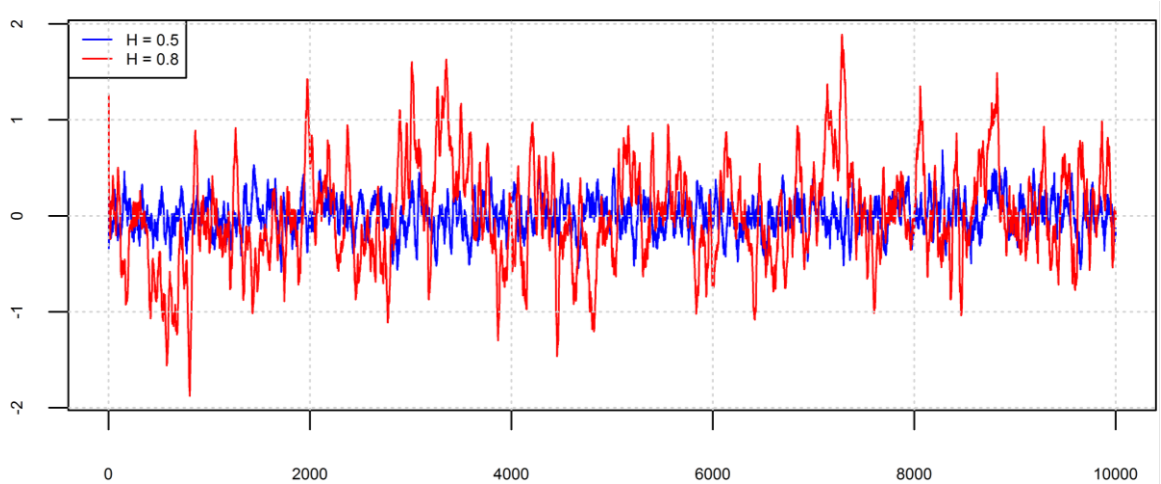


Figure 1.1. Moving average for 30 points for a simulated HKp with $\mu = 0$, $\sigma = 1$ and (a) $H = 0.5$ and (b) $H = 0.8$.

The fact that a HKp is stationary is emphasized also. Koutsoyiannis (2006b) shows that the

HKp can reproduce climatic trends due to its scaling behaviour. These trends are then considered as large-scale fluctuations. Paradoxical results obtained by modelling long-term terms as deterministic components are avoided, e.g the case study examined by Koutosyiannis (2006b, Section 2.6)). Koutsoyiannis (2006b) explains that the nonstationarity modelling approach is contradictory in its rationale and its terminology, implying misleading perception of the phenomena and uncertainty estimation.

1.3 A Bayesian framework on the prediction of climate

Assuming that the chosen statistical model is appropriate to model the data, the problem of predicting future variables conditional on their observed values belongs to the branch of parametric statistics. A Bayesian approach is suitable to solve the problem. For a defence of the Bayesian choice the interested reader is referred to Robert (2007 p.507-518).

1.3.1 Definition of the Bayesian statistical model

We assume that there is a record of n observations which we write as a vector $\boldsymbol{x}_{1:n} = (x_1,\ldots,x_n)^T$ (where the superscript T is used to denote the transpose of a vector or matrix and vectors and matrices are bolded, see also Appendix A for more on the notations). Furthermore we define the random variable $\underline{\boldsymbol{x}}_{1:n} := (\underline{x}_1,\ldots,\underline{x}_n)^T$, where

$$\underline{\boldsymbol{x}}_{1:n} \sim f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta}) \tag{1.11}$$

where f is a probability distribution with unknown parameter $\boldsymbol{\theta}$ which belongs to a vector space Θ of finite dimension. Then a parametric statistical model consists of $\boldsymbol{x}_{1:n}$ and $\underline{\boldsymbol{x}}_{1:n}$ (Robert, 2007 p.7).

We assume now that the uncertainty of the parameter $\boldsymbol{\theta}$ is modelled by a probability distribution π on Θ, called prior distribution, such that

$$\underline{\boldsymbol{\theta}} \sim \pi(\boldsymbol{\theta}) \tag{1.12}$$

Notice that we generally use the symbol π for probability density functions of parameters. Then the distribution of $\underline{\boldsymbol{\theta}}$ conditional on $\boldsymbol{x}_{1:n}$ is called posterior distribution and is used to make inference on $\underline{\boldsymbol{\theta}}$.

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:n}) = \frac{f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \tag{1.13}$$

Then the Bayesian statistical model consists of $\boldsymbol{x}_{1:n}$, $\underline{\boldsymbol{x}}_{1:n}$ and π (Robert 2007 p.9).

Notice that:

- The term $\int f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ in (1.13) does not depend on $\boldsymbol{\theta}$. Furthermore for the given record of observations it is a constant. Thus $\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})$ is proportional to $f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta})\ \pi(\boldsymbol{\theta})$. An important consequence is that to calculate $\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})$ the calculation of the integral term is not necessary.

- The analysis on $\underline{\boldsymbol{\theta}}$ is performed, conditional upon $\boldsymbol{x}_{1:n}$ (Robert 2007 p.529). For instance after modelling $\boldsymbol{x}_{1:n}$ with the parametric statistical model, an inference on $\boldsymbol{\theta}$ is made, calculating its confidence intervals. The typical approach of the construction of confidence intervals is justified on a long-term basis (Robert 2007 p.16). On the other hand the Bayesian statistical model proposes a procedure for the problem at hand (Robert 2007 p.513).

- Bayesian inference obeys the likelihood principle, according to which the information brought by $\boldsymbol{x}_{1:n}$ about $\boldsymbol{\theta}$ is entirely contained in the likelihood function $l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})$. For more information on the likelihood principle and its application in Bayesian statistics see Robert (2007 p.15).

- At the cost of an additional assumption, i.e. the introduction of the prior distribution of $\underline{\boldsymbol{\theta}}$, inference on $\underline{\boldsymbol{\theta}}$ or some future variables reduces to a simple technical problem. Furthermore the cost of this assumption can be reduced using a noninformative prior distribution, i.e. a distribution derived from f (Robert 2007 p.127) when this is the only available information. Sometimes this automatic derivation leads to improper (or generalized, Robert 2007 p.27) distributions for $\underline{\boldsymbol{\theta}}$. In these cases $\underline{\boldsymbol{\theta}}$ cannot be considered as a random variable. As pointed by Robert (2007 p.10) "*the importance of the prior distribution in a Bayesian statistical analysis is not at all that the parameter of interest $\boldsymbol{\theta}$ can (or cannot) be perceived as generated from π or even as a random variable, but rather that the use of a prior distribution is the best way to summarize the available information (or even the lack of information) about this parameter, as well as the residual uncertainty, thus allowing for incorporation of this imperfect information in the decision process*". Furthermore if the integral $\int f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ is finite, the distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})$ is well defined and can be used as a regular probability distribution to describe the properties of $\boldsymbol{\theta}$. For consistency reasons $\underline{\boldsymbol{\theta}}$ will be handled as a random variable (Robert 2007 p.165).

1.3.2 Parameter estimation and confidence regions

The available information on $\underline{\boldsymbol{\theta}}$ is summarized by its posterior distribution. However, sometimes an estimate of $\boldsymbol{\theta}$ is required. Bayesian point estimation is based on decision theory. A loss function $L(\boldsymbol{\theta},\delta)$ is selected. This function evaluates the error $L(\boldsymbol{\theta},\delta)$ associated with the decision δ (i.e. the estimate of $\boldsymbol{\theta}$) when the parameter takes the value $\boldsymbol{\theta}$ (Robert 2007 p.52). A

Bayes rule $\delta^{\pi}(\boldsymbol{x}_{1:n})$, which is the value of δ that minimizes the function $\mathrm{E}^{\pi}[L(\boldsymbol{\theta},\delta)|\boldsymbol{x}_{1:n}]$ for the given π, $\boldsymbol{x}_{1:n}$ and $\underline{\boldsymbol{x}}_{1:n}$, is defined as a Bayesian estimate of $\boldsymbol{\theta}$ (Robert 2007 p.173).

Furthermore after the computation of $\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})$ confidence regions on $\underline{\boldsymbol{\theta}}$ are available. In the Bayesian formulation $\underline{\boldsymbol{\theta}}$ has a given probability to belong to a fixed confidence region. On the other hand confidence intervals have a given probability to contain the parameter $\boldsymbol{\theta}$. The former notion is more intuitive (Robert 2007 p.260).

1.3.3 Predictive distribution

In our particular problem, i.e. the prediction of a future variable $\underline{\boldsymbol{y}}$ conditional on $\boldsymbol{x}_{1:n}$ the posterior predictive distribution defined by (Robert 2007 p.22)

$$g(\boldsymbol{y}|\boldsymbol{x}_{1:n}) = \int g(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x}_{1:n})\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})\mathrm{d}\boldsymbol{\theta} \tag{1.14}$$

solves the problem. The calculation of the integral (1.14) is not necessary, since $g(\boldsymbol{y}|\boldsymbol{x}_{1:n})$ can be simulated from the mixture of the conditional distributions $\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})$ and $g(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x}_{1:n})$ (Gelman et al. 2004 p.74). The future variable can be any variable of interest, e.g. $\underline{\boldsymbol{y}} = \underline{\boldsymbol{x}}_{n_1:n_2}$ for $n_1 > n$ where

$$\underline{\boldsymbol{x}}_{n_1:n_2} := (\underline{x}_{n_1},\ldots,\underline{x}_{n_2})^{\mathrm{T}} \tag{1.15}$$

One could claim that instead of using the posterior distribution of $\underline{\boldsymbol{\theta}}$, it would suffice to substitute $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in $g(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x}_{1:n})$. As will be shown in Chapter 4 neglecting the uncertainty in the estimation of $\boldsymbol{\theta}$ for the problem at hand, valuable information is lost.

1.4 Objectives and research questions

1.4.1 The broader perspective

To summarize the discussion, this thesis focuses on the prediction of hydroclimatic variables including temperature and rainfall, using a stochastic framework. Handling this problem using stochastics is not the state-of-the-art for the climatology scientific community, however it seems to be the only reasonable option for an adequate answer to this problem. Not surpringly, after completing the technical part of the thesis the author came up to an article pointing that the number of climatologists supporting this options steadily increases (Macilwain 2014). A viable answer to the question, which is the most appropriate stochastic model for studying climate has not yet been given (Keenan 2014). In this thesis it is assumed that geophysical processes exhibit HK behaviour and are modelled correspondingly.

Owing to their structure, GCMs do not regard the climate variables as random. Thus climate projections are point estimates. Contrary to the aforementioned approach, stochastic models treat climate variables as random. Hence all the information about the climate variables is included in their distribution. It is crucial to find this distribution. The most practical parametric approach is the Bayesian one, which reduces the complexity of the problem to technical issues, with the cost of just one additional assumption, i.e. the assignment of a prior distribution to $\boldsymbol{\theta}$.

1.4.2 Research objectives

The objective of this thesis is to provide some tools towards the development of a stochastic framework for the prediction of hydroclimatic variables. The main components of this framework are the stochastic model and the data. Topics such as the estimation of the parameters of the model, the uncertainty of the estimation of the parameters and the incorporation of this uncertainty in the prediction uncertainty are examined. An additional component, namely the output of GCMs is incorporated in the model. Information gained from the deterministic models is assessed and is used to update the inference of the stochastic model.

The development of these tools should contribute in quantifying the uncertainty in the prediction of climate. Uncertainty quantification contrary to point estimation may explain climate variability.

1.4.3 Research questions

The main questions that are addressed in the manuscript are:

- How can the uncertainty in the estimation of the parameters be integrated in the uncertainty of the prediction?
- How can the data be used for the prediction?
- Which is an appropriate framework to gain from available information from deterministic models?

To answer these questions, a previous typical statistical approach of the problem is investigated and is justified theoretically. However this approach is not adequate to solve the problem, owing to its indirect but encouraging results. Thus to continue the research, another way is chosen, namely the Bayesian approach. Some results on the estimation of parameters using a maximum likelihood estimator further strengthen the Bayesian choice.

The main problem i.e. the prediction, is solved in a Bayesian way, revealing uncertainties not previously paid attention to. The thesis concludes with an attempt to improve prediction using deterministic information, incorporating this in a stochastic model with dependence structure more complicated than a Markovian one.

1.5 Thesis outline

This thesis examines several issues associated with the climate stochastic prediction. In Chapter 2 a general algorithm for the estimation of confidence intervals of parameters of interest is investigated. This algorithm was discovered heuristically in a previous study and was used for estimating hydroclimatic uncertainty in a stochastic framework. Here it is compared with other general algorithms and it is found that it performs satisfactorily. Properties of the algorithm are discovered within an analytical framework strengthening the arguments in favour of its use in this earlier study.

Following the encouraging results of Chapter 2 we decide to head to the study of stochastic models, however following the Bayesian paradigm. In Chapter 3 the properties of three estimators of the parameters of the HKp are examined analytically. One estimator is novel. A simulation study is performed and it is shown that these estimators have some optimal properties. The optimal properties of the maximum likelihood estimator are encouraging in the sense that the likelihood principle is followed by the Bayesian approach. Furthermore it is shown that the parameters of the HKp must be handled simultaneously, and we should avoid examining them separately.

Chapter 4 is the main part of this thesis. The posterior predictive distributions of the climate variables are calculated conditional on past observations within a Bayesian stochastic framework. This framework contains a stationary stochastic process which exhibits HK behaviour. The examined variables are assumed to be normal or truncated normal. The results are compared with cases where some of the parameters are considered known and the effect of the uncertainty in their estimation is shown. The framework is applied to temperature, rainfall and runoff data from Greece and Europe.

An attempt to include the outputs of a deterministic model within the framework of the stochastic model is displayed in Chapter 5. To this end properties of the maximum likelihood estimator of the bivariate HKp are analysed. In this case the parameters of the framework are considered known and equal to their estimates. The framework is applied to global temperature and rainfall data and their corresponding GCMs prediction. Chapter 6

summarizes the analytical results on the technical level providing also some insights from their application in climate data.

2. An algorithm to construct Monte Carlo confidence intervals for an arbitrary function of probability distribution parameters

In this Chapter[1] an algorithm for calculating an exact confidence interval for a parameter of the location or scale family, based on a two-sided hypothesis test on the parameter of interest, using some pivotal quantities is analysed. We use this algorithm to calculate approximate confidence intervals for the parameter or a function of the parameter of one-parameter continuous distributions. After appropriate heuristic modifications of the algorithm we use it to obtain approximate confidence intervals for a parameter or a function of parameters for multi-parameter continuous distributions. The advantage of the algorithm is that it is general and gives a fast approximation of an exact confidence interval. Some asymptotic (analytical) results are shown which validate the use of the method under certain regularity conditions. In addition, numerical results of the method compare well with those obtained by other known methods of the literature on the exponential, the normal, the gamma and the Weibull distribution.

The algorithm of the method was derived by Koutsoyiannis and Kozanis (2005) and is a main tool of the statistical software Hydrognomon (Itia research group 2009-2012). Koutsoyiannis et al. (2007) used the algorithm as an intuitive tool without mathematical proofs in an attempt to form a stochastic framework for climate prediction and quantification of the prediction uncertainty.

2.1 Introduction

Various general methods for the calculation of a confidence interval for a parameter of interest exist. Casella and Berger (2001 p.496-497) suggest the use of the asymptotic distribution of the maximum likelihood estimator (MLE) to construct a confidence interval for a function of the parameter of a one-parameter distribution. Wilks (1938) constructs a confidence interval based on the score statistic (see also Casella and Berger 2001 p.498). Kite (1988) gives approximate confidence intervals for the parameters of various distributions, by performing separate analyses for each distribution and each parameter estimation method. Garthwaite and Buckland (1992) make a new use of the Robbins-Monro search process to generate Monte Carlo confidence intervals for a one-parameter probability distribution. The Jacknife method is another general technique to obtain confidence intervals (see e.g. Román-

[1] Based on: Tyralis et al. (2013)

Montoya et al. 2008). Ripley (1987 p.176-178) constructs simple Monte Carlo confidence intervals which depend on the type of local properties (location or scale) of the parameter of interest.

In this Chapter we generalize the method proposed by Ripley (1987), retaining its simplicity. The method we study here incorporates Ripley's two suggested equations into one new equation. The method has a general character and does not make a distinction between the location and scale families, while other methods make such distinction. It provides single results without requiring user choices. These are strong advantages which make the proposed method a useful statistical computation tool.

Initially, we show that our algorithm is asymptotically equivalent to a Wald-type interval, i.e. an interval resulting from the inversion of a Wald test (Casella and Berger 2001 p.499) of a parameter or a function of a parameter of any one-parameter probability distribution. We also show how this algorithm works for certain distributions. Then we generalize this new algorithm to construct confidence intervals for the parameters or functions of parameters for multi-parameter probability distributions. We show that these intervals are asymptotically equivalent to Wald-type intervals. We also show analytically how this algorithm works for the normal distribution. We compare the results of the algorithm with those obtained by other exact and approximate methods for the exponential, normal, gamma and Weibull distributions, and it turns out that the algorithm works well even for small samples. The approximate methods described here include Wald-type intervals given in the literature or derived using the formula in Casella and Berger (2001 p.497), Ripley's two equations, and bias-corrected and accelerated (BCa) bootstrap non-parametric intervals (see also DiCiccio 1984; Di Ciccio and Efron 1996; Di Ciccio and Romano 1995; Efron 1987; Efron and Tibshirani 1993; Hall 1988; Kisielinska 2012).

The proposed algorithm is partly heuristic and simultaneously so general that it needs no assumptions about the statistical behaviour of the statistics under study, i.e. it can perform for any continuous distribution with any number of parameters, and for any distributional or derivative parameter. Only the theoretical probabilistic model is needed and all other calculations are done by a number of Monte Carlo simulations without additional assumptions.

2.2 Terminology and notation

We use the terminology of Casella and Berger (2001). We recall that an interval estimate of a parameter $\theta \in R$ is any pair of functions, $l(\boldsymbol{x})$ and $u(\boldsymbol{x})$, of a sample $\boldsymbol{x}_{1:n} = (x_1,\ldots,x_n)$ that satisfy $l(\boldsymbol{x}_{1:n}) \leq u(\boldsymbol{x}_{1:n})$ for all $\boldsymbol{x}_{1:n}$. If $\underline{\boldsymbol{x}}_{1:n}$ is the random variable, consisting of i.i.d random variables, whose realization is $\boldsymbol{x}_{1:n}$, the inference $l(\boldsymbol{x}_{1:n}) \leq \theta \leq u(\boldsymbol{x}_{1:n})$ is made. The random interval $[l(\underline{\boldsymbol{x}}_{1:n}),u(\underline{\boldsymbol{x}}_{1:n})]$ is called an interval estimator.

The following result from Casella and Berger (2001 p.421,422) is necessary for the proofs of Section 2.3 and shows how we can construct a confidence interval from a hypothesis testing procedure:

For each $\theta_0 \in \Theta \subseteq R$, let $A(\theta_0)$ be the acceptance region of a level α test of H_0: $\theta = \theta_0$. For each $\boldsymbol{x}_{1:n}$, we define an interval $C(\boldsymbol{x}_{1:n})$ in the parameter space by

$$C(\boldsymbol{x}_{1:n}) = \{\theta_0: \boldsymbol{x}_{1:n} \in A(\theta_0)\} \tag{2.1}$$

Then the random set $C(\underline{\boldsymbol{x}}_{1:n})$ is a $1 - \alpha$ confidence interval. Conversely, let $C(\underline{\boldsymbol{x}}_{1:n})$ be a $1 - \alpha$ confidence interval. For any $\theta_0 \in \Theta$, we define

$$A(\theta_0) = \{\boldsymbol{x}_{1:n}: \theta_0 \in C(\boldsymbol{x}_{1:n})\} \tag{2.2}$$

Then $A(\theta_0)$ is the acceptance region of a level α test of H_0: $\theta = \theta_0$. For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a level α test if $\sup_{\theta\in\Theta_0} \beta(\theta) \leq \alpha$. If $\sup_{\theta\in\Theta_0} \beta(\theta) = \alpha$ then this is a size α test which is a special case of the level α test (Casella and Berger 2001 p.385). Note that the above terminology is not precise when the test is randomized (Shao 2003 p.477).

2.3 Construction of confidence intervals for one-parameter distributions

Now we proceed to the construction of a confidence interval for one-parameter continuous probability distributions. The following result which is a consequence of (2.1) and (2.2) is necessary for the construction of the confidence interval.

We suppose that $\underline{b} := b(\underline{\boldsymbol{x}}_{1:n})$ is a MLE of the parameter θ of a one-parameter distribution with density $f(x|\theta)$. Then $\hat{\theta} = b(\boldsymbol{x}_{1:n})$ is the estimate of the parameter. We suppose now that the probability density of the statistic $b(\underline{\boldsymbol{x}}_{1:n})$ is $g(b|\theta)$. Then we seek two functions $\lambda(\theta)$, $\upsilon(\theta)$ such that:

$$P\{\lambda(\theta) \leq b(\underline{\boldsymbol{x}}_{1:n}) \leq \upsilon(\theta)\} = 1 - \alpha \tag{2.3}$$

We define $\lambda(\theta)$, $\upsilon(\theta)$ as those functions that satisfy:

$$\Pr\{b(\underline{\boldsymbol{x}}_{1:n}) < \lambda(\theta)\} = \Pr\{b(\underline{\boldsymbol{x}}_{1:n}) > \upsilon(\theta)\} = \alpha/2 \tag{2.4}$$

The above equation implies that:

$$\lambda(\theta) = G^{-1}(\alpha/2|\theta) \text{ and } \upsilon(\theta) = G^{-1}(1-\alpha/2|\theta) \tag{2.5}$$

where $G^{-1}(\cdot|\theta)$ denotes the inverse of the cumulative distribution function (or distribution function from this point forward) G of the statistic $b(\underline{\boldsymbol{x}}_{1:n})$.

Now we construct a test H_0: $\theta = \hat{\theta}$ vs H_1: $\theta \neq \hat{\theta}$ with acceptance region:

$$A(\hat{\theta}) = \{\boldsymbol{x}_{1:n}: G^{-1}(\alpha/2|\hat{\theta}) \leq b(\boldsymbol{x}) \leq G^{-1}(1-\alpha/2|\hat{\theta})\} \tag{2.6}$$

which is a size α test because the value of the power function at $\hat{\theta}$ is $\beta(\hat{\theta})$, given by (2.7).

$$\beta(\hat{\theta}) = 1 - \Pr\{G^{-1}(\alpha/2|\hat{\theta}) \leq b(\underline{\boldsymbol{x}}_{1:n}) \leq G^{-1}(1-\alpha/2|\hat{\theta})|\theta = \hat{\theta})\} \tag{2.7}$$

Thus

$$\beta(\hat{\theta}) = 1 - [\mathrm{G}(G^{-1}(1-\alpha/2|\hat{\theta})|\hat{\theta}) - \mathrm{G}(G^{-1}(\alpha/2|\hat{\theta})|\hat{\theta})] = 1 - (1-\alpha/2-\alpha/2) = \alpha \tag{2.8}$$

From this test and according to (2.1) and (2.2) we obtain the following $1-\alpha$ confidence interval for θ:

$$C(\boldsymbol{x}_{1:n}) = \{\hat{\theta}: G^{-1}(\alpha/2|\hat{\theta}) \leq b(\boldsymbol{x}_{1:n}) \leq G^{-1}(1-\alpha/2|\hat{\theta})\} \tag{2.9}$$

After rewriting (2.9) we obtain the following $1-\alpha$ confidence interval for θ:

$$C(\boldsymbol{x}_{1:n}) = \{\theta: G^{-1}(\alpha/2|\theta) \leq b(\boldsymbol{x}_{1:n}) \leq G^{-1}(1-\alpha/2|\theta)\} \tag{2.10}$$

Now we define l and u as the solutions of the equations:

$$\upsilon(l) = b(\boldsymbol{x}_{1:n}) \text{ and } \lambda(u) = b(\boldsymbol{x}_{1:n}) \tag{2.11}$$

From the above equation we obtain that:

$$G^{-1}(\alpha/2|u) = b(\boldsymbol{x}_{1:n}) \text{ and } G^{-1}(1-\alpha/2|l) = b(\boldsymbol{x}_{1:n}) \tag{2.12}$$

We assume that $C(\boldsymbol{x}_{1:n}) = [\theta_1,\theta_2]$ where θ_1,θ_2 are the solutions of the equations

$$G^{-1}(\alpha/2|\theta_2) = b(\boldsymbol{x}_{1:n}) \text{ and } G^{-1}(1-\alpha/2|\theta_1) = b(\boldsymbol{x}_{1:n}) \tag{2.13}$$

Now it is obvious that $[l,u]$ is a $1-\alpha$ confidence interval estimate for θ.

2.3.1 Construction of the confidence interval

Having proved that $[l,u]$ is a $1-\alpha$ confidence interval estimate for θ, we can use it to construct an approximate confidence interval that can be easily computed numerically. From Figure 2.1 we observe that

$$\frac{\upsilon(\hat{\theta})-\hat{\theta}}{\hat{\theta}-l}=\frac{\mathrm{CA}}{\mathrm{CB}}\approx\left(\frac{\mathrm{d}\upsilon}{\mathrm{d}\theta}\right)_{|\theta=\hat{\theta}} \tag{2.14}$$

Solving for l we find

$$l\approx\hat{\theta}+\frac{\hat{\theta}-\upsilon(\hat{\theta})}{(\mathrm{d}\upsilon/\mathrm{d}\theta)_{|\theta=\hat{\theta}}} \tag{2.15}$$

and in a similar way we find that

$$u\approx\hat{\theta}+\frac{\hat{\theta}-\lambda(\hat{\theta})}{(\mathrm{d}\lambda/\mathrm{d}\theta)_{|\theta=\hat{\theta}}} \tag{2.16}$$

We can thus claim that

$$[l(\underline{\boldsymbol{x}}),u(\underline{\boldsymbol{x}})]=[b(\underline{\boldsymbol{x}})+\frac{b(\underline{\boldsymbol{x}})-\upsilon(b(\underline{\boldsymbol{x}}))}{(\mathrm{d}\upsilon/\mathrm{d}\theta)_{|\theta=b(\underline{\boldsymbol{x}})}},\ b(\underline{\boldsymbol{x}})+\frac{b(\underline{\boldsymbol{x}})-\lambda(b(\underline{\boldsymbol{x}}))}{(\mathrm{d}\lambda/\mathrm{d}\theta)_{|\theta=b(\underline{\boldsymbol{x}})}}] \tag{2.17}$$

is an approximate $1-\alpha$ confidence interval for θ.

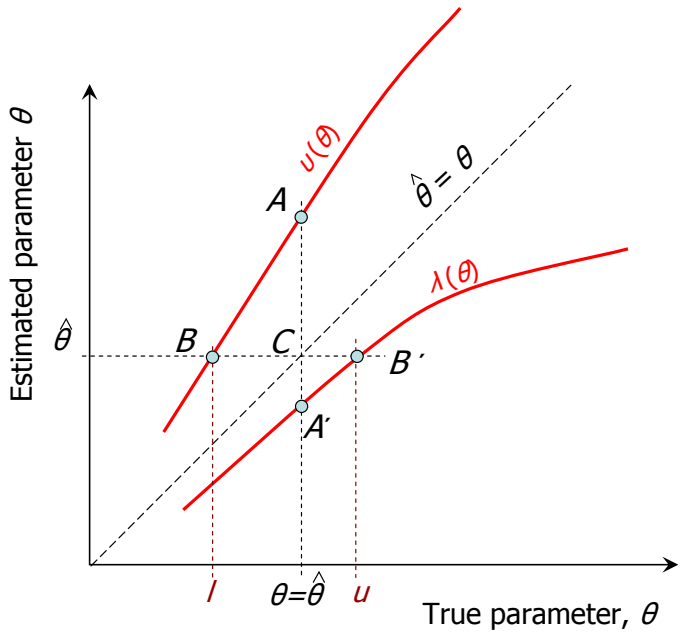


Figure 2.1. Sketch explaining the determination of confidence limits l and u from an inversion of a hypothesis test.

Under suitable regularity conditions (i.e. Casella and Berger 2001 p.516) the density of the

MLE is given by Hillier and Armstrong (1999). The necessary conditions for the equations (2.15) and (2.16) to hold are that λ and υ are continuous and differentiable at a region of $\hat{\theta}$. The validity of these assumptions for certain cases could be investigated, using Hillier and Armstrong (1999) formula, but at such situations this is not always possible.

2.3.2 Some theoretical results

It is useful to find cases where the above derived confidence interval is exact (i.e. 2.15 and 2.16 are exact). We can easily prove that this happens in the case where $\upsilon(\theta) = c_1\theta + c_2$, where c_1 and c_2 are any real numbers:

$$(\frac{\mathrm{d}\upsilon}{\mathrm{d}\theta})_{|\theta=\hat{\theta}} = c_1, \text{ and } \frac{\upsilon(\hat{\theta}) - \hat{\theta}}{\hat{\theta} - l} = \frac{c_1\hat{\theta} + c_2 - \hat{\theta}}{\hat{\theta} - [(\hat{\theta} - c_2)/c_1]} = c_1 \tag{2.18}$$

(The proof for u can be conducted in a similar way and is omitted). Special cases of this are (i) when $\upsilon(\theta) = \theta + c$, and (ii) when $\upsilon(\theta) = c\theta$. These two correspond to the first and second methods described by Ripley (1987) respectively (p.176, eq.3 and p.177, eq.6, after substitution of $(\frac{\mathrm{d}\upsilon}{\mathrm{d}\theta})_{|\theta=\hat{\theta}} = c = \upsilon(\theta)/\theta$ in (2.17)). We can also easily prove that location families correspond to the first case and scale families correspond to the second case. The proof is given below:

(a) For location families the quantity $\underline{\mu} - \mu$ (where $\underline{\mu}$ is a MLE of the location parameter μ) is a pivotal quantity, i.e. its distribution does not depend on unknown parameters (see Lawless 2003 p.562). Then from (2.4) we have that

$$\Pr\{\underline{\mu} < \lambda(\mu)\} = \alpha/2 \tag{2.19}$$

which implies that

$$\Pr\{\underline{\mu} - \mu < \lambda(\mu) - \mu\} = \alpha/2 \tag{2.20}$$

and we obtain that

$$\lambda(\mu) = \mu + G^{-1}(\alpha/2) \tag{2.21}$$

where G is the distribution function of $\underline{\mu} - \mu$ that does not depend on μ. In a similar way we obtain that

$$\upsilon(\mu) = \mu + G^{-1}(1 - \alpha/2) \tag{2.22}$$

Now it is obvious from (i) above that the confidence interval obtained by (2.17) is an exact confidence interval.

(b) For scale families the quantity $\underline{\sigma}/\sigma$ (where $\underline{\sigma}$ is a MLE of the location parameter σ) is a pivotal quantity (see Lawless 2003 p.562). Then from (2.4) we have that

$$\Pr\{\underline{\sigma} < \lambda(\sigma)\} = \alpha/2 \tag{2.23}$$

which implies that

$$\Pr\{\underline{\sigma}/\sigma < \lambda(\sigma)/\sigma\} = \alpha/2 \tag{2.24}$$

and we obtain that

$$\lambda(\sigma) = \sigma\, G^{-1}(\alpha/2) \tag{2.25}$$

where G is the distribution function of $\underline{\sigma}/\sigma$ and is independent of σ. In a similar way we obtain that

$$\upsilon(\sigma) = \sigma\, G^{-1}(1 - \alpha/2) \tag{2.26}$$

Now it is obvious from (ii) above that the confidence interval obtained by (2.17) is an exact confidence interval.

While in the above cases our method provides exact confidence intervals, when the equation $\upsilon(\theta) = c_1\theta + c_2$ does not hold, it can only provide approximate confidence intervals, where the level of approximation depends on the form of λ and υ and for certain cases will be examined in the next Sections. It is also easy to prove that the confidence interval given by (2.17) is asymptotically equivalent to a Wald-type interval for any function of the parameter θ (and hence for the parameter itself) under certain regularity conditions. The proof is given below.

We want to find a confidence interval for a function $h(\theta)$ of θ. We assume that $\underline{\theta}$ is a MLE of θ. Then according to Casella and Berger (2001 p.497) and Efron and Hinkley (1978), the variance of the function $h(\underline{\theta})$ can be approximated by

$$\hat{\mathrm{Var}}[h(\underline{\theta})|\theta] \approx \frac{[h'(\theta)]^2|_{\theta=\hat{\theta}}}{-\dfrac{\partial^2}{\partial\theta^2}\ln l(\theta|\boldsymbol{x}_{1:n})|_{\theta=\hat{\theta}}} \tag{2.27}$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ and $l(\theta|\boldsymbol{x}_{1:n})$ the likelihood function of θ. Now according to Casella and Berger (2001 p.497) we have the following result:

$$\frac{h(\underline{\theta}) - h(\theta)}{\sqrt{\hat{\mathrm{V}}\mathrm{ar}[h(\underline{\theta})|\theta]}} \rightarrow \mathrm{N}(0,1) \tag{2.28}$$

Then from (2.4) we get that

$$\Pr\{h(\underline{\theta}) < \lambda(\theta)\} = \alpha/2 \tag{2.29}$$

and

$$\Pr\{h(\underline{\theta}) > \upsilon(\theta)\} = \alpha/2 \tag{2.30}$$

which imply that

$$\lambda(\theta) = h(\theta) + \sqrt{\hat{\mathrm{V}}\mathrm{ar}[h(\underline{\theta})|\theta]}\Phi^{-1}(\alpha/2) \text{ and } \upsilon(\theta) = h(\theta) + \sqrt{\hat{\mathrm{V}}\mathrm{ar}[h(\underline{\theta})|\theta]}\Phi^{-1}(1-\alpha/2) \tag{2.31}$$

where Φ^{-1} denotes the inverse of the standard normal distribution function. If we substitute θ for $h(\theta)$ then (2.28) becomes identical to case (i) above.

2.3.3 Construction of the algorithm

Having found an expression for the confidence interval, we can construct a Monte Carlo algorithm to calculate it when there do not exist analytical expressions for the functions of interest. The algorithm has the following steps:

Step 1. We find the maximum likelihood estimate say $\hat{\theta}$.

Step 2. We produce k samples of size n, from $f(x|\hat{\theta})$.

Step 3. We use these k samples to compute $\lambda(\hat{\theta})$ and $\upsilon(\hat{\theta})$.

Step 4. We produce additional k samples of size n, from $f(x|\hat{\theta}+\delta\theta)$, where $\delta\theta$ is a small increment.

Step 5. We use these additional k samples to compute $\lambda(\hat{\theta}+\delta\theta)$ and $\upsilon(\hat{\theta}+\delta\theta)$.

Step 6. We substitute $(\frac{\mathrm{d}\upsilon}{\mathrm{d}\theta})_{|\theta = \hat{\theta}}$ of (2.15) with $[\upsilon(\hat{\theta}+\delta\theta) - \upsilon(\hat{\theta})]/\delta\theta$, and $(\frac{\mathrm{d}\lambda}{\mathrm{d}\theta})_{|\theta = \hat{\theta}}$ of (2.16) with $[\lambda(\hat{\theta}+\delta\theta) - \lambda(\hat{\theta})]/\delta\theta$.

Step 7. We compute l and u from (2.15) and (2.16).

We conclude based on the construction of the algorithm that it could be applied to cases where θ is estimated by a different estimator. Below we give an application of the algorithm on the normal distribution where we used the unbiased estimator of θ and obtained good results.

2.4 Construction of confidence intervals for multi-parameter probability distributions

We assume now that we have a multi-parameter probability distribution with density $f(x|\boldsymbol{\theta})$ and parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$, whose estimator is $\underline{\boldsymbol{T}} = (\underline{T}_1, \underline{T}_2, \ldots, \underline{T}_k)$. We wish to calculate a $1 - \alpha$ confidence interval for a scalar function $\beta := h(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$. If we assume that $\underline{\boldsymbol{T}}$ is a MLE then $b(\underline{\boldsymbol{x}})$ given by (2.32)

$$b(\underline{\boldsymbol{x}}) := h(\underline{\boldsymbol{T}}) \tag{2.32}$$

is a MLE for $h(\boldsymbol{\theta})$ and $b(\boldsymbol{x}_{1:n})$ given by (2.33)

$$b(\boldsymbol{x}_{1:n}) = h(\boldsymbol{t}) \tag{2.33}$$

is its estimate. To extend the method, described by (2.15) and (2.16) in the multiple parameter case, the derivatives $\mathrm{d}\lambda/\mathrm{d}\theta$ and $\mathrm{d}\upsilon/\mathrm{d}\theta$ should be evaluated in appropriate directions $\boldsymbol{d}_\lambda$ and $\boldsymbol{d}_\upsilon$.

Let $\boldsymbol{\gamma}$ be defined by (2.34)

$$\boldsymbol{\gamma} := (\lambda, \beta, \upsilon)^{\mathrm{T}} \tag{2.34}$$

where λ, υ have been defined by (2.4) and let

$$\mathrm{Var}[\underline{\boldsymbol{T}}] = \mathrm{diag}(\mathrm{Var}[\underline{T}_1], ..., \mathrm{Var}[\underline{T}_k]) \tag{2.35}$$

$\mathrm{Var}[\underline{\boldsymbol{T}}]$ can be easily computed during the same Monte Carlo simulation that is performed to compute $\boldsymbol{\gamma}$. It is reasonable to assume that $\boldsymbol{d}_\lambda$ and $\boldsymbol{d}_\upsilon$ will depend on $\mathrm{Var}[\underline{\boldsymbol{T}}]$ as well as on the matrix of derivatives of $\boldsymbol{\gamma}$,

$$\frac{\mathrm{d}\boldsymbol{\gamma}}{\mathrm{d}\boldsymbol{\theta}} = \begin{bmatrix} \dfrac{\mathrm{d}\lambda}{\mathrm{d}\boldsymbol{\theta}} \\ \dfrac{\mathrm{d}\beta}{\mathrm{d}\boldsymbol{\theta}} \\ \dfrac{\mathrm{d}\upsilon}{\mathrm{d}\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial\lambda}{\partial\theta_1} & \dfrac{\partial\lambda}{\partial\theta_2} & \cdots & \dfrac{\partial\lambda}{\partial\theta_k} \\ \dfrac{\partial\beta}{\partial\theta_1} & \dfrac{\partial\beta}{\partial\theta_2} & \cdots & \dfrac{\partial\beta}{\partial\theta_k} \\ \dfrac{\partial\upsilon}{\partial\theta_1} & \dfrac{\partial\upsilon}{\partial\theta_2} & \cdots & \dfrac{\partial\upsilon}{\partial\theta_k} \end{bmatrix} \tag{2.36}$$

Heuristically, we can assume a simple relation of the form

$$\boldsymbol{d}_\lambda = \mathrm{Var}[\underline{\boldsymbol{T}}] \left(\frac{\mathrm{d}\boldsymbol{\gamma}}{\mathrm{d}\boldsymbol{\theta}}\right)^{\mathrm{T}} \boldsymbol{e}_\lambda \tag{2.37}$$

where $\boldsymbol{e}_\lambda$ is a size 3 vector of constants needed to transform the matrix product of the first two terms of the right hand side into a vector. The elements of this vector could be thought of as weights corresponding to each of the derivatives of the three elements of $\boldsymbol{\gamma}$. The simplest choice is to assume equal weights, i.e.

$$\boldsymbol{e}_\lambda = (1, 1, 1)^{\mathrm{T}} \tag{2.38}$$

However, numerical investigations showed that (2.39)

$$\boldsymbol{e}_\lambda = (0, 1, 1)^{\mathrm{T}} \tag{2.39}$$

yields better approximations and the theoretical analysis below showed that it yields asymptotically good results under certain regularity conditions.

The derivatives of λ and β with respect to $\boldsymbol{\theta}$ on direction $\boldsymbol{d}_\lambda$ will then be

$$\left(\frac{\mathrm{d}\lambda}{\mathrm{d}\boldsymbol{\theta}}\right)\boldsymbol{d}_\lambda = \left(\frac{\mathrm{d}\lambda}{\mathrm{d}\boldsymbol{\theta}}\right)\mathrm{Var}[\underline{\boldsymbol{T}}]\left(\frac{\mathrm{d}\boldsymbol{\gamma}}{\mathrm{d}\boldsymbol{\theta}}\right)^{\mathrm{T}}\boldsymbol{e}_\lambda, \quad \left(\frac{\mathrm{d}\beta}{\mathrm{d}\boldsymbol{\theta}}\right)\boldsymbol{d}_\lambda = \left(\frac{\mathrm{d}\beta}{\mathrm{d}\boldsymbol{\theta}}\right)\mathrm{Var}[\underline{\boldsymbol{T}}]\left(\frac{\mathrm{d}\boldsymbol{\gamma}}{\mathrm{d}\boldsymbol{\theta}}\right)^{\mathrm{T}}\boldsymbol{e}_\lambda \tag{2.40}$$

and are both scalars, so by taking their ratio we can calculate $\mathrm{d}\lambda/\mathrm{d}\beta$. By symmetry, similar relationships can be written for υ and $\boldsymbol{d}_\upsilon$ with

$$\boldsymbol{e}_\upsilon = (1, 1, 0)^{\mathrm{T}} \tag{2.41}$$

The two groups of relationships can be unified in terms of the 3×3 matrix $\boldsymbol{q}$ defined as

$$\boldsymbol{q} := \frac{\mathrm{d}\boldsymbol{\gamma}}{\mathrm{d}\boldsymbol{\theta}}\mathrm{Var}[\underline{\boldsymbol{T}}]\left(\frac{\mathrm{d}\boldsymbol{\gamma}}{\mathrm{d}\boldsymbol{\theta}}\right)^{\mathrm{T}} \tag{2.42}$$

It can then be easily shown that on the directions $\boldsymbol{d}_\lambda$ and $\boldsymbol{d}_\upsilon$,

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\beta} = \frac{q_{12} + q_{13}}{q_{22} + q_{23}}, \quad \frac{\mathrm{d}\upsilon}{\mathrm{d}\beta} = \frac{q_{31} + q_{32}}{q_{21} + q_{22}} \tag{2.43}$$

In Section 2.7 we show that the confidence interval for the parameter μ of a normal distribution $\mathrm{N}(\mu,\sigma^2)$ is asymptotically equivalent to a Wald-type interval. We also show that the confidence interval obtained by our method is asymptotically equivalent to a Wald-type interval for two-parameter regular distributions and hence for any multi-parameter distribution.

2.4.1 Construction of the algorithm

Now the algorithm for the calculation of the intervals follows:

Step 1. We find the MLE of $\boldsymbol{\theta}$, namely $\underline{\boldsymbol{\theta}}$, and its maximum likelihood estimate say $\hat{\boldsymbol{\theta}}$.

Step 2. The MLE of β is $h(\underline{\boldsymbol{\theta}})$, and its maximum likelihood estimate is $h(\hat{\boldsymbol{\theta}})$.

Step 3. We produce m samples of size n, from $f(x|\hat{\boldsymbol{\theta}})$.

Step 4. We use these m samples to compute $\lambda(\hat{\boldsymbol{\theta}})$, $\upsilon(\hat{\boldsymbol{\theta}})$, $h(\hat{\boldsymbol{\theta}})$ and $\mathrm{Var}[\underline{\boldsymbol{T}}]$.

Step 5. We produce additional m samples of size n, from $f(x|\hat{\boldsymbol{\theta}}+\delta\boldsymbol{\theta}_i)$, where $\delta\boldsymbol{\theta}_i$ is a vector with all elements zero except the ith element, which is a small quantity $\delta\theta_i$.

Step 6. We use these additional m samples to compute $\lambda(\hat{\boldsymbol{\theta}}+\delta\boldsymbol{\theta}_i)$, $\upsilon(\hat{\boldsymbol{\theta}}+\delta\boldsymbol{\theta}_i)$ and $h(\hat{\boldsymbol{\theta}}+\delta\boldsymbol{\theta}_i)$.

Step 7. We repeat steps 4 and 5 for $i = 1,2,\ldots,k$.

Step 8. We substitute in (2.36) $[\lambda(\hat{\boldsymbol{\theta}}+\delta\boldsymbol{\theta}_i) - \lambda(\hat{\boldsymbol{\theta}})]/\delta\theta_i$ for $\frac{\partial\lambda}{\partial\theta_i}$, $[\upsilon(\hat{\boldsymbol{\theta}}+\delta\boldsymbol{\theta}_i) - \upsilon(\hat{\boldsymbol{\theta}})]/\delta\theta_i$ for $\frac{\partial\upsilon}{\partial\theta_i}$ and $[h(\hat{\boldsymbol{\theta}}+\delta\boldsymbol{\theta}_i) - h(\hat{\boldsymbol{\theta}})]/\delta\theta_i$ for $\frac{\partial h}{\partial\theta_i}$.

Step 9. We calculate $\boldsymbol{q}$ from (2.42).

Step 10. We compute l and u from (2.15) and (2.16).

2.5 Simulation results

To test the algorithm in specific cases, we construct confidence intervals for the scale parameter of the exponential distribution, the location parameter and the pth percentile of the normal distribution, the scale and shape parameter of the gamma distribution and the scale parameter and the pth percentile of the Weibull distribution. Then we compare the numerical results with known, mostly analytical, results from the literature. Various methods are first compared using a single sample but the ranking based on visual inspection could be considered as subjective. Thus coverage probabilities using Monte Carlo methods are also calculated to obtain a more objective inference.

2.5.1 Confidence interval for the scale parameter of the exponential distribution

The density of the exponential distribution is

$$f_{\text{EXP}}(x|\sigma) = (1/\sigma)\exp(-x/\sigma),\ x \geq 0,\ \sigma > 0 \tag{2.44}$$

The MLE of σ is

$$\underline{\sigma} = \underline{x}_1^{(n)} \tag{2.45}$$

A $1-\alpha$ Wald-type confidence interval (Papoulis and Pillai 2002 p.310), is

$$[l(\underline{\boldsymbol{x}}_{1:n}),u(\underline{\boldsymbol{x}}_{1:n})] = [\frac{\underline{x}_1^{(n)}}{1+\Phi^{-1}(1-\alpha/2)/\sqrt{n}},\ \frac{\underline{x}_1^{(n)}}{1-\Phi^{-1}(1-\alpha/2)/\sqrt{n}}] \tag{2.46}$$

We find a $1-\alpha$ exact confidence interval, using the pivotal quantity $\underline{\sigma}/\sigma$. The distribution of $\underline{\sigma}$ is given by (2.47)

$$\underline{\sigma} \sim \text{gamma}(n,n/\sigma) \tag{2.47}$$

and a $1-\alpha$ exact confidence interval is obtained by the following equations.

$$F_G(x_1^{(n)}|n,n/l) = 1-\alpha/2,\ F_G(x_1^{(n)}|n,n/u) = \alpha/2 \tag{2.48}$$

where $F_G(x|\alpha,\beta)$ is the gamma distribution whose density is

$$f_G(x|\alpha,\beta) = \beta^{\alpha}\,[\Gamma(\alpha)]^{-1}\,x^{\alpha-1}\exp(-\beta x),\ x>0 \tag{2.49}$$

where $\alpha > 0$ is the shape parameter and $1/\beta > 0$ is the scale parameter.

The confidence interval obtained by (2.48) is exact and the confidence interval obtained by (2.46) is Wald-type. These two are intercompared also with the BCa bootstrap non-parametric interval, designated as "bootstrap", the two confidence intervals obtained by the two Ripley's methods, designated as "Ripley location" and "Ripley scale", respectively, and the confidence interval obtained by our algorithm, designated as MCCI (Monte Carlo Confidence Interval). Figure 2.2 compares the confidence intervals obtained by all six methods for a simulated sample with 50 elements from an exponential distribution with $\sigma = 1$. For this sample $\hat{\sigma} = 1.002$. As we see, MCCI is close to the exact and the "Ripley scale" and gives a better approximation than the Wald-type, the "bootstrap" and the "Ripley location".

2.5.2 Confidence interval for the location parameter of the normal distribution

The density of the normal distribution is

$$f_N(x|\mu,\sigma) = (2\pi\sigma^2)^{-1/2}\exp[(-1/2\sigma^2)\,(x-\mu)^2] \tag{2.50}$$

where μ is the location parameter, and $\sigma > 0$ is the scale parameter. A $1-\alpha$ exact confidence interval (Papoulis and Pillai 2002 p.309) is

$$[l(\boldsymbol{x}_{1:n}),u(\boldsymbol{x}_{1:n})] = [x_1^{(n)} - F^{-1}_{T(n-1)}(1-\alpha/2)\,\frac{s_n}{\sqrt{n}},\ x_1^{(n)} + F^{-1}_{T(n-1)}(1-\alpha/2)\,\frac{s_n}{\sqrt{n}}] \tag{2.51}$$

where

$$s_n := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - x_1^{(n)})^2} \tag{2.52}$$

A $1-\alpha$ Wald-type confidence interval (Papoulis and Pillai 2002 p.309) is

$$[l(\boldsymbol{x}_{1:n}),u(\boldsymbol{x}_{1:n})] = [x_1^{(n)} - \Phi^{-1}(1-\alpha/2)\,\frac{s_n}{\sqrt{n}},\ x_1^{(n)} + \Phi^{-1}(1-\alpha/2)\,\frac{s_n}{\sqrt{n}}] \tag{2.53}$$

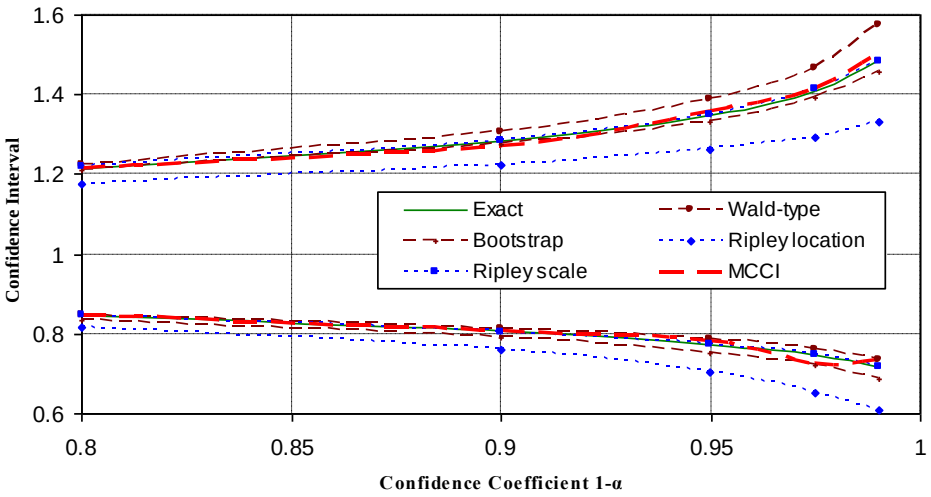


Figure 2.2. Confidence intervals for the scale of an exponential distribution with $n = 50$, $\hat{\sigma} = 1.002$. Here the number of samples $k = 50\,000$ for MCCI, "Ripley location" and "Ripley scale" cases and $\delta\sigma = 0.05$.

We compare the MCCI with the exact interval obtained by (2.51), as well as with the Wald-type interval, the BCa interval and the intervals obtained by Ripley's two methods. Figure 2.3 compares the confidence intervals obtained by the six methods for a simulated sample with 10 elements from a normal distribution with $\mu = 0$ and $\sigma = 1$. For this sample $\hat{\mu} = 0.026$ and $\hat{\sigma} = 1.023$. In this case for the calculation of the confidence interval we use the unbiased estimators of μ and σ^2 (instead of the MLE). As we see, MCCI gives a better approximation than the other four approximate methods.

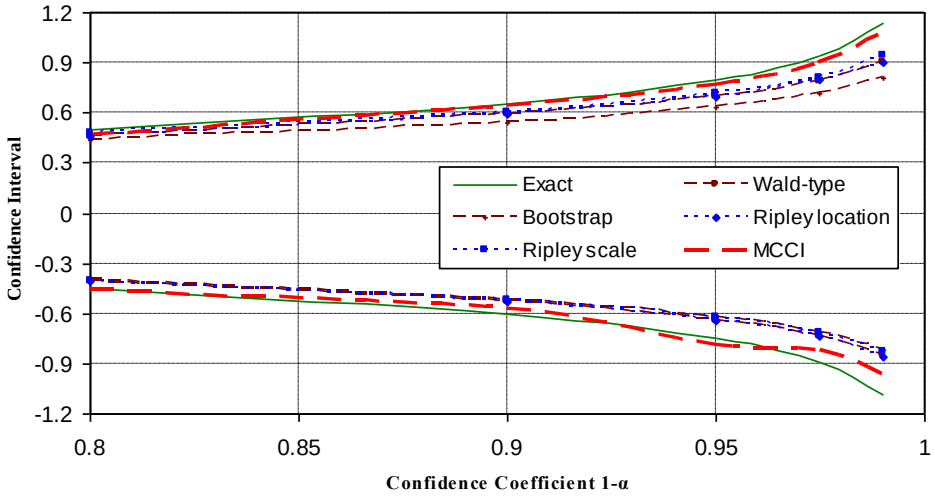


Figure 2.3. Confidence intervals for the location parameter of a normal distribution with $n = 10$, $\hat{\mu} = 0.026$ and $\hat{\sigma} = 1.023$. Here the number of samples $k = 100\,000$ for MCCI, "Ripley location" and "Ripley scale" cases, $\delta\mu = 0.1$ and $\delta\sigma = 0.1$.

2.5.3 Confidence interval for the percentile of the normal distribution

The pth percentile t_p is defined by (2.54)

$$t_p := \mu + z_p\sigma \tag{2.54}$$

where z_p is pth percentile of the standard normal distribution. A $1-\alpha$ Wald-type confidence interval estimate is given by the following equation (e.g. Koutsoyiannis 1997 p.69)

$$[l(\boldsymbol{x}_{1:n}),u(\boldsymbol{x}_{1:n})] = [x_1^{(n)}+z_p s_n-\Phi^{-1}(1-\alpha/2)\frac{s_n}{\sqrt{n}}\sqrt{1+z_p^2/2},x_1^{(n)}+z_p s_n+\Phi^{-1}(1-\alpha/2)\frac{s_n}{\sqrt{n}}\sqrt{1+z_p^2/2}] \tag{2.55}$$

Another way to obtain a confidence interval is by using Bayesian analysis (see Gelman et al. 2004 p.75,76). Then if we chose a prior π defined by (2.56)

$$\pi(\mu,\sigma) \propto 1/\sigma^2 \tag{2.56}$$

we can construct a sampler based on the following mixture.

$$\underline{\sigma}^2|\boldsymbol{x}_{1:n} \sim \text{Inv-}\chi^2(n-1,s_n^2) \text{ and } \underline{\mu}|\sigma^2,\boldsymbol{x}_{1:n} \sim \text{N}(x_1^{(n)},\sigma^2/n) \tag{2.57}$$

Thus, here we compare six confidence intervals, the Bayesian confidence region, the Wald-type of equation (2.55), the BCa interval, the intervals obtained by Ripley's two methods, and the MCCI. Figure 2.4 compares the confidence intervals for $\mu + 2\sigma$ obtained by the six methods for a simulated sample with 50 elements from a normal distribution with μ = 0 and σ = 1. For this sample $\hat{\mu}$ = −0.027 and $\hat{\sigma}$ = 0.998. As we see, Bayesian and MCCI are almost indistinguishable, and MCCI is better when compared to "Ripley location" and "Ripley scale". "Ripley location" is close to the Wald-type and the "bootstrap". The same holds for Figure 2.5 which compares all methods on the calculation of a 1−0.01 confidence interval, when z_p varies from −3 to 3.

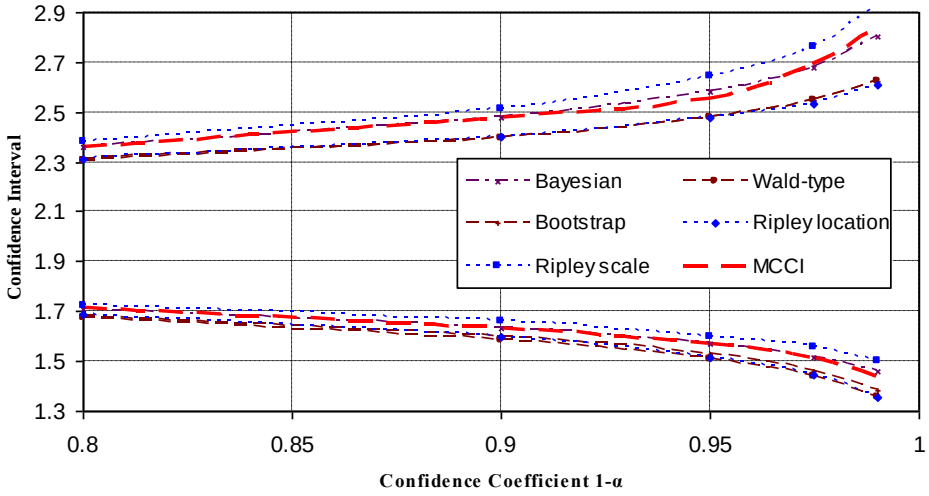


Figure 2.4. Confidence intervals for $\mu + 2\sigma$ of a normal distribution with n = 50, $\hat{\mu}$ = −0.027 and $\hat{\sigma}$ = 0.998. Here the number of samples m = 50 000 for MCCI, "Ripley location" and "Ripley scale" cases, $\delta\mu$ = 0.1 and $\delta\sigma$ = 0.1.

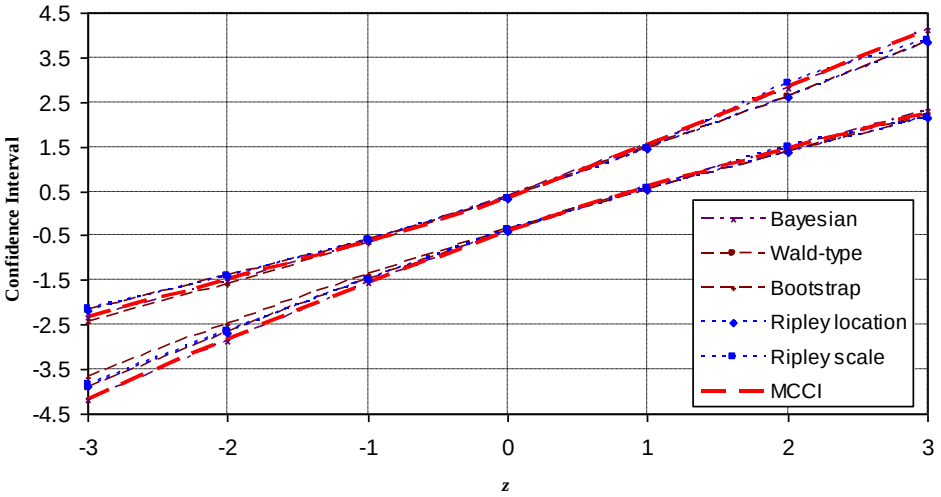


Figure 2.5. Confidence intervals with confidence coefficient 1−0.01 for $\mu + z\sigma$ of a normal distribution with $n = 50$, $\hat{\mu} = -0.027$ and $\hat{\sigma} = 0.998$. Here the number of samples $m = 50\,000$ for MCCI, "Ripley location" and "Ripley scale" cases, $\delta\mu = 0.1$ and $\delta\sigma = 0.1$.

2.5.4 Confidence interval for the scale parameter of the gamma distribution

First we show how we can calculate an approximate confidence interval for the scale parameter of the gamma distribution. We define $\underline{R}_n$ by (2.58)

$$\underline{R}_n := \ln(\underline{x}_1^{(n)}/\tilde{\underline{x}}) \tag{2.58}$$

where $\tilde{\underline{x}}$ is the geometric mean of a size-n sample, which, according to Bhaumik et al. (2009) and Bain and Engelhardt (1975), has a distribution independent of the scale parameter $\sigma = 1/\beta$. The maximum likelihood estimates of α and σ according to Bhaumik et al. (2009) and Choi and Wette (1969), denoted by $\hat{\alpha}$ and $\hat{\sigma}$ are the solutions of the equations (2.59)

$$R_n = \ln(\alpha) - \psi(\alpha) \text{ and } \alpha\sigma = x_1^{(n)} \tag{2.59}$$

where ψ denotes the digamma function.

We have that

$$\mathrm{E}[\underline{R}_n] = -\ln(n) - \psi(\alpha) + \psi(n\alpha) \text{ and } \mathrm{Var}[\underline{R}_n] = (1/n)\psi'(\alpha) - \psi'(n\alpha) \tag{2.60}$$

We also define as c and v, functions of α and n, the solutions of the system of equations (2.61)

$$2n\alpha\, \mathrm{E}[\underline{R}_n] = cv \text{ and } (2n\alpha)^2\mathrm{Var}[\underline{R}_n] = 2c^2v \tag{2.61}$$

From (2.61) we obtain

$$c = \frac{n\alpha\mathrm{Var}[\underline{R}_n]}{\mathrm{E}[\underline{R}_n]} \text{ and } v = \frac{2\mathrm{E}^2[\underline{R}_n]}{\mathrm{Var}[\underline{R}_n]} \tag{2.62}$$

For the construction of the confidence interval see Bhaumik et al. (2009) and Engelhardt and Bain (1977). The statistic $\underline{z}$ defined by (2.63)

$$\underline{z} = 2\underline{x}/\sigma \tag{2.63}$$

has approximately a chi-square distribution with $2n\hat{\alpha}$ degrees of freedom, specifically

$$\underline{z} \sim \chi^2(2n\hat{\alpha}) \tag{2.64}$$

We define the statistic $\underline{T}_1$ by (2.65).

$$\underline{T}_1 = 2n\hat{\alpha}\underline{R}_n/c + \underline{z} \tag{2.65}$$

Then $\underline{T}_1$ is approximately distributed according to (2.66).

$$\underline{T}_1 \sim \chi^2(v+2n\hat{\alpha}) \tag{2.66}$$

Now using the $\underline{T}_1$ statistic we obtain the following $1 - \alpha$ confidence interval for the scale parameter σ.

$$[l(\underline{\mathbf{x}}_{1:n}),u(\underline{\mathbf{x}}_{1:n})] = [\frac{2n\underline{x}_1^{(n)}}{F_{\chi^2}^{-1}(1-\alpha/2|v+2n\hat{k}) - 2n\hat{\alpha}\, R_n/c}, \frac{2n\underline{x}_1^{(n)}}{F_{\chi^2}^{-1}(\alpha/2|v+2n\hat{k}) - 2n\hat{\alpha}\, R_n/c}] \tag{2.67}$$

We will designate the confidence interval obtained by (2.67) as "approximate". Another way to obtain a confidence interval is by using Bayesian analysis (See Robert 2007). According to Son and Oh (2006), if we choose a prior $\pi(\alpha,\sigma) \propto 1/\sigma$, we construct a Gibbs sampler using the following equations

$$\underline{\sigma}|\alpha,\mathbf{x}_{1:n} \sim \text{Inv-gamma}(n\alpha, \sum_{i=1}^{n} x_i) \tag{2.68}$$

Also

$$\pi(\alpha|\sigma,\mathbf{x}_{1:n}) \propto [\Gamma(\alpha)]^{-n}\sigma^{-n\alpha}\prod_{i=1}^{n} x_i^{\alpha-1} \tag{2.69}$$

A Wald-type interval is calculated, using the formula in Casella and Berger (2001 p.497)

$$[l(\mathbf{x}_{1:n}),u(\mathbf{x}_{1:n})] = [\hat{\sigma} - \Phi^{-1}(1-\alpha/2)\sqrt{-l''(\hat{\sigma})}, \hat{\sigma} + \Phi^{-1}(1-\alpha/2)\sqrt{-l''(\hat{\sigma})}] \tag{2.70}$$

where $-l''(\hat{\sigma})$ is an estimate of the Hessian at $(\hat{\alpha},\hat{\sigma})$, when optimizing the log-likelihood function.

We designate the confidence region obtained by (2.68) and (2.69) as Bayesian, the BCa interval as "bootstrap", the confidence interval obtained by Ripley's two methods as "Ripley location" and "Ripley scale" and the confidence interval obtained by our algorithm as MCCI. Figure 2.6 compares the confidence intervals obtained by all seven methods for a simulated sample with 50 elements from a gamma distribution with $\alpha = 2$ and $\sigma = 3$. For this sample $\hat{\alpha} = 1.979$ and $\hat{\sigma} = 3.007$. As we can see, the MCCI, "Ripley scale" and "bootstrap" limits are close to the Bayesian ones, but the approximate, "Wald-type and "Ripley location" limits lie far apart, which shows that they do not provide a satisfactory approximation (perhaps owing to too many assumptions involved in their derivation).

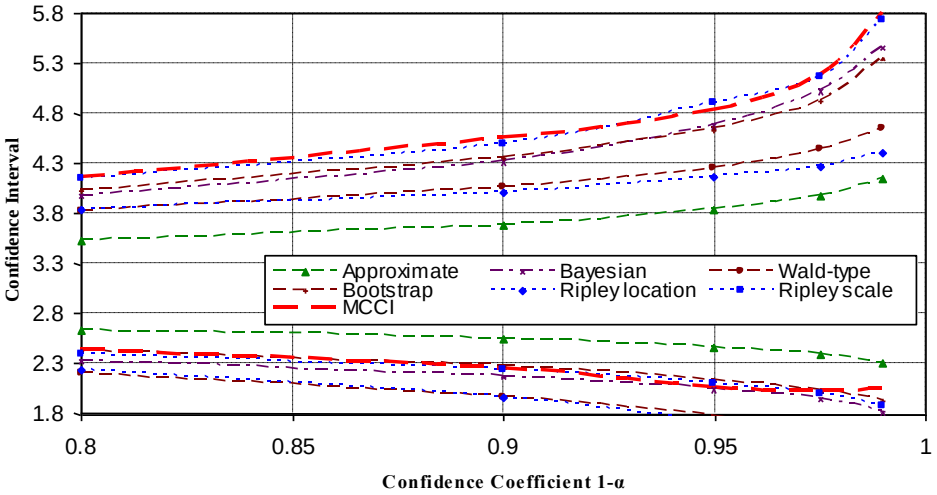


Figure 2.6. Confidence intervals for the scale parameter of a gamma distribution with $n = 50$, $\hat{\alpha} = 1.979$ and $\hat{\sigma} = 3.007$. Here the number of samples $m = 20\,000$ for MCCI, "Ripley location" and "Ripley scale" cases, $\delta\alpha = 0.3$ and $\delta\sigma = 0.3$.

2.5.5 Confidence interval for the shape parameter of the gamma distribution

To obtain a $1 - \alpha$ confidence interval for the shape parameter α, according to Bhaumik et al. (2009; see also Engelhardt and Bain 1978), we use the statistic $\underline{T}_1$ defined by (2.71).

$$\underline{T}_1 = 2n\alpha\underline{R}_n \tag{2.71}$$

Then $\underline{T}_1$ is approximately distributed according to (2.72).

$$\underline{T}_1 \sim c\chi^2(\nu) \tag{2.72}$$

Then a $1 - \alpha$ confidence interval corresponds to the following inequality

$$\frac{\mathrm{Var}[\underline{R}_n]}{\mathrm{E}[\underline{R}_n]} F^{-1}_{\chi^2}(\alpha/2|\nu) < 2R_n < \frac{\mathrm{Var}[\underline{R}_n]}{\mathrm{E}[\underline{R}_n]} F^{-1}_{\chi^2}(1 - \alpha/2|\nu) \tag{2.73}$$

where we solve for α.

A Wald-type interval is calculated, using the formula in Casella and Berger (2001 p.497)

$$[l(\boldsymbol{x}_{1:n}),u(\boldsymbol{x}_{1:n})] = [\hat{\alpha} - \Phi^{-1}(1-\alpha/2)\sqrt{-l''(\hat{\alpha})}, \hat{\alpha} + \Phi^{-1}(1-\alpha/2)\sqrt{-l''(\hat{\alpha})}] \qquad (2.74)$$

where $-l''(\hat{\alpha})$ is an estimate of the Hessian at $(\hat{\alpha}, \hat{\sigma})$, when optimizing the log-likelihood function.

We designate the confidence interval obtained by (2.73) as "approximate", the confidence region obtained by (2.68), (2.69) as Bayesian, the confidence interval obtained by (2.74) as Wald-type, the BCa confidence interval as "bootstrap", the confidence intervals obtained by the two Ripley's methods as "Ripley location" and "Ripley scale" and the confidence interval obtained by our algorithm as MCCI.

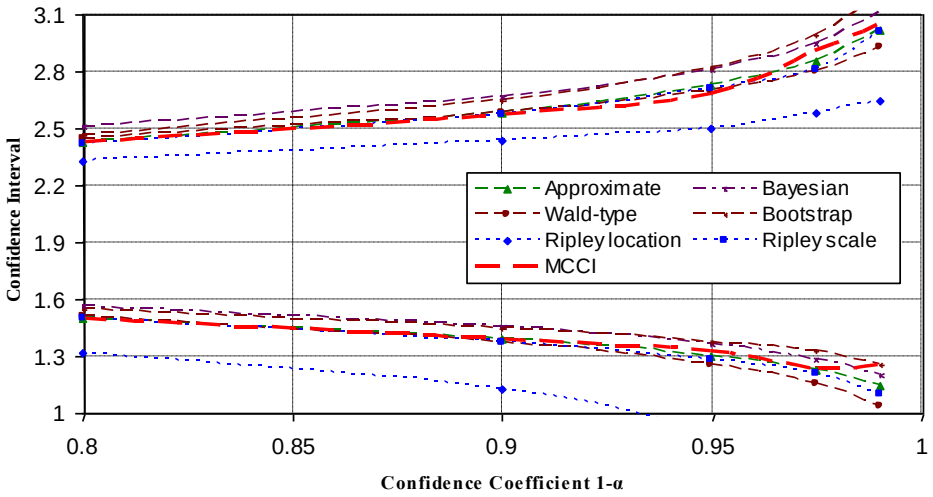


Figure 2.7. Confidence intervals for the shape parameter of a gamma distribution with $n = 50$, $\hat{\alpha} = 1.979$ and $\hat{\sigma} = 3.007$. Here the number of samples $m = 20\,000$ for MCCI, "Ripley location" and "Ripley scale" cases, $\delta\alpha = 0.3$ and $\delta\sigma = 0.3$.

Figure 2.7 compares the confidence intervals obtained by all seven methods for a simulated sample with 50 elements from a gamma distribution with $k = 2$ and $\theta = 3$. For this sample $\hat{\alpha} = 1.979$ and $\hat{\sigma} = 3.007$. As we can see, the "approximate", Wald-type, "Ripley location" and MCCI confidence intervals are close. The Bayesian confidence region is close to the "approximate" which, in our opinion, gives a good approximation of the exact confidence interval. "Ripley location" is far from the other intervals.

2.5.6 Confidence interval for the scale parameter of the Weibull distribution

The density of the Weibull distribution is

$$f_W(x|a,b) = (b/a)\,(x/a)^{b-1}\exp(-(x/a)^b),\ x > 0 \qquad (2.75)$$

where $a > 0$ is the scale parameter and $b > 0$ is the shape parameter. According to Yang et al. (2007) first we must find a modified MLE of b, according to the following equation, which is a modification of the equations discussed in Cohen (1965).

$$l(b) := \frac{n-2}{b} - (n\sum_{i=1}^{n} x_i^b \ln x_i)(\sum_{i=1}^{n} x_i^b)^{-1} + \sum_{i=1}^{n} \ln x_i = 0 \tag{2.76}$$

We denote $\hat{b}$ the modified maximum likelihood estimate given by (2.76) and $\hat{a}$ the modified maximum likelihood estimate given by the following equation.

$$\hat{a} = [(1/n)\sum_{i=1}^{n} x_i^{\hat{b}}]^{1/\hat{b}} \tag{2.77}$$

We define $S(b)$ and c_1 by (2.78).

$$S(b) := \sum_{i=1}^{n} x_i^b,\ c_1 := \sqrt{1 + 0.607927 \cdot 0.422642^2} \tag{2.78}$$

Then a $1 - \alpha$ confidence interval estimate is given by the (2.79).

$$[l(\boldsymbol{x}_{1:n}), u(\boldsymbol{x}_{1:n})] = [(\frac{2S(\hat{b})}{c_1 F_{\chi 2}^{-1}(1-\alpha/2|2n) - 2n(c_1-1)})^{1/\hat{b}}, (\frac{2S(\hat{b})}{c_1 F_{\chi 2}^{-1}(\alpha/2|2n) - 2n(c_1-1)})^{1/\hat{b}}] \tag{2.79}$$

A Wald-type interval is calculated, using the formula in Casella and Berger (2001 p.497)

$$[l(\boldsymbol{x}_{1:n}), u(\boldsymbol{x}_{1:n})] = [\hat{a} - \Phi^{-1}(1-\alpha/2)\sqrt{-l''(\hat{a})},\ \hat{a} + \Phi^{-1}(1-\alpha/2)\sqrt{-l''(\hat{a})}] \tag{2.80}$$

where $-l''(\hat{a})$ is an estimate of the Hessian at $(\hat{a}, \hat{b})$, when optimizing the log-likelihood function.

We designate the interval obtained by (2.79) as "approximate", the interval obtained by (2.80) as Wald-type, the BCa interval as "bootstrap", the confidence interval obtained by Ripley's two method as "Ripley location" and "Ripley scale" and the confidence interval obtained by our algorithm as MCCI.

Figure 2.8 compares the confidence intervals obtained by the six methods for a simulated sample with 50 elements from a Weibull distribution with $a = 2$ and $b = 3$. For this sample $\hat{a} = 2.022$ and $\hat{b} = 3.097$. As we can see, the "approximate", "Ripley scale" and MCCI confidence intervals are almost indistinguishable and the Wald-type, "bootstrap" and "Ripley location" are far from the previous intervals.

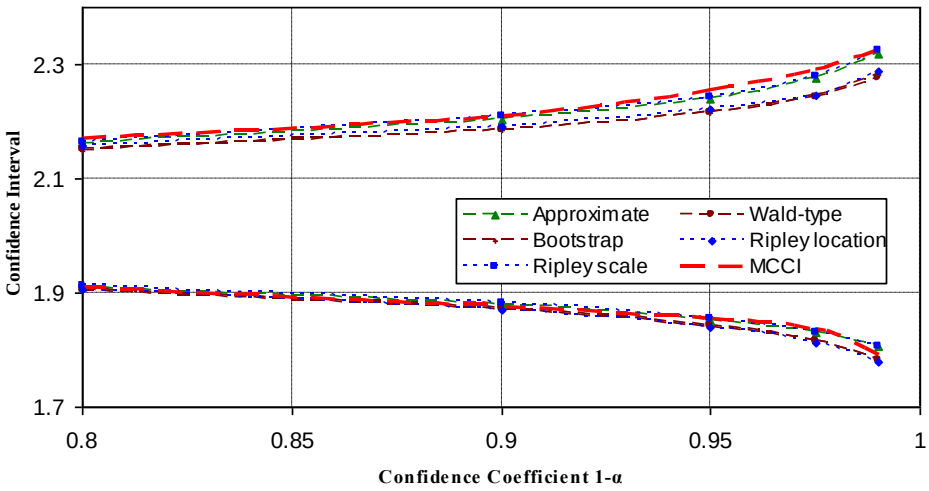


Figure 2.8. Confidence intervals for the scale parameter of a Weibull distribution with n = 50, $\hat{a}$ = 2.022 and $\hat{b}$ = 3.097. Here the number of samples m = 20 000 for MCCI, "Ripley location" and "Ripley scale" cases, $\delta a = 0.1$ and $\delta b = 0.1$.

2.5.7 Confidence interval for the pth percentile of the Weibull distribution

According to Yang et al. (2007) the pth percentile of the Weibull distribution is

$$t_p = a[-\ln(1-p)]^{1/b} \tag{2.81}$$

Then a $1 - \alpha$ approximate confidence interval estimate is given by the following equation.

$$[l(\boldsymbol{x}_{1:n}),u(\boldsymbol{x}_{1:n})] = [(-\frac{2S(\hat{b})\ln(1-p)}{c_2\, F^{-1}_{\chi^2}(1-\alpha/2|2n) - 2n(c_2-1)})^{1/\hat{b}},(-\frac{2S(\hat{b})\ln(1-p)}{c_2\, F^{-1}_{\chi^2}(\alpha/2|2n) - 2n(c_2-1)})^{1/\hat{b}}] \tag{2.82}$$

where c_2 is defined by (2.83).

$$c_2 := \sqrt{1 + 0.607927\cdot\{0.422642 - \ln[-\ln(1-p)]\}^2} \tag{2.83}$$

Figure 2.9 compares the confidence intervals obtained by the five methods, the "approximate", the "bootstrap", the "Ripley location", the "Ripley scale" and the MCCI for a simulated sample with 50 elements from a Weibull distribution with a = 2 and b = 3. For this sample $\hat{a}$ = 2.022 and $\hat{b}$ = 3.097. "Bootstrap" and "Ripley location" are close to each other but far from the other three confidence intervals.

2.5.8 Summary results

Table 2.1 shows the results of all previous methods summarized. MCCI is similar to "exact" (when "exact" can be calculated analytically, cases 1,2). MCCI is also similar to "approximate", in cases 5,6,7. In these cases "approximate" seems to be a good approximation

of an exact confidence interval. This implies that MCCI is a good approximation of an exact confidence interval.

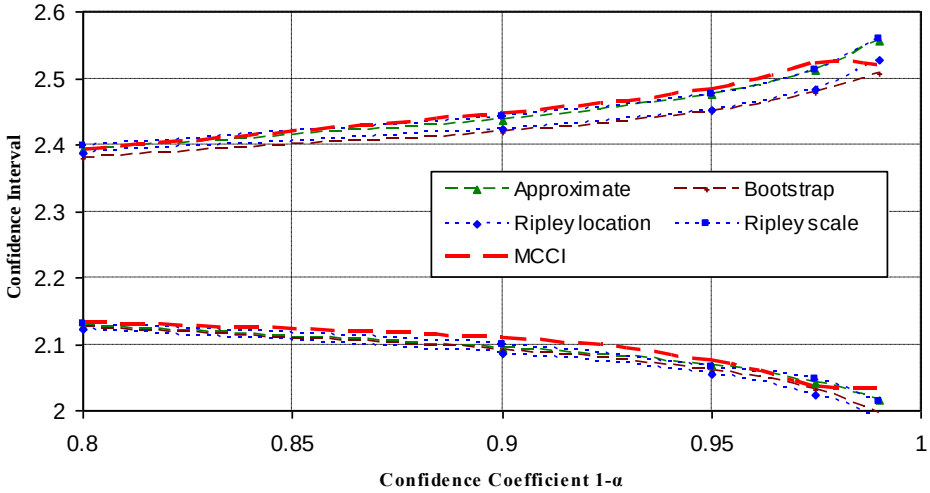


Figure 2.9. Confidence intervals for the 75th percentile of a Weibull distribution with n = 50, $\hat{a}$ = 2.022 and $\hat{b}$ = 3.097. Here the number of samples m = 20 000 for MCCI, "Ripley location" and "Ripley scale" cases, $\delta a = 0.1$ and $\delta b = 0.1$.

On the other hand in case 3 MCCI is almost identical to "Bayesian" which we think is a good property. In case 4, MCCI is closer to "Bayesian" than the "approximate". We believe that the "approximate" is not a good approximation of an exact confidence interval, because it involves a lot of assumptions and transformations. MCCI was also better in our opinion than Wald-type and bootstrap intervals in all cases. We should also keep in mind that confidence intervals and Bayesian "confidence regions" are not directly comparable (see also the Chapter dedicated to matching priors in Robert 2007 p.137).

Table 2.1. Summary results of the case studies examined. Smaller numbers mean that the corresponding result is better. Equal numbers mean that there is a similarity between the different results. For example, in the case of the percentile of the normal distribution, MCCI, "Ripley scale" and "Bayesian" methods (marked as 1) gave similar results, whereas Wald-type, "bootstrap" and "Ripley location" methods (marked as 2, 3 and 3 correspondingly) gave results worse than the former methods.

				Methods							
Case	Figure No	Distribution	Parameter	Exact	Bayesian	Approximate	Ripley location	Ripley scale	Wald-type	Bootstrap	MCCI
1	2.2	Exponential	Scale	1			4	1	3	2	1
2	2.3	Normal	Location	1			2	2	2	3	1
3	2.4	Normal	Percentile		1		3	1	2	3	1
4	2.6	Gamma	Scale		1	3	2	1	2	1	1
5	2.7	Gamma	Shape		2	1	3	1	1	2	1
6	2.8	Weibull	Scale			1	2	1	2	2	1
7	2.9	Weibull	Percentile			1	2	1		2	1

As an additional means of intercomparison, coverage probabilities using Monte Carlo methods were calculated for all methods except for the Bayesian confidence regions and the

algorithm behaved relatively well in all cases (Table 2.2). MCCI was better when estimating the confidence intervals for the normal and the gamma distribution parameters, and had the best mean rank for all the examined cases.

An application of the algorithm, using historical river flow data is given in Section 2.8.

Table 2.2. Monte Carlo coverage probabilities and rank of each method when calculating 0.975 confidence intervals after 10 000 iterations (rank 1 is assigned to the method of best performance).

							Coverage probabilities (with ranks in parentheses) for all methods				
Case	Distribution	Parameter	Sample size	Parameter value	Parameter value	Approximate	Ripley location	Ripley scale	Wald-type	Bootstrap	MCCI
1	Exponential	Scale	10	$\sigma = 2$			0.889 (5)	0.977 (2)	0.975 (1)	0.916 (4)	0.966 (3)
2	Normal	Location	10	$\mu = 0$	$\sigma = 1$		0.946 (3)	0.946 (3)	0.947 (2)	0.931 (5)	0.968 (1)
3	Normal	Percentile	10	$\mu = 0$	$\sigma = 1$		0.919 (4)	0.929 (2)	0.929 (2)	0.867 (5)	0.973 (1)
4	Gamma	Scale	50	$\alpha = 2$	$\sigma = 3$	0.753	0.923 (5)	0.976 (1)	0.940 (4)	0.957 (3)	0.974 (1)
5	Gamma	Shape	50	$\alpha = 2$	$\sigma = 3$	0.976	0.948 (5)	0.972 (2)	0.978 (2)	0.956 (4)	0.974 (1)
6	Weibull	Scale	50	$a = 2$	$b = 3$	0.971	0.969 (3)	0.970 (2)	0.966 (4)	0.965 (5)	0.973 (1)
7	Weibull	Percentile	50	$a = 2$	$b = 3$	0.971	0.968 (3)	0.970 (1)		0.961 (4)	0.969 (2)
	mean rank						4.000	1.857	2.500	4.286	1.429

2.6 Sensitivity to the choice of the increment and the simulation sample size

In this Section we test the sensitivity of the algorithm on the choice of the increments $\delta\mu$ and $\delta\sigma$ and the simulated sample size in the case of the location parameter and the percentile of the normal distribution.

Figure 2.10 tests the sensitivity of the algorithm to the choice of the increments $\delta\mu$ and $\delta\sigma$ in the cases of the location and the percentile parameters of the normal distribution, for $n = 10$ (upper panel) and $n = 50$ (lower panel), where for the calculation of the confidence interval the unbiased estimators of μ and σ^2 were used. As we see, the algorithm gives good approximations, regardless of the choice of $\delta\mu$ and $\delta\sigma$. For small n, a slight problem appears if $\delta\mu$ is too small (< 0.5). Figure 2.11 describes the convergence of the algorithm for the same cases. The speed of convergence is low since ~50 000 iterations are needed for its stabilization, although reasonable results are obtained even for ~10 000 iterations.

2.7 Some theoretical results

First we show that the confidence interval for the parameter μ of a normal distribution $N(\mu,\sigma^2)$ is asymptotically equivalent to a Wald-type interval. For the normal distribution we define

$$\boldsymbol{\theta} := (\mu, \sigma) \tag{2.84}$$

$$\boldsymbol{T}(\underline{\boldsymbol{x}}) := (T_1(\underline{\boldsymbol{x}}), T_2(\underline{\boldsymbol{x}})) \tag{2.85}$$

where $T_1(\underline{\boldsymbol{x}}) = \underline{\mu}$, and $T_2(\underline{\boldsymbol{x}}) = \underline{\sigma}$ are the MLE of μ and σ respectively.

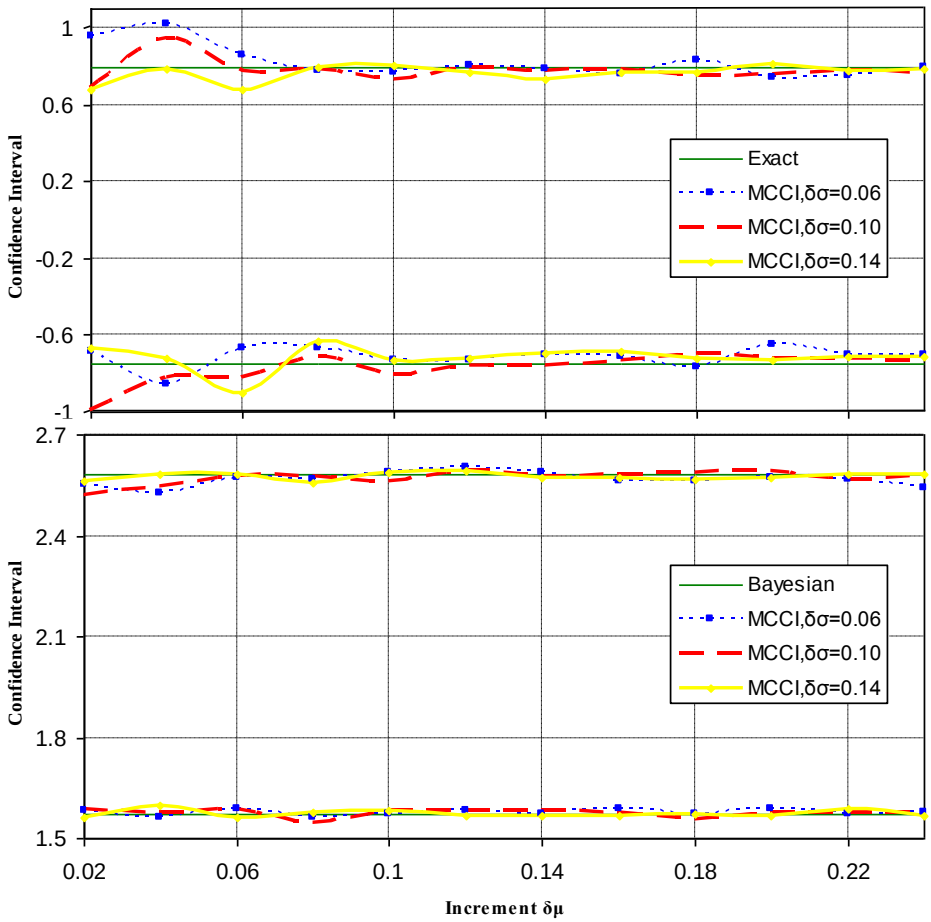


Figure 2.10. 0.95 confidence intervals for a normal distribution estimated for different $\delta\mu$ and $\delta\sigma$ (the parameter increments denoted in text as $\delta\theta_i$): (upper) confidence interval for the location parameter μ from a sample with n = 10, $\hat{\mu}$ = 0.026 and $\hat{\sigma}$ = 1.023 and number of samples drawn m = 100 000; (lower) confidence interval for the quantity $\mu + 2\sigma$ from a sample with n = 50, $\hat{\mu}$ = −0.027 and $\hat{\sigma}$ = 0.998 and number of samples drawn m = 50 000.

Then, following the notation of the preceding Sections we have

$$\beta := h(\mu,\sigma) = \mu \tag{2.86}$$

$$h(\boldsymbol{T}) = T_1 \tag{2.87}$$

and

$$P(b(\underline{\mathbf{x}}) < \lambda(\boldsymbol{\theta})) = \alpha/2,\ P(b(\underline{\mathbf{x}}) > \upsilon(\boldsymbol{\theta})) = \alpha/2 \tag{2.88}$$

which imply that

$$\lambda = \mu + \Phi^{-1}(\alpha/2)\sigma/\sqrt{n} \text{ and } \upsilon = \mu + \Phi^{-1}(1-\alpha/2)\sigma/\sqrt{n} \tag{2.89}$$

Now from (2.36) we obtain

$$\frac{d\boldsymbol{\gamma}}{d\boldsymbol{\theta}} = \begin{bmatrix} \frac{d\lambda}{d\boldsymbol{\theta}} \\ \frac{d\beta}{d\boldsymbol{\theta}} \\ \frac{d\upsilon}{d\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} 1 & \Phi^{-1}(\alpha/2)/\sqrt{n} \\ 1 & 0 \\ 1 & \Phi^{-1}(1-\alpha/2)/\sqrt{n} \end{bmatrix} \tag{2.90}$$

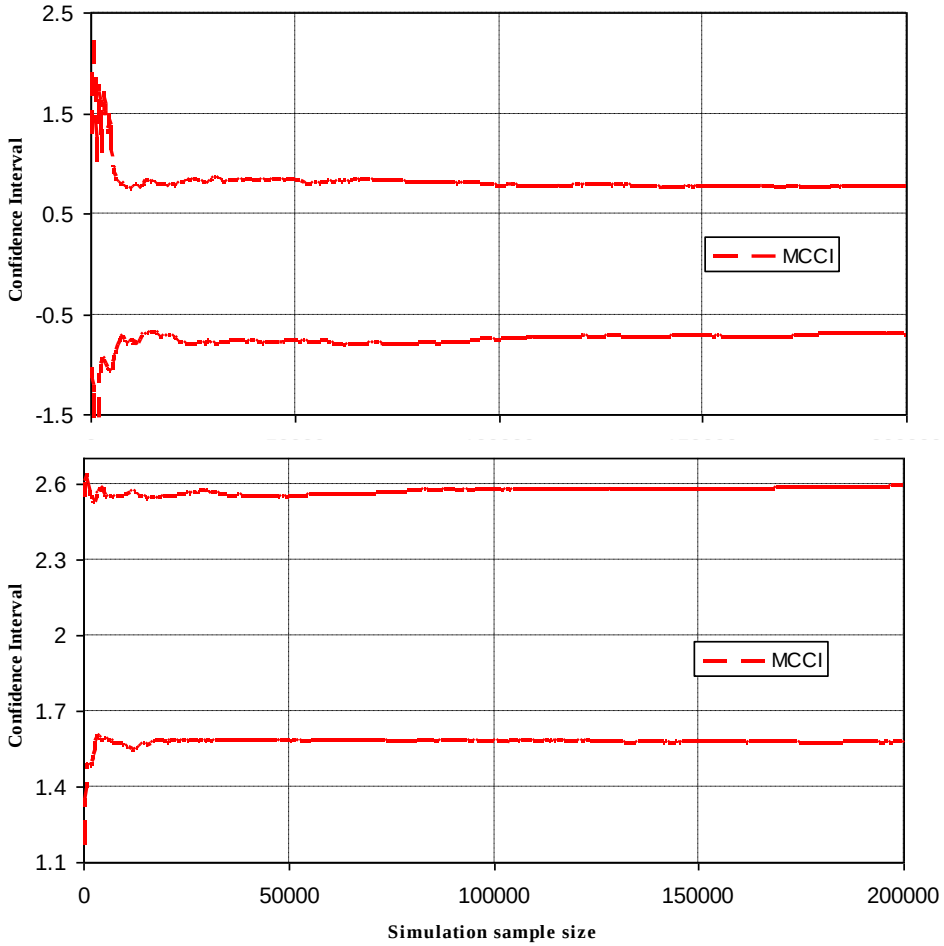


Figure 2.11. 0.95 confidence intervals for a normal distribution estimated for varying simulation sample size: (upper) confidence interval for the location parameter μ from a sample with $n = 10$, $\hat{\mu} = 0.026$ and $\hat{\sigma} = 1.023$; (lower) confidence interval for the quantity $\mu + 2\sigma$ from a sample with $n = 50$, $\hat{\mu} = -0.027$ and $\hat{\sigma} = 0.998$.

It is also easy to prove that asymptotically

$$\begin{bmatrix} \underline{\mu} - \mu \\ \underline{\sigma} - \sigma \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{\sigma^2}{n} \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \right) \tag{2.91}$$

thus

$$\underline{\mu} \sim \mathrm{N}(\mu, \sigma^2/n) \text{ and } \underline{\sigma} \sim \mathrm{N}(\sigma, \sigma^2/2n) \tag{2.92}$$

We also have that

$$\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(\alpha/2) \tag{2.93}$$

From (2.15), (2.16) we derive

$$l = \underline{\mu} - \frac{\underline{\sigma}\Phi^{-1}(1-\alpha/2)}{\sqrt{n}\,\frac{d\upsilon}{d\mu}} \text{ and } u = \underline{\mu} + \frac{\underline{\sigma}\Phi^{-1}(1-\alpha/2)}{\sqrt{n}\,\frac{d\lambda}{d\mu}} \tag{2.94}$$

From (2.43) we have that

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} = \frac{\mathrm{d}\upsilon}{\mathrm{d}\mu} = 1 - \frac{(\Phi^{-1}(1-\alpha/2))^2}{4n} \tag{2.95}$$

A $1-\alpha$ confidence interval for μ is $(\underline{\mu} - F_{\mathrm{T}(n-1)}(1-\alpha/2)\frac{\underline{\sigma}}{\sqrt{n}}, \underline{\mu} + F_{\mathrm{T}(n-1)}(1-\alpha/2)\frac{\underline{\sigma}}{\sqrt{n}})$ (e.g. Papoulis and Pillai 2002 p.309). Now we have that

$$\lim_{n\to\infty} \frac{\Phi^{-1}(1-\alpha/2)/\left(\frac{\mathrm{d}\upsilon}{\mathrm{d}\mu}\right)}{F_{\mathrm{T}(n-1)}(1-\alpha/2)} = 1 \tag{2.96}$$

which proves that the confidence interval obtained by (2.17) is asymptotically exact.

We will also show that the confidence interval obtained by our method is asymptotically equivalent to a Wald-type interval for two-parameter regular distributions. According to Casella and Berger (2001 p.472)

$$\sqrt{n}(\underline{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathrm{d}} \mathrm{N}(\mathbf{0}, \boldsymbol{I}^{-1}) \tag{2.97}$$

where $\underline{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$, and $\boldsymbol{I}$ is the Fisher Information Matrix with elements

$$\boldsymbol{I}_{jk} = \mathrm{E}\left[-\frac{\partial^2 \ln f(x|\boldsymbol{\theta})}{\partial\theta_j\theta_k}\right] \tag{2.98}$$

This means that

$$\sqrt{n}(\underline{\theta}_1 - \theta_1) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \boldsymbol{I}_{11}^{-1}) \text{ and } \sqrt{n}(\underline{\theta}_2 - \theta_2) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \boldsymbol{I}_{22}^{-1}) \tag{2.99}$$

We conclude that

$$\sqrt{n}(\underline{\beta} - \beta) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma_\beta^2) \tag{2.100}$$

where σ_β^2 depends only on θ_1 and θ_2. Suppose that we seek a $1-\alpha$ confidence interval for β. Then it is easy to show that asymptotically

$$\lambda(\beta) = \beta - \Phi^{-1}(1-\alpha/2)\sigma_\beta/\sqrt{n},\ \upsilon(\beta) = \beta + \Phi^{-1}(1-\alpha/2)\sigma_\beta/\sqrt{n} \tag{2.101}$$

Now we have

$$\mathrm{Var}[\underline{\theta}_1] = \boldsymbol{I}_{11}^{-1}/n,\ \mathrm{Var}[\underline{\theta}_2] = \boldsymbol{I}_{22}^{-1}/n \tag{2.102}$$

and

$$\frac{\mathrm{d}\boldsymbol{\gamma}}{\mathrm{d}\boldsymbol{\theta}} = \begin{bmatrix} \frac{\mathrm{d}\lambda}{\mathrm{d}\boldsymbol{\theta}} \\ \frac{\mathrm{d}\beta}{\mathrm{d}\boldsymbol{\theta}} \\ \frac{\mathrm{d}\upsilon}{\mathrm{d}\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \frac{\partial\beta}{\partial\theta_1} - \Phi^{-1}(1-\alpha/2)\frac{\partial\sigma_\beta}{\partial\theta_1}/\sqrt{n} & \frac{\partial\beta}{\partial\theta_2} - \Phi^{-1}(1-\alpha/2)\frac{\partial\sigma_\beta}{\partial\theta_2}/\sqrt{n} \\ \frac{\partial\beta}{\partial\theta_1} & \frac{\partial\beta}{\partial\theta_2} \\ \frac{\partial\beta}{\partial\theta_1} + \Phi^{-1}(1-\alpha/2)\frac{\partial\sigma_\beta}{\partial\theta_1}/\sqrt{n} & \frac{\partial\beta}{\partial\theta_2} + \Phi^{-1}(1-\alpha/2)\frac{\partial\sigma_\beta}{\partial\theta_2}/\sqrt{n} \end{bmatrix} \tag{2.103}$$

$$\frac{q_{31}+q_{32}}{q_{21}+q_{22}} = \frac{[\Phi^{-1}(1-\alpha/2)]^2[(\frac{\partial\sigma_\beta}{\partial\theta_1})^2\boldsymbol{I}_{11}^{-1} + (\frac{\partial\sigma_\beta}{\partial\theta_2})^2\boldsymbol{I}_{22}^{-1}] - \Phi^{-1}(1-\alpha/2)\sqrt{n}[\frac{\partial\sigma_\beta}{\partial\theta_1}\frac{\partial\beta}{\partial\theta_1}\boldsymbol{I}_{11}^{-1} + \frac{\partial\sigma_\beta}{\partial\theta_2}\frac{\partial\beta}{\partial\theta_2}\boldsymbol{I}_{22}^{-1}]}{\sqrt{n}[\Phi^{-1}(1-\alpha/2)(\frac{\partial\sigma_\beta}{\partial\theta_1}\frac{\partial\beta}{\partial\theta_1} + \frac{\partial\sigma_\beta}{\partial\theta_2}\frac{\partial\beta}{\partial\theta_2}) - 2\sqrt{n}((\frac{\partial\beta}{\partial\theta_1})^2\boldsymbol{I}_{11}^{-1} + (\frac{\partial\beta}{\partial\theta_2})^2\boldsymbol{I}_{22}^{-1})]} -$$

$$- \frac{2n((\frac{\partial\beta}{\partial\theta_1})^2\boldsymbol{I}_{11}^{-1} + (\frac{\partial\beta}{\partial\theta_2})^2\boldsymbol{I}_{22}^{-1})}{\sqrt{n}[\Phi^{-1}(1-\alpha/2)(\frac{\partial\sigma_\beta}{\partial\theta_1}\frac{\partial\beta}{\partial\theta_1} + \frac{\partial\sigma_\beta}{\partial\theta_2}\frac{\partial\beta}{\partial\theta_2}) - 2\sqrt{n}((\frac{\partial\beta}{\partial\theta_1})^2\boldsymbol{I}_{11}^{-1} + (\frac{\partial\beta}{\partial\theta_2})^2\boldsymbol{I}_{22}^{-1})]} \tag{2.104}$$

It is obvious that

$$\lim_{n\to\infty}\frac{\mathrm{d}\upsilon}{\mathrm{d}\beta} = \lim_{n\to\infty}\frac{q_{31}+q_{32}}{q_{21}+q_{22}} = 1 \tag{2.105}$$

In a similar way we can find that

$$\lim_{n\to\infty}\frac{\mathrm{d}\lambda}{\mathrm{d}\beta} = \lim_{n\to\infty}\frac{q_{12}+q_{13}}{q_{22}+q_{23}} = 1 \tag{2.106}$$

Now substituting to (2.15), (2.16) we obtain

$$l = \underline{\beta} - \Phi^{-1}(1-\alpha/2)\sigma_\beta/\sqrt{n},\ u = \underline{\beta} + \Phi^{-1}(1-\alpha/2)\sigma_\beta/\sqrt{n} \tag{2.107}$$

which is an asymptotically equivalent to a Wald-type interval according to Casella and Berger (2001 p.497).

Repeating the same procedure for three-parameter distributions, we obtain the same results.

2.8 Application of the algorithm to a historical river flows dataset

In this Section we apply the algorithm on a historical river flow data set using the hydrological statistical software Hydrognomon (Itia research group 2009-2012), suitable for the processing and the analysis of hydrological time series, which has already incorporated

the proposed method. The case study is performed on an important basin in Greece, which is currently part of the water supply system of Athens and has a history, as regards hydraulic infrastructure and management, that goes back to at least 3500 years ago. Modelling attempts with good performance have already been carried out on the hydrosystem (Rozos et al., 2004). A long-term dataset of the catchment runoff, extending from 1906 to 2008, is available. The example presented in Figure 2.12 is for the January monthly flow record at the Boeoticos Kephisos river outlet at the Karditsa station measured in hm^3. The gamma distribution is often used to model monthly river flows. Confidence limits of quantiles of distributions are of interest to hydrologists. Here we derived confidence intervals for the scale and the shape parameters of the gamma distribution. Comparison of the results of the different methods used show that the MCCI and "Ripley scale" limits are close to the Bayesian ones. In addition, Figure 2.13 gives confidence limits of the distribution percentiles using the same dataset, this time constructed using Hydrognomon (Itia research group 2009-2012).

2.9 Conclusions

By modifying two Monte Carlo methods used by Ripley (1987), associated with the computation of a confidence interval for a parameter of a probability distribution, we derive a new equation and a general algorithm which gives a single solution for a confidence interval, which combines the advantages of these two methods without requiring discrimination for the type of parameter. We show that this algorithm is exact for a single parameter of distribution of either location or scale family. It is also asymptotically equivalent to a Wald-type interval for parameters of regular continuous distributions.

After appropriate modification of the algorithm we make it appropriate for calculating confidence intervals for a parameter of multi-parameter distributions. We show that this algorithm is asymptotically equivalent to a Wald-type interval for regular distributions.

We tested the algorithm in seven cases, namely the construction of a confidence interval for the scale parameter of the exponential distribution, the location parameter and the *p*th percentile of the normal distribution, the scale and shape parameter of a gamma distribution, and the scale parameter and the *p*th percentile of the Weibull distribution. We found that in general this algorithm works well and results in correct coverage probabilities.

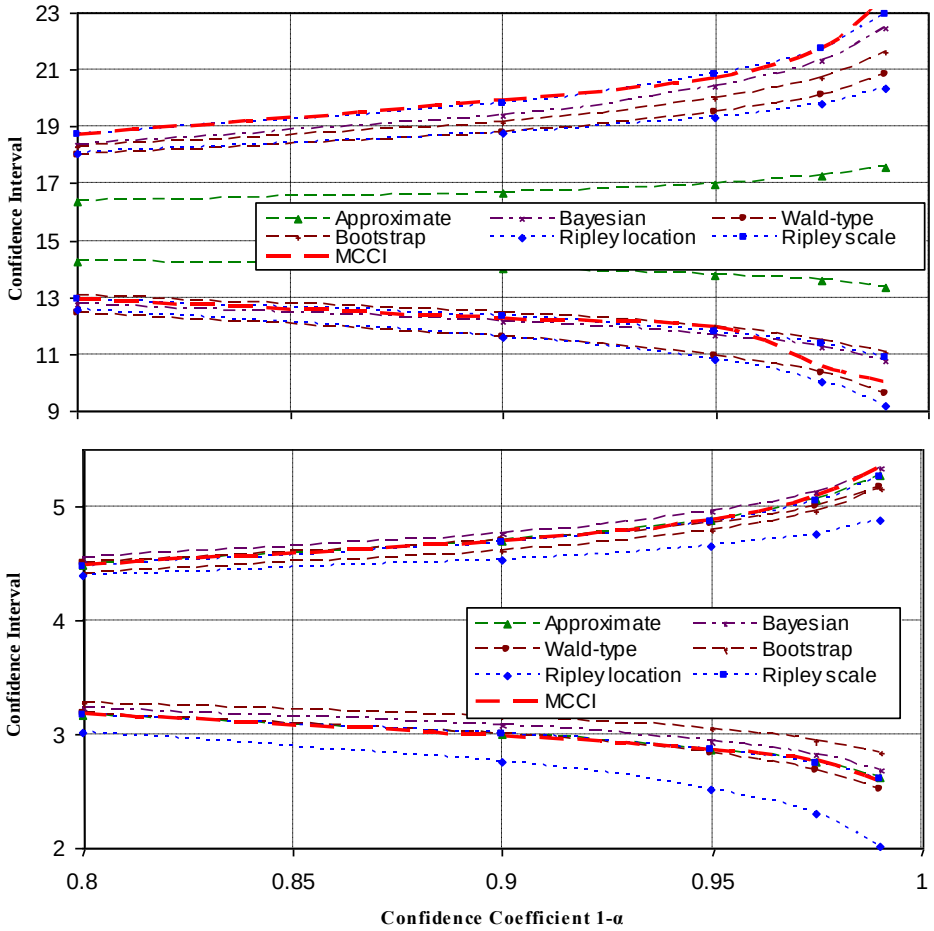


Figure 2.12. Confidence intervals for the scale (upper) and shape (lower) parameter of a gamma distribution, used to model the Boeoticos Kephisos river January monthly flows with $n = 102$, $\hat{\alpha} = 3.842$ and $\hat{\sigma} = 15.218$. Here the number of samples $m = 120\,000$ for MCCI and $m = 60\,000$ for the "Ripley location" and "Ripley scale" cases, $\delta\alpha = 0.3$ and $\delta\sigma = 0.3$.

We propose the use of the algorithm for an approximation of a confidence interval of any parameter for any continuous distribution because it is easily applicable in every case and gives better approximations than other known algorithms as shown in specific cases above. An additional advantage compared to Ripley's two methods is that it is not needed to select one of the methods. Our algorithm worked equally well or better from the best of Ripley's methods in all the examined cases. Thanks to its generality, the algorithm has been implemented in the hydrometeorological software package Hydrognomon (Itia research group 2009-2012), which fits various distributions in data records and calculates point and interval estimates for parameters and distribution quantiles, which are then used for hydrological design.

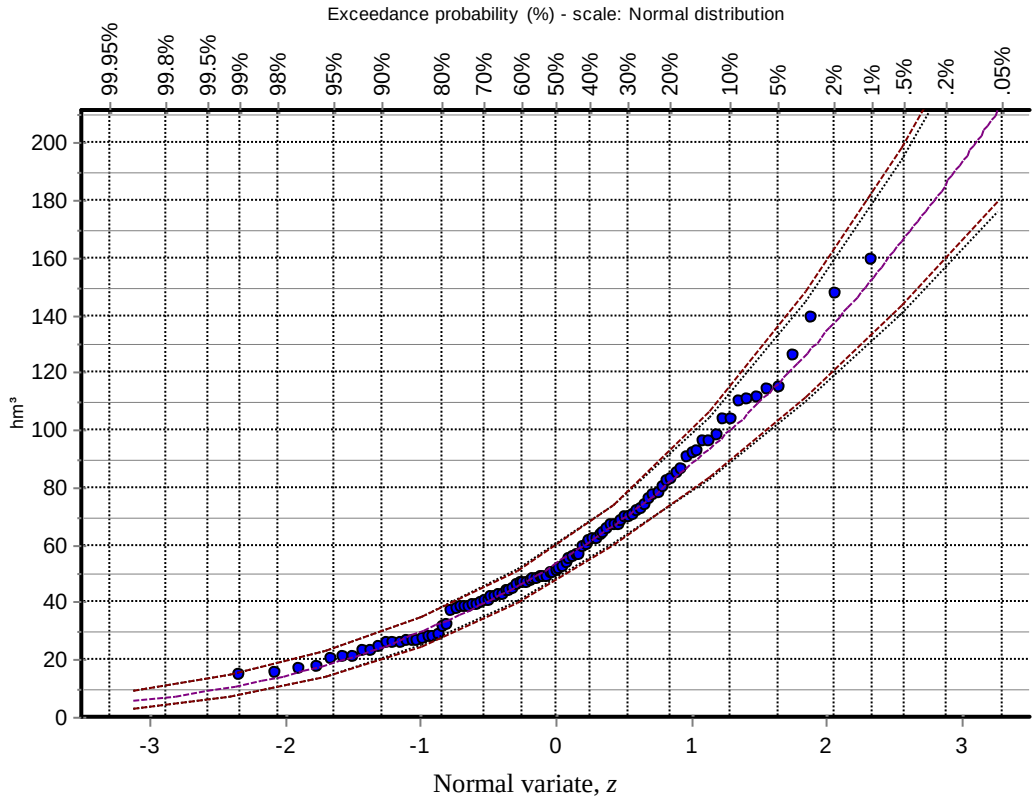


Figure 2.13. A graph (normal probability plot) produced by the Hydrognomon software referring to the monthly flow of Boeoticos Kephisos river for the month of January (1993-2006). The sample (dots plotted using Weibul plotting positions) was modelled by a gamma distribution (central line) with $\hat{\alpha}$ = 3.842 and $\hat{\sigma}$ = 15.218. Dotted lines represent 95% prediction intervals for these parameter values (denoted as λ and υ in the text) and dashed lines represent 95% confidence intervals (MCCI denoted as l and u in the text) for the distribution percentiles.

The confidence intervals obtained by the algorithm are approximate and the algorithm was not developed with the intention to replace the exact confidence intervals, when their calculation is possible. Further research is needed to evaluate the influence of the choice of the numerical parameters (increments $\delta\boldsymbol{\theta}_i$ and the simulation sample size) to the results of the algorithm. A disadvantage of the algorithm is that a lot of repetitions are needed to converge.

3. Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process

Of critical importance[1] in analyzing hydrological and geophysical time series is the estimation of the strength of the HK behaviour. The parameter *H* of the HKp arises naturally from the study of self-similar processes and expresses the strength of the HK behaviour. A number of estimators of *H* have been proposed. These are usually validated by an appeal to some aspect of self-similarity, or by an asymptotic analysis of the distributional properties of the estimator as the length of the time series converges to infinity and only estimate the self-similarity parameter.

Here we show that the estimation of *H* affects the estimation of the standard deviation, a fact that was not given appropriate attention in the literature. We propose the Least Squares based on Variance estimator, and we investigate numerically its performance, which we compare to the Least Squares based on Standard Deviation estimator, as well as the maximum likelihood estimator after appropriate streamlining of the latter. These three estimators rely on the structure of the HKp and estimate simultaneously its Hurst parameter and standard deviation. In addition, we test the performance of the three methods for a range of sample sizes and *H* values, through a simulation study and we compare it with other estimators of the literature.

3.1 Introduction

Rea et al. (2013) present an extensive literature review dealing with the properties of these estimators. They also examine the properties of twelve estimators, i.e. the nine estimators (aggregated variance, differencing the variance, absolute values of the aggregated series, Higuchi's method, residuals of regression, R/S method, periodogram method, modified periodogram method, Whittle estimator) discussed in Taqqu et al. (1995) plus the wavelet, GPH and Haslett-Raftery estimator. Weron (2002) discusses the properties of residuals of regression, R/S method and periodogram method. Grau-Carles (2005) also analyzes the behaviour of the residuals of regression, the R/S method and the GPH.

Additionally, new estimators are proposed, for example Guerrero and Smith (2005) presented a maximum likelihood based estimator, while Coeurjolly (2008) presented estimators based on convex combinations of sample quantiles of discrete variations of a

sample path over a discrete grid of the interval [0, 1]. Some authors propose improvements of existing estimators. For example, Mielniczuk and Wojdyllo (2007) improve the R/S method. Other authors like Esposti et al. (2008) propose techniques which use more than one methods simultaneously to estimate the *H* parameter.

Because the finite sample properties of these estimators can be quite different from their asymptotic properties, some authors have undertaken empirical comparisons of estimators of *H*. The nine classical estimators were discussed in some detail by Taqqu et al. (1995) who carried out an empirical study of these estimators for a single series length of 10 000 data points, 5 values of *H*, and 50 replications. All twelve estimators above were discussed in more detail by Rea et al. (2013) who carried out an empirical study of these estimators for series lengths between 100 and 10 000 data points in steps of 100, *H* values between 0.55 and 0.90 in steps of 0.05 and 1000 replications. Rea et al. (2013) also presented an extensive literature review about the same kind of empirical studies.

These studies did not include two methods. The maximum likelihood (ML) method discussed by McLeod and Hippel (1978) and McLeod et al. (2007), probably due to computational problems (Beran 1994 p.109), and the method by Koutsoyiannis (2003), hereinafter referred to as the LSSD (Least Squares based on Standard Deviation) method, which was also articulated recently by Ehsanzadeh and Adamowski (2010). The ML method estimates the Hurst parameter based on the whole structure of the process, i.e. its joint distribution function. The LSSD method relies on the self-similarity property of the process. One common characteristic of the ML and LSSD methods is that they estimate simultaneously the Hurst parameter *H* and the standard deviation σ of the process. This is of great importance, because both parameters are essential for the construction of the model and, as we will show below (see also Koutsoyiannis 2003) their estimators generally are not independent of each other. In addition, the classical statistical estimator of σ encompasses strong bias if applied to a series with HK behaviour (Koutsoyiannis 2003; Koutsoyiannis and Montanari 2007). It is thus striking that some of the existing methods do not remedy or even pose this problem at all, and estimate *H* independently of σ and vice versa, e.g. assuming that σ can be estimated using its classical statistical estimator, which does not depend on *H*.

The focus of this Chapter is the simultaneous estimation of the parameters *H* and σ of the HKp. We use the ML and LSSD methods that have the capacity for simultaneous estimation,

[1] Based on: Tyralis and Koutsoyiannis (2011)

after appropriate streamlining of the former in a more practical form, and we propose a third method which is an improvement of the LSSD method (referred to as LSV method —Least Squares based on Variance) retaining the simultaneous parameter estimation attitude. We apply the three methods to evaluate their performance in a Monte Carlo simulation framework and we compare the results with those of the estimators presented in Taqqu et al. (1995) with the exception of the Whittle estimator, which we replaced by the local Whittle estimator presented in Robinson (1995).

3.2 Definitions

We assume that $\{\underline{x}_t\}$, $t = 1, 2, \ldots$ is an HKp. We also define the aggregated stochastic process for every time scale:

$$\underline{z}_t^{(\kappa)} := \sum_{l = (t-1)\,\kappa + 1}^{t\,\kappa} \underline{x}_l = \kappa\, \underline{x}_t^{(\kappa)} \tag{3.1}$$

For this process the following relationships hold:

$$\mathrm{E}[\underline{z}_t^{(\kappa)}] = \kappa\,\mu,\ \gamma_0^{(\kappa)} = \mathrm{Var}[\underline{z}_t^{(\kappa)}] = \kappa^{2\cdot H}\,\gamma_0,\ \sigma^{(\kappa)} = (\gamma_0^{(\kappa)})^{1/2} \tag{3.2}$$

The autocorrelation function of either of $\underline{x}_t^{(\kappa)}$ and $\underline{z}_t^{(\kappa)}$, for any aggregated timescale κ, is independent of κ, and given by

$$\rho_k^{(\kappa)} = \rho_k = |k+1|^{2H} / 2 + |k-1|^{2H} / 2 - |k|^{2H},\ k = 0, 1, \ldots \tag{3.3}$$

3.3 Methods

3.3.1 Maximum likelihood estimator

In this Section the method of maximum likelihood is employed for the estimation of the parameters of HKp, namely H, σ, μ. For a given record $\boldsymbol{x}_{1:n}$ the likelihood of $\boldsymbol{\theta} := (\mu, \sigma, H)$ takes the general form (McLeod and Hippel 1978):

$$l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n}) = \frac{1}{(2\pi)^{n/2}}\ |\sigma^2\, \boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \exp[-1/(2\sigma^2)\ (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\ (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)] \tag{3.4}$$

where

$$\boldsymbol{e}_n = (1,1,\ldots,1)^{\mathrm{T}} \tag{3.5}$$

is a column vector with n elements, $\boldsymbol{R}_{[1:n]\ [1:n]}$ is the autocorrelation matrix, i.e., a n-by-n matrix with elements $r_{ij} = \rho_{|i-j|}$, and $|\cdot|$ denotes the determinant of a matrix.

Then a maximum likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{H})$, as shown in Section 3.4, consists of the following relationships:

$$\hat{\mu} = \frac{\boldsymbol{x}_{1:n}^{\mathrm{T}}\, \hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\, \boldsymbol{e}_n}{\boldsymbol{e}_n^{\mathrm{T}}\, \hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\, \boldsymbol{e}_n}, \tag{3.6}$$

$$\hat{\sigma} = \sqrt{\frac{(\boldsymbol{x}_{1:n} - \hat{\mu}\, \boldsymbol{e}_n)^{\mathrm{T}}\, \hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\, (\boldsymbol{x}_{1:n} - \hat{\mu}\, \boldsymbol{e}_n)}{n}} \tag{3.7}$$

and $\hat{H}$ can be obtained from the maximization of the single-variable function $g_1(H)$ defined as:

$$g_1(H) := -\frac{n}{2} \ln\left[(\boldsymbol{x}_{1:n} - \frac{\boldsymbol{x}_{1:n}^{\mathrm{T}}\, \hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\, \boldsymbol{e}_n}{\boldsymbol{e}_n^{\mathrm{T}}\, \hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\, \boldsymbol{e}_n} \boldsymbol{e}_n)^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} (\boldsymbol{x}_{1:n} - \frac{\boldsymbol{x}_{1:n}^{\mathrm{T}}\, \hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\, \boldsymbol{e}_n}{\boldsymbol{e}_n^{\mathrm{T}}\, \hat{\boldsymbol{R}}_{[1:n]\,[1:n]}^{-1}\, \boldsymbol{e}_n} \boldsymbol{e}_n)\right] - \frac{1}{2} \ln(|\boldsymbol{R}_{[1:n]\,[1:n]}|) \tag{3.8}$$

3.3.2 LSSD method

This method was proposed by Koutsoyiannis (2003). In his paper after a systematic Monte Carlo study he found an estimator $\underline{\tilde{s}}_n$ of σ, approximately unbiased for known H and for normal distribution of $\underline{x}_i$, where

$$\underline{\tilde{s}}_n := \sqrt{\frac{n - 1/2}{n - n^{2H-1}}}\, \underline{s}_n = \sqrt{\frac{n - 1/2}{(n-1)\,(n - n^{2H-1})}} \sqrt{\sum_{i=1}^{n} (\underline{x}_i - \underline{x}_1^{(n)})^2} \tag{3.9}$$

$$\underline{s}_n := \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (\underline{x}_i - \underline{x}_1^{(n)})^2} \tag{3.10}$$

This algorithm is based on classical sample estimates $s^{(\kappa)}$ of standard deviations $\sigma^{(\kappa)}$ for timescales κ ranging from 1 to a maximum value $\kappa' = [n/10]$. This maximum value was chosen so that $s^{(\kappa)}$ can be estimated from at least 10 data values.

Combining (3.2) and (3.9), assuming $\mathrm{E}[\underline{\tilde{s}}] = \sigma$ and using the self-similarity property of the process one obtains

$$\mathrm{E}[\underline{s}_n^{(\kappa)}] \approx \mathrm{c}_\kappa(H)\, \kappa^H\, \sigma \tag{3.11}$$

with

$$\mathrm{c}_\kappa(H) := \sqrt{\frac{n/\kappa - (n/\kappa)^{2H-1}}{n/\kappa - 1/2}} \tag{3.12}$$

Then the algorithm minimizes a fitting error $\mathrm{er}^2(\sigma, H)$:

$$\mathrm{er}^2(\sigma, H) := \sum_{\kappa=1}^{\kappa'} \frac{[\ln \mathrm{E}[\underline{s}_n^{(\kappa)}] - \ln s_n^{(\kappa)}]^2}{\kappa^p} + \frac{H^{q+1}}{q+1} = \sum_{\kappa=1}^{\kappa'} \frac{[\ln\sigma + H \cdot \ln\kappa + \ln c_\kappa(H) - \ln s_n^{(\kappa)}]^2}{\kappa^p} + \frac{H^{q+1}}{q+1} \quad (3.13)$$

where a weight equal to $1/\kappa^p$ is assigned to the partial error of each scale κ. For $p = 0$ the weights are equal whereas for $p = 1, 2, \ldots$, decreasing weights are assigned to increasing scales; this is reasonable because at larger scales the sample size is smaller and thus the uncertainty larger. Using Monte Carlo experiments it was found that, although differences in estimates caused by different values of p in the range 0 to 2 are not so important, $p = 2$ results in slightly more efficient estimates (i.e., with smaller variation) and thus is preferable. A penalty factor $H^{q+1}/(q+1)$ has been included in er^2 in (3.13) for a high q, say 50. The effect of this factor is that it excludes the value $\hat{H} = 1$ and forces $\hat{H}$ to slightly smaller values when it is close to 1. As a consequence this factor helps get rid of an infinite $\hat{\sigma}$ also forcing to smaller values for $\hat{H}$ close to 1 (see Section 3.5).

An analytical procedure to locate the minimum is not possible. Therefore, minimization of $\mathrm{er}^2(\sigma, H)$ is done numerically and several numerical procedures can be devised for this purpose. A detailed iterative procedure is given in Koutsoyiannis (2003).

3.3.3 LSV method

In Section 3.3.2 an approximately unbiased estimator $\tilde{\underline{s}}_n$ of σ was found after a systematic Monte Carlo simulation. However, if σ^2 is used instead of σ, we have the advantage that there exists a theoretically consistent expression, which determines $\mathrm{E}[\underline{s}^2]$ as a function of σ and H. This is the basis to form a modified version of the LSSD method, the LSV method. From the general relationship (Beran 1994 p.9)

$$\mathrm{E}[\underline{s}^2] = (1 - \frac{\delta_n(\rho)}{n-1})\,\sigma^2, \text{ where } \delta_n(\rho) := (1/n) \sum_{i \neq j} \rho(i, j) = 2 \sum_{k=1}^{n-1} (1 - \frac{k}{n})\,\rho_k \quad (3.14)$$

we easily obtain that for an HKp:

$$\mathrm{E}[\underline{s}_n^2] = \frac{n - n^{2H-1}}{n-1}\,\sigma^2 \quad (3.15)$$

Due to the self-similarity property of the process the following relationship holds:

$$\mathrm{E}[\underline{s}_n^{2(\kappa)}] = \frac{(n/\kappa) - (n/\kappa)^{2H-1}}{(n/\kappa) - 1}\,\gamma_0^{(\kappa)} = \frac{(n/\kappa) - (n/\kappa)^{2H-1}}{(n/\kappa) - 1}\,\kappa^{2H}\,\sigma^2 = c_\kappa(H)\,\kappa^{2H}\,\sigma^2 \quad (3.16)$$

where

$$c_\kappa(H) := \frac{(n/\kappa) - (n/\kappa)^{2H-1}}{(n/\kappa) - 1} \text{ and } s_n^{2(\kappa)} = \frac{1}{n/\kappa - 1} \sum_{i=1}^{n/\kappa} (z_i^{(\kappa)} - k\, x_1^{(n)})^2. \tag{3.17}$$

Thus, the following error function should be minimized in order to obtain an estimation of H and σ:

$$er^2(\sigma, H) := \sum_{\kappa=1}^{\kappa'} \frac{[E[s_n^{2(\kappa)}] - s_n^{2(\kappa)}]^2}{\kappa^p} = \sum_{\kappa=1}^{\kappa'} \frac{[c_\kappa(H)\, \kappa^{2H}\, \sigma^2 - s_n^{2(\kappa)}]^2}{\kappa^p},\ \kappa' = [n/10] \tag{3.18}$$

Taking partial derivatives, i.e.,

$$\frac{\partial er^2(\sigma, H)}{\partial \sigma^2} = 2\, [\sigma^2\, \alpha_{11}(H) - \alpha_{12}(H)] \tag{3.19}$$

where:

$$\alpha_{11}(H) := \sum_{\kappa=1}^{\kappa'} \frac{c_\kappa^2(H)\, \kappa^{4H}}{\kappa^p},\ \alpha_{12}(H) := \sum_{\kappa=1}^{\kappa'} \frac{c_\kappa(H)\, \kappa^{2H}\, s_n^{2(\kappa)}}{\kappa^p} \tag{3.20}$$

and equating to zero we obtain an estimate of σ:

$$\hat{\sigma} = \sqrt{\alpha_{12}(\hat{H})/\alpha_{11}(\hat{H})} \tag{3.21}$$

An estimate of H can be obtained by minimizing the single-variable function:

$$g_2(H) := \sum_{\kappa=1}^{\kappa'} \frac{s_n^{4(\kappa)}}{\kappa^p} - \frac{\alpha_{12}^2(H)}{\alpha_{11}(H)},\ 0 < H < 1 \tag{3.22}$$

We prove in Section 3.5 that $er^2(\sigma, H)$ attains its minimum for $H \le 1$. However, when $\hat{H} = 1$, then from equations (3.21) and (3.30) we obtain that $\hat{\sigma} = \infty$. Accordingly, to avoid such behaviour (values of $\hat{\sigma}$ tending to infinity), a penalty factor $H^{q+1}/(q+1)$ for a high q is added again, as in method LSSD, to the error function.

So the function to be minimized becomes:

$$er^2(\sigma, H) := \sum_{\kappa=1}^{\kappa'} \frac{[c_\kappa(H)\, \kappa^{2H}\, \sigma^2 - s_n^{2(\kappa)}]^2}{\kappa^p} + \frac{H^{q+1}}{q+1} \tag{3.23}$$

An estimate of H can be obtained by the minimization of the single-variable function:

$$g_2(H) := \sum_{\kappa=1}^{\kappa'} \frac{s_n^{4(\kappa)}}{\kappa^p} - \frac{\alpha_{12}^2(H)}{\alpha_{11}(H)} + \frac{H^{q+1}}{q+1},\ 0 < H < 1 \tag{3.24}$$

and σ is again estimated from (3.21).

3.4 Proof of equations (3.6) and (3.8)

From equation (3.4) we obtain:

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

$$+ \frac{\boldsymbol{e}_n^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n\,\boldsymbol{x}_{1:n}^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{x}_{1:n} - (\boldsymbol{x}_{1:n}^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n)^2}{\boldsymbol{e}_n^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n})] \tag{3.25}$$

Since $\boldsymbol{e}_n^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n > 0$ ($\boldsymbol{R}_{[1:n]\,[1:n]}$ is positive definite matrix) the maximum of $l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})$ is achieved when

$$\hat{\mu} = \frac{\boldsymbol{x}_{1:n}^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n}{\boldsymbol{e}_n^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n} \tag{3.26}$$

For that value of μ, taking the logarithm of the posterior density we obtain:

$$\ln[l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})] = -(n/2)\ln(2\pi) - n\ln\sigma - (1/2)\ln|\boldsymbol{R}_{[1:n]\,[1:n]}| - \frac{1}{2\,\sigma^2}(\boldsymbol{x}_{1:n} - \hat{\mu}\boldsymbol{e}_n)^{\mathrm{T}}\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}(\boldsymbol{x}_{1:n} - \hat{\mu}\boldsymbol{e}_n) \tag{3.27}$$

$$\frac{\partial \ln[l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})]}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\,(\boldsymbol{x}_{1:n} - \hat{\mu}\boldsymbol{e}_n)^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,(\boldsymbol{x}_{1:n} - \hat{\mu}\boldsymbol{e}_n) \tag{3.28}$$

Thus, the logarithm of the maximum posterior density is maximized when $\partial\ln[l(\boldsymbol{\theta}|\boldsymbol{x}_n)]/\partial\sigma = 0$. The solution of this equation proves equation (3.6) and gives the ML estimator of σ.

Substituting the values of μ and σ from equation (3.6), we obtain:

$$\ln[l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})] = \frac{n}{2}\ln(\frac{n}{2\pi}) - \frac{n}{2} - \frac{n}{2}\ln[(\boldsymbol{x}_{1:n} - \frac{\boldsymbol{x}_{1:n}^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n}{\boldsymbol{e}_n^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n}\boldsymbol{e}_n)^{\mathrm{T}}\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}(\boldsymbol{x}_{1:n} - \frac{\boldsymbol{x}_{1:n}^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n}{\boldsymbol{e}_n^{\mathrm{T}}\,\boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\,\boldsymbol{e}_n}\boldsymbol{e}_n)] -$$

$$-\frac{1}{2}\cdot\ln[\det(\boldsymbol{R}_{[1:n]\,[1:n]})] = \frac{n}{2}\ln(\frac{n}{2\pi}) - \frac{n}{2} + g_1(H) \tag{3.29}$$

which is a function of H through the matrix $\boldsymbol{R}_{[1:n]\,[1:n]}$. So we maximize the above single-variable function, or equivalently the function $g_1(H)$, and find $\hat{H}$.

We may observe that it is not necessary to form the entire matrix $\boldsymbol{R}_{[1:n]\,[1:n]}$ and invert it to compute $g_1(H)$ (It suffices to form a column $(\rho_0\ \ldots\ \rho_{n-1})^{\mathrm{T}}$). Since $\boldsymbol{R}_{[1:n]\,[1:n]}$ is a positive

definite Toeplitz matrix we can use the Levinson-Trench-Zohar algorithm (Musicus, 1988). This algorithm can solve the problem of calculating $\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n$ and $\ln|\boldsymbol{R}_{[1:n]\,[1:n]}|$ using only $O(n^2)$ operations and $O(n)$ storage. In contrast, standard methods such as Gaussian elimination or Choleski decomposition generally require $O(n^3)$ operations and $O(n^2)$ storage. This is of critical importance when the time series size is large and computer memory capacity restricts its ability to solve the problem.

3.5 Proof of boundedness of the LSV estimate of *H* in (0, 1]

In order to examine the behaviour of $\hat{\sigma}$ and $g_2(H)$ from equations (3.21) and (3.22) we calculate the following limits:

$$\lim_{H\to 1}\frac{\alpha_{12}^2(H)}{\alpha_{11}(H)}=\left[\sum_{\kappa=1}^{\kappa'}\frac{\ln(n/\kappa)\,\kappa^2\,s_n^{2(\kappa)}}{\kappa^p}\right]^2\Big/\sum_{\kappa=1}^{\kappa'}\frac{\ln(n/\kappa)\,\kappa^2}{\kappa^p}>0 \text{ and } \lim_{H\to 1}\frac{\alpha_{12}(H)}{\alpha_{11}(H)}=\infty \qquad (3.30)$$

Therefore, there is a possibility that $g_2(H)$ could have a minimum for $H = 1$ and $\sigma = \infty$, when σ tends to infinity from this path: $\sigma = \sqrt{\alpha_{12}(H)/\alpha_{11}(H)}$.

Then

$$\lim_{H\to 1} g_2(H)=\sum_{\kappa=1}^{\kappa'}\frac{s_n^{4(\kappa)}}{\kappa^p}-\Big(\sum_{\kappa=1}^{\kappa'}\frac{\ln(n/\kappa)\,\kappa^2\,s_n^{2(\kappa)}}{\kappa^p}\Big)^2\Big/\Big(\sum_{\kappa=1}^{\kappa'}\frac{\ln(n/\kappa)\,\kappa^2}{\kappa^p}\Big) \qquad (3.31)$$

Now we prove $er^2(\sigma, H)$ attains its minimum for $H \le 1$. The proof is given bellow:

Suppose that $H_2 > 1$ and $\sigma_2 > 0$ (It's easy to prove that an estimated $\hat{\sigma} > 0$ always). Now for any $H_1 \in (0, 1)$ we can always find a $\sigma_1 > 0$, such that $c_\kappa(H_1)\,\kappa^{2H_1}\,\sigma_1^2 - s_n^{2(\kappa)} < 0$ for every κ. For these values of H_1 and σ_1: $|c_\kappa(H_1)\,\kappa^{2\cdot H_1}\,\sigma_1^2 - s_n^{2(\kappa)}| < |c_\kappa(H_2)\,\kappa^{2\cdot H_2}\,\sigma_2^2 - s_n^{2(\kappa)}|$ for every κ. This proves that $er^2(\sigma_1, H_1) < er^2(\sigma_2, H_2)$. Thus, $er^2(\sigma, H)$ attains its minimum for $H \le 1$.

3.6 Calculation of Fisher Information Matrix's elements

We can easily calculate the $\boldsymbol{I}_{12}(\boldsymbol{\theta})$, $\boldsymbol{I}_{13}(\boldsymbol{\theta})$ and $\boldsymbol{I}_{23}(\boldsymbol{\theta})$ elements of the Fisher Information Matrix (Robert 2007 p.129):

$$\frac{\partial\ln[l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})]}{\partial\mu}=-\frac{1}{\sigma^2}\left(\boldsymbol{e}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n\,\mu-\boldsymbol{x}_{1:n}^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n\right) \qquad (3.32)$$

$$\frac{\partial\ln[l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})]}{\partial\sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\left(\boldsymbol{x}_{1:n}-\mu\,\boldsymbol{e}_n\right)^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\left(\boldsymbol{x}_{1:n}-\mu\,\boldsymbol{e}_n\right) \qquad (3.33)$$

$$\frac{\partial^2 \ln[l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})]}{\partial\mu\,\partial\sigma} = \frac{2}{\sigma^3}\,(\boldsymbol{e}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n\,\mu - \boldsymbol{x}^{\mathrm{T}}_{1:n}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n) \tag{3.34}$$

$$\frac{\partial^2 \ln[l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})]}{\partial\mu\,\partial H} = \frac{1}{\sigma^2}\,(\mu\,\boldsymbol{e}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\frac{\partial \boldsymbol{R}_{[1:n]\,[1:n]}}{\partial H}\,\cdot$$

$$\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n - \boldsymbol{x}_n^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\frac{\partial \boldsymbol{R}_{[1:n]\,[1:n]}}{\partial H}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{e}_n) \tag{3.35}$$

$$\frac{\partial^2 \ln[l(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})]}{\partial\sigma\,\partial H} = -\frac{1}{\sigma^3}\,(\boldsymbol{x}_{1:n} - \mu\,\boldsymbol{e}_n)^{\mathrm{T}}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\frac{\partial \boldsymbol{R}_{[1:n]\,[1:n]}}{\partial H}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,(\boldsymbol{x}_{1:n} - \mu\,\boldsymbol{e}_n) \tag{3.36}$$

The expectations of the above expressions are easily calculated and give the corresponding elements of the Fisher Information Matrix $\boldsymbol{I}(\boldsymbol{\theta})$.

3.7 Results

The three H and σ estimators, namely ML, LSSD and LSV are implemented in the computational software Matlab. We evaluated each estimator's performance in estimating H and σ for simulated HKp. HKp series were generated using Stoev (2008) function. This function generates "exact" paths of HKp by using circulant embedding. We ran 200 replications of simulated HKp series with eight different lengths and five different H values. The lengths were 64, 128, 256, 512, 1 024, 2 048, 4 096 and 8 192 data points. The H values were 0.60, 0.70, 0.80, 0.90 and 0.95. Without loss of generality, in all cases the true (population) value of σ was assumed 1.00.

For each series, H and σ were estimated by each of these three estimators. For each H value and series length we estimated from the simulated data the median, 75% and 95% confidence intervals and the square root of the mean square error (Taqqu et al. 1995)

$$\mathrm{RMSE} := \sqrt{\frac{1}{200}\sum_{k=1}^{200}(H_k - H)^2} \tag{3.37}$$

The H or σ estimates were sorted into ascending order and the median (50th percentile) was obtained after replacement of the 100th and 101st values by their arithmetic average. Similar calculations were done for the 75% and 95% parametric bootstrap confidence intervals, based on symmetric upper and lower sample quantile values.

Figures 3.1-3.5 depict some of the results in graphical form. (To present the results in tabular form would require a very large amount of space). In Figures 3.1-3.3 the vertical axis ranges between −0.3 and 0.2 for ΔH (the estimated H minus the true H) to facilitate

comparisons among the estimators' standard deviation of their estimates. Figure 3.4 shows the RMSE as a function of the series length. Again all vertical axes have the same range to facilitate comparisons. Figure 3.5 presents RMSE as a function of series length. Figures 3.1-3.5 also depict corresponding results for the σ estimators.

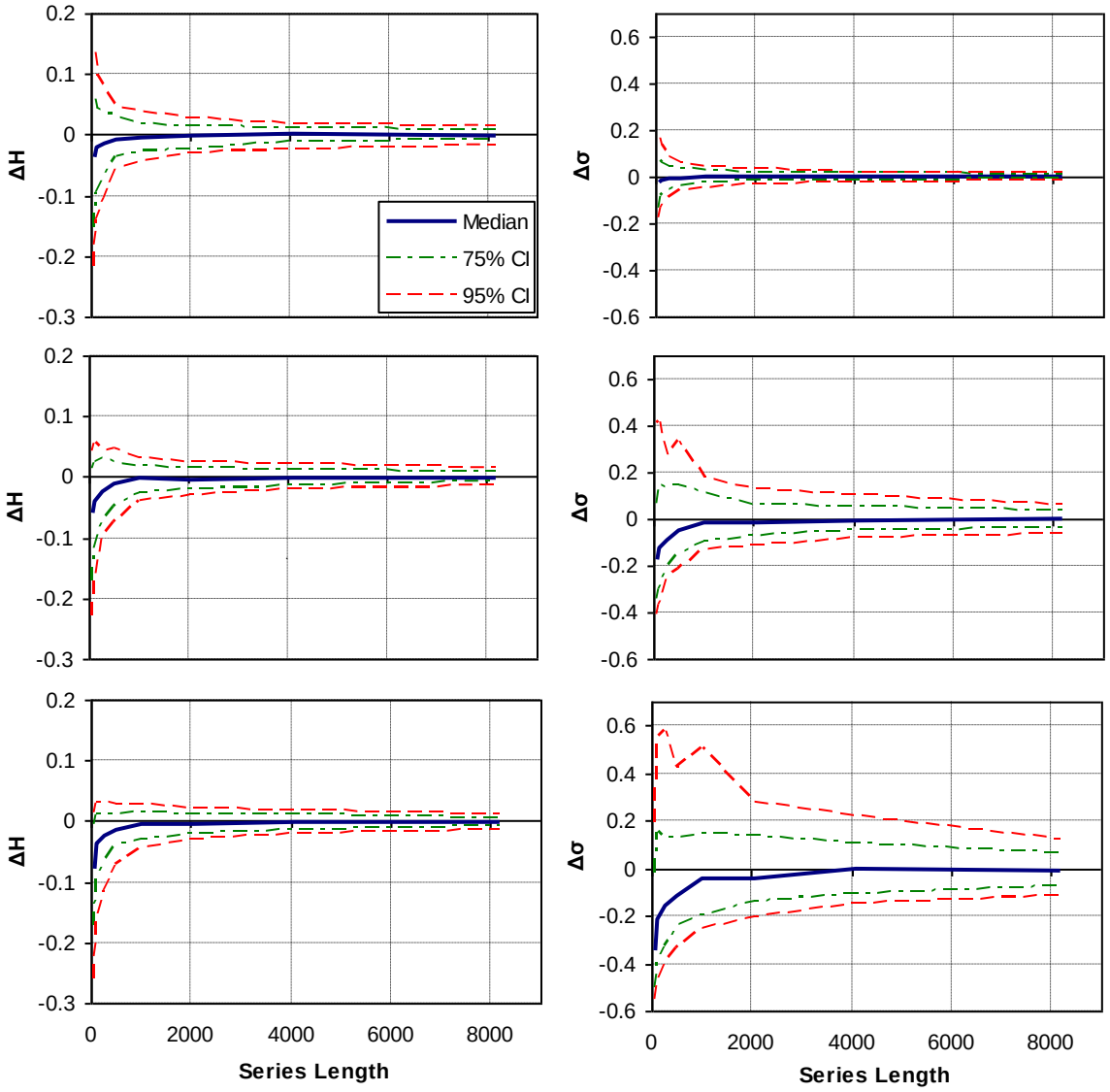


Figure 3.1. Monte Carlo confidence intervals for the H and σ estimates with true H = 0.60, H = 0.90 and H = 0.95 (upper to lower panels), where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, for the ML estimator.

The results for the ML method are shown in Figure 3.1. The ML method is unbiased for H at all series lengths when true H = 0.6, but becomes biased and underestimated H when H increases, for low length of time series. This method is unbiased for σ at all series lengths when true H = 0.6 but becomes biased and underestimates σ when H increases. But even for values of H over 0.9, the method becomes unbiased when the time series length increases.

The results for the LSSD method are presented in Figure 3.2. The LSSD method was

unbiased for H and σ at all series lengths when true $H \leq 0.9$, but became biased and underestimated H and σ when true $H = 0.95$. We observed the same results for the LSV method (Figure 3.3), but this method was slightly worse compared with the previous method. The 75% confidence intervals all contain the true values, except when true $H = 0.95$ and the LSSD or LSV method is used to estimate H or σ.

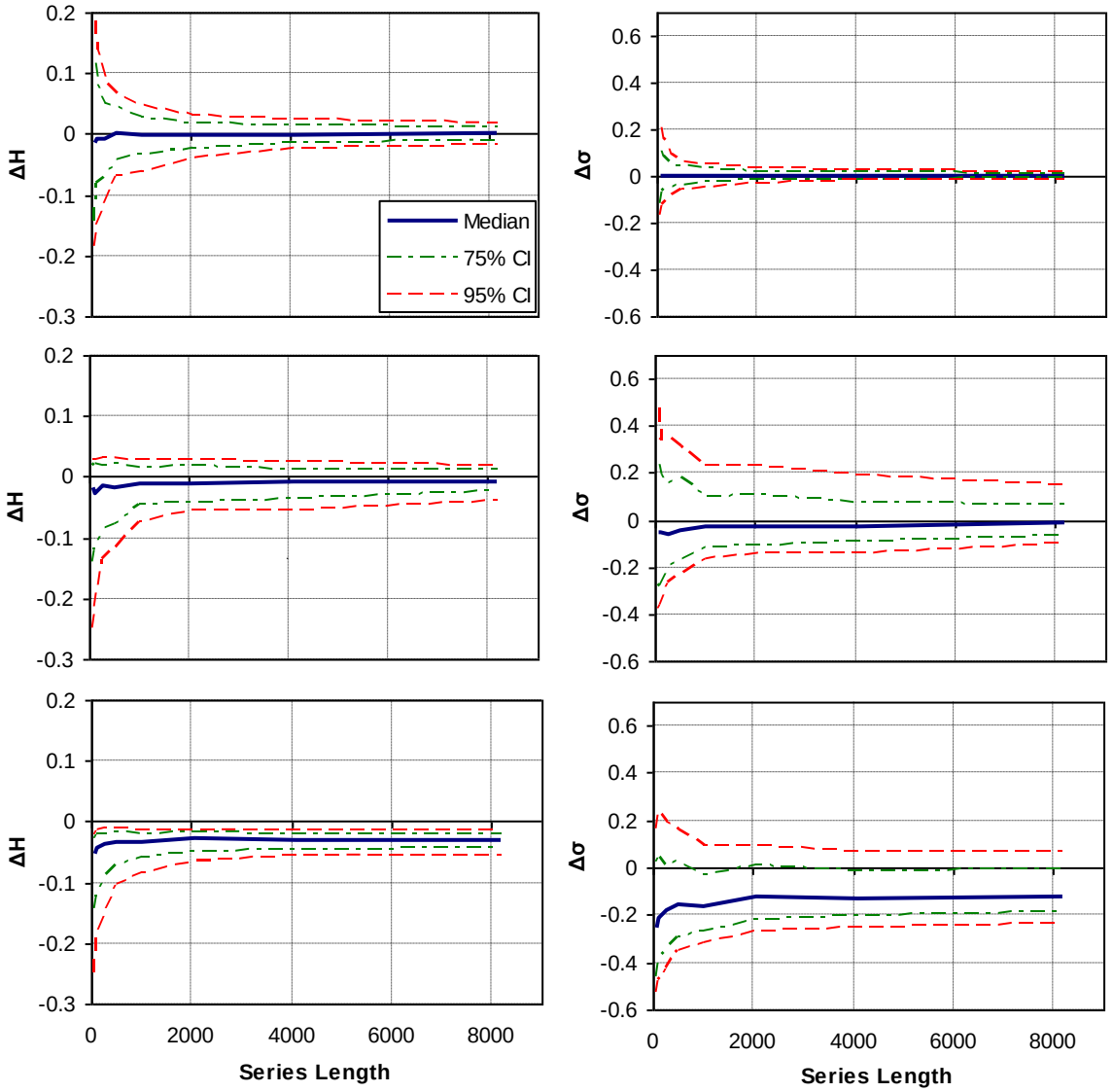


Figure 3.2. Monte Carlo confidence intervals for the H and σ estimates with true $H = 0.60$, $H = 0.90$ and $H = 0.95$ (upper to lower panels), where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, for the LSSD estimator.

Figure 3.4 compares the RMSE of all three methods. We observe that when estimating H the ML method is best, followed by the LSV method, for all values of H. The same holds when estimating σ, except that the LSSD method behaves better than the LSV method.

Figure 3.5 presents the variation of RMSE when H increases. We observe that when estimating H the RMSE increases when H increases for the LSSD and LSV methods but it

remains stable for the maximum likelihood method. However, when estimating σ, the RMSE increases for increasing H in all methods.

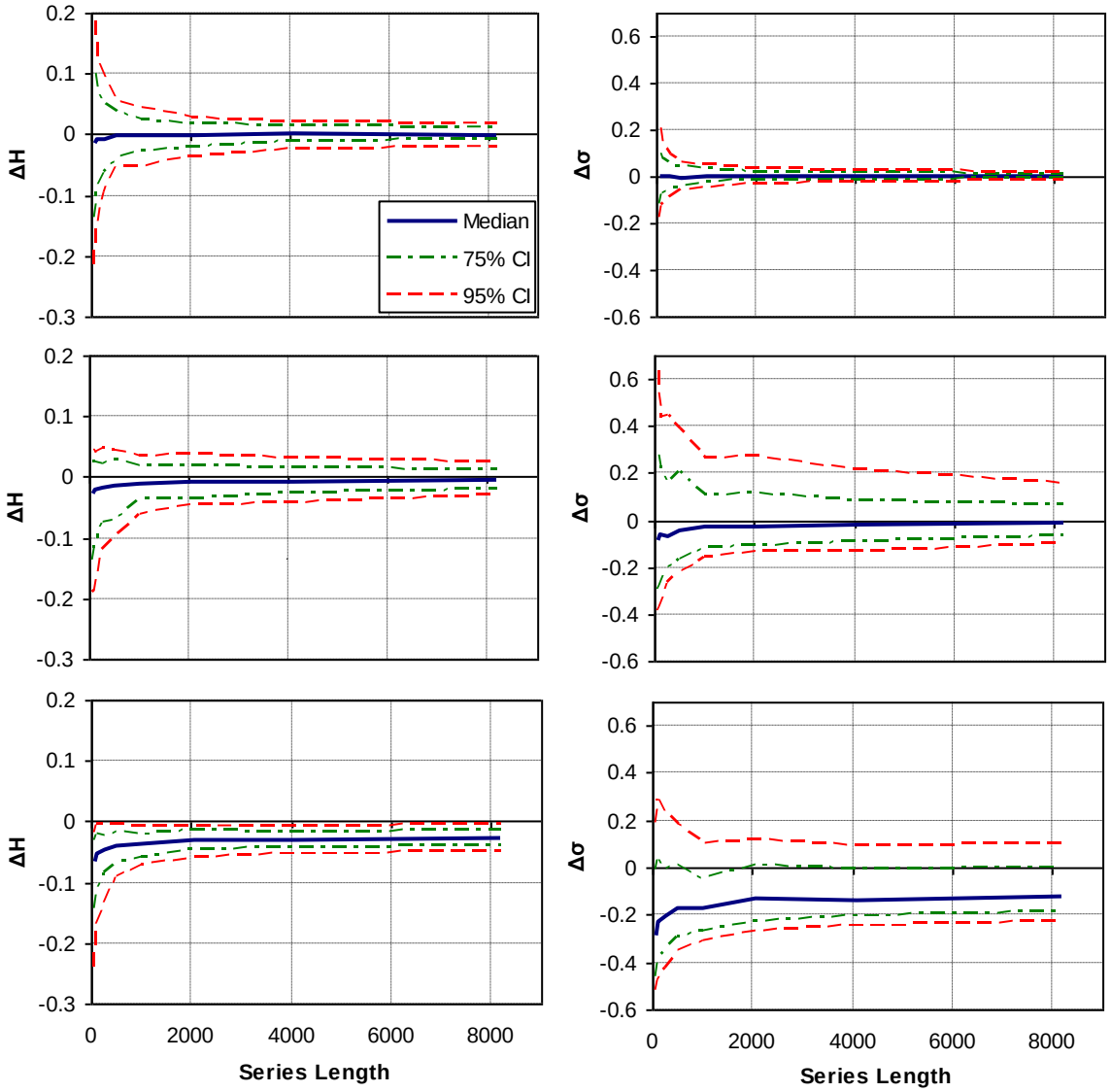


Figure 3.3. Monte Carlo confidence intervals for the H and σ estimates with true H = 0.60, H = 0.90 and H = 0.95 (upper to lower panels), where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, for the LSV estimator.

Figure 3.6 presents the correlation between $\hat{H}$ and $\hat{\mu}$, for nominal H = 0.8. $\hat{H}$ does not seem to affect $\hat{\mu}$, in terms of bias and this holds for every time series length.

Figure 3.7 presents the correlation between $\hat{H}$ and $\hat{\sigma}$, for nominal H = 0.6 and 0.8. It seems that an increase of nominal H results in an increase of the correlation between $\hat{H}$ and $\hat{\sigma}$. We can see that a high $\hat{H}$ results in a high $\hat{\sigma}$, and a low $\hat{H}$ results in a low $\hat{\sigma}$.

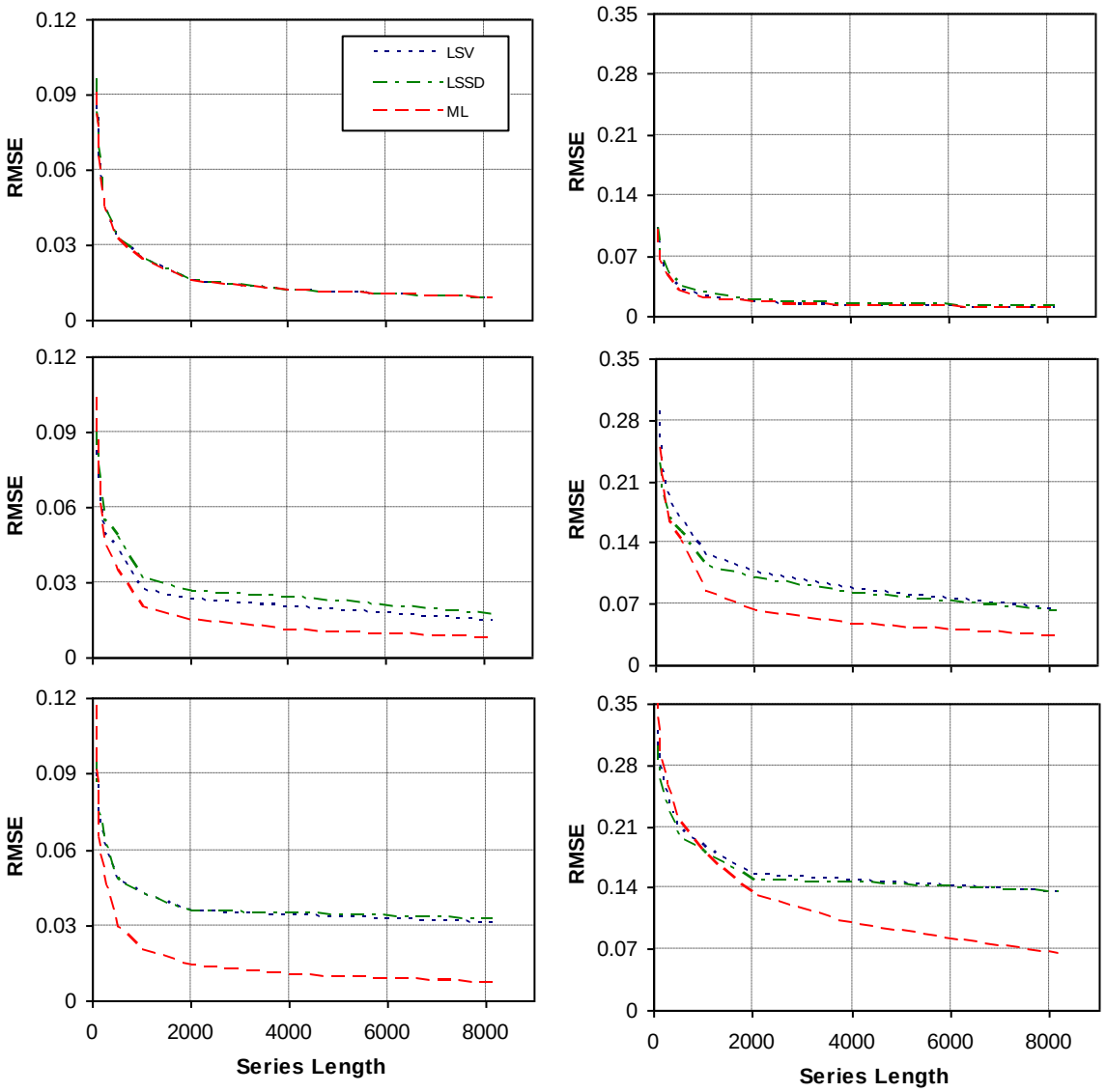


Figure 3.4. Root mean square error (RMSE) (left of the estimated H and right of the estimated σ) as a function of series length for all three estimators, with $H = 0.60$, $H = 0.90$ and $H = 0.95$ (upper to lower panels) and $\sigma = 1$.

A proof of the kind of dependence between the maximum likelihood estimates of the parameters could be given by the use of the Fisher Information Matrix $\boldsymbol{I}(\boldsymbol{\theta})$ with elements $\boldsymbol{I}_{ij}(\boldsymbol{\theta}) := -\mathrm{E}[\frac{\partial^2 \ln[l(\boldsymbol{\theta}|\boldsymbol{x}_n)]}{\partial\theta_i\,\partial\theta_j}]$, where $\boldsymbol{\theta} := (\theta_1, \theta_2, \theta_3) \equiv (\mu, \sigma, H)$. We easily calculate $\boldsymbol{I}_{12}(\boldsymbol{\theta}) = \boldsymbol{I}_{13}(\boldsymbol{\theta}) = 0$ and $\boldsymbol{I}_{23}(\boldsymbol{\theta}) = (1/\sigma)\,\mathrm{Tr}(\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\frac{\partial \boldsymbol{R}_{[1:n]\,[1:n]}}{\partial H}) \neq 0$ (see Section 3.6). Thus $\hat{\mu}$ and $\hat{H}$ are orthogonal and so are $\hat{\mu}$ and $\hat{\sigma}$, but not $\hat{\sigma}$ and $\hat{H}$.

Figure 3.8 presents the mean of the estimated ΔH and $\Delta\sigma$ along with their corresponding standard deviations from the ensemble versus q. An increase of q results to a decrease of bias when estimating H and an increase in the corresponding variance. The minimum bias when estimating σ, is achieved for values of q around 50 depending on the actual values of H, but

there is also an increase in the corresponding variance when q increases as expected from equations (3.21) and (3.30). It should be noted that a change of q does not influence the estimates when H is low, because $H^{q+1} / (q+1)$ is negligible for values of H near 0.5.

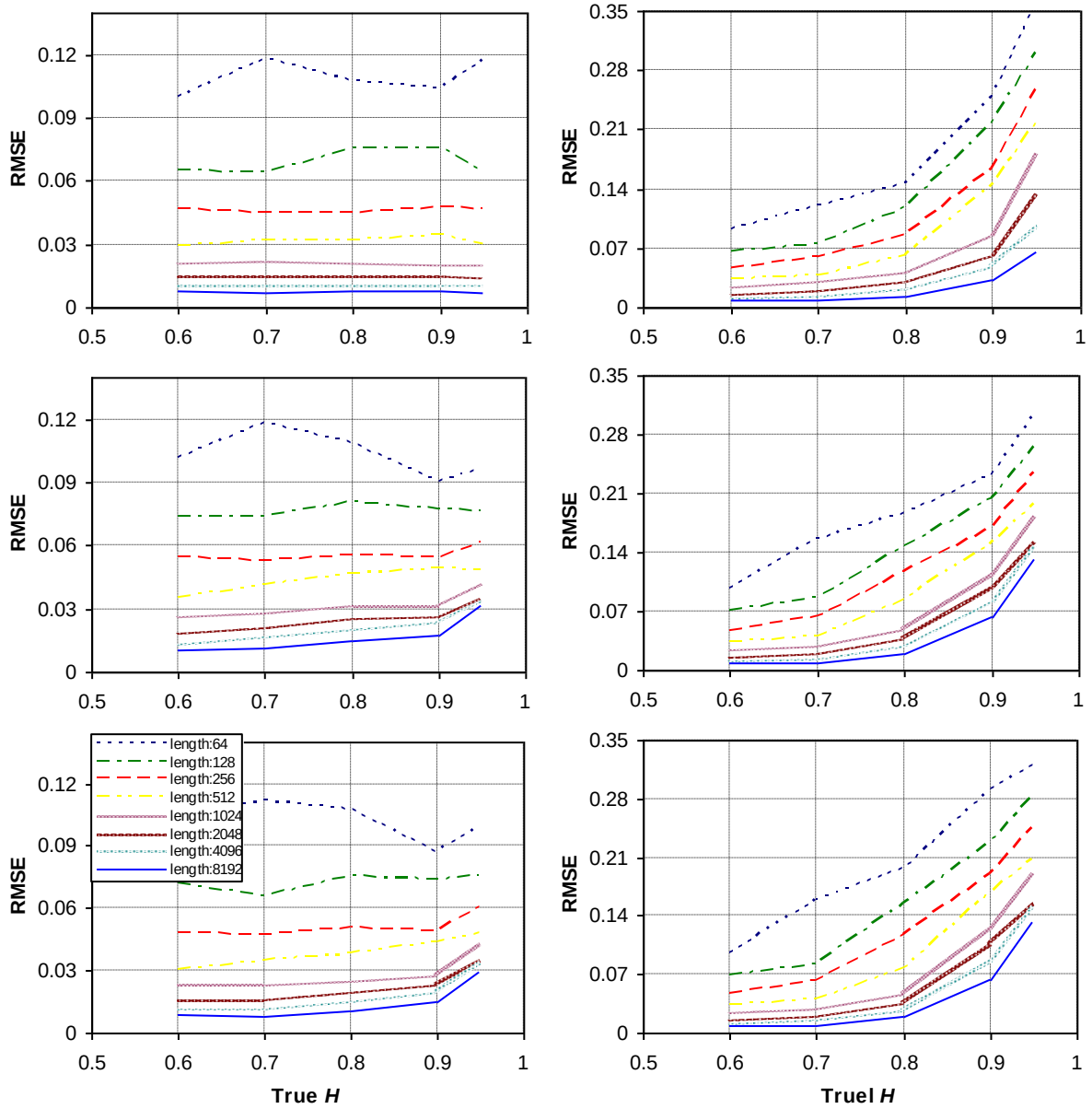


Figure 3.5. Root mean square error (RMSE) of H (left) and σ (right) as a function of true H for all lengths. Upper to lower panels correspond to ML, LSSD and LSV methods.

Figure 3.9 presents the mean of the estimated ΔH and $\Delta\sigma$ along with their corresponding standard deviations from the ensemble versus p. There is a range of p between 5 and 6, where we achieve minimum bias when estimating H or σ, but the corresponding variance decreases when p increases. We also note the irregularity between the graphs, caused by the presence of q, which gives smaller standard deviation of estimator for a high $H = 0.95$ rather than smaller H (e.g. $H = 0.90$).

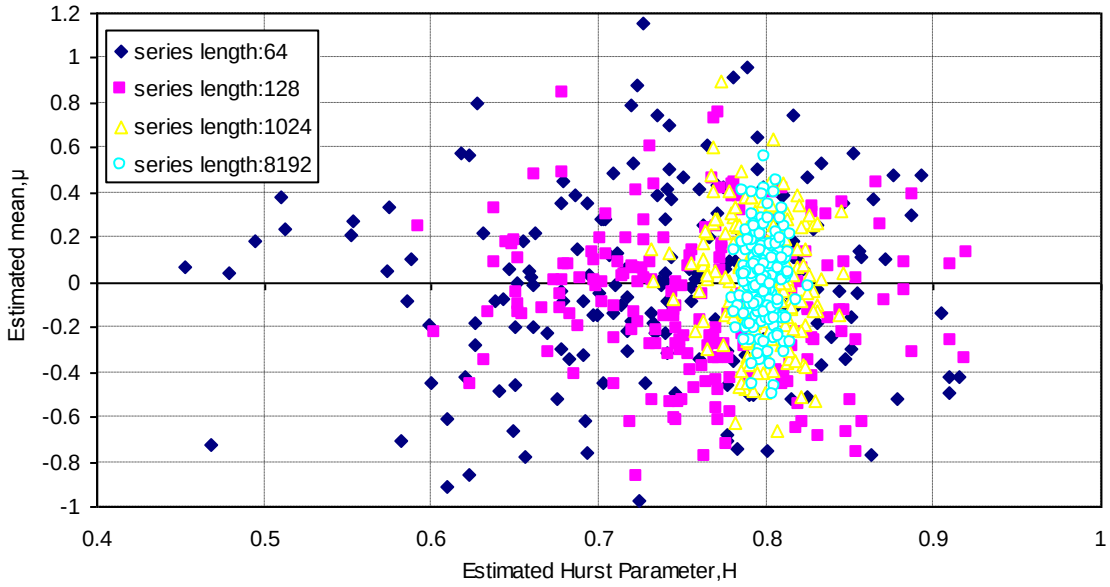


Figure 3.6. Estimated Hurst parameter H versus estimated mean μ from the ML method from 200 ensembles of synthetic time series with various lengths for true $H = 0.8$.

Figure 3.10 presents the mean of the estimated ΔH and $\Delta\sigma$ along with their corresponding standard deviations from the ensemble versus $m := n/\kappa'$. We observe that up to a value of $m = 10 \approx 1024/100$ the results remain the same, while for values of m more than 10 there is a higher bias and lower variance.

Finally we can see from Table 3.1 and Table 3.2 that these three methods perform better than the eight methods discussed in Taqqu et al. (1995) and the local Whittle estimator discussed in Robinson (1995).

3.8 Conclusions

It is clear from the simulations that the three estimators (ML, LSSD and LSV) are not equivalent, when compared to each other. Compared to other estimators of the literature, when estimating H, they seem to be more accurate and have a low error. This holds, because they have lower variance for large time series length and the other estimators rely on some asymptotic properties, whereas these estimators rely mostly on the structure of the HKp.

Table 3.1. Estimation results for H using 200 independent realizations 8 192 long where τ is the standard deviation of the sample containing the estimated H's. H's were estimated using Chen (2008) package, except the local Whittle estimates (Shimotsu 2004).

Estimation method		Nominal H				Estimation method	Nominal H			
		0.6	0.7	0.8	0.9		0.6	0.7	0.8	0.9
Variance	$\hat{H}$	0.595	0.687	0.775	0.850	R/S	0.619	0.706	0.784	0.854
	τ	0.027	0.027	0.026	0.027		0.031	0.032	0.031	0.032
	RMSE	0.027	0.030	0.036	0.057		0.036	0.033	0.035	0.055
DiffVar	$\hat{H}$	0.567	0.667	0.771	0.864	Periodogram	0.604	0.708	0.809	0.912
	τ	0.073	0.068	0.067	0.061		0.024	0.023	0.025	0.024
	RMSE	0.080	0.076	0.073	0.070		0.024	0.024	0.026	0.027
Absolute	$\hat{H}$	0.594	0.686	0.775	0.849	Modified Periodogram	0.565	0.661	0.752	0.847
	τ	0.028	0.027	0.028	0.029		0.037	0.038	0.037	0.034
	RMSE	0.029	0.031	0.038	0.059		0.051	0.054	0.060	0.063
Higuchi	$\hat{H}$	0.599	0.696	0.798	0.888	Local Whittle	0.601	0.700	0.804	0.902
	τ	0.028	0.029	0.040	0.044		0.023	0.023	0.022	0.021
	RMSE	0.028	0.029	0.040	0.046		0.023	0.023	0.023	0.021
Var. of Residuals	$\hat{H}$	0.600	0.702	0.801	0.896					
	τ	0.024	0.028	0.030	0.027					
	RMSE	0.024	0.028	0.030	0.027					

Note: *Variance*: a method based on aggregated variance; *DiffVar*: a method based on differencing the variance; *Absolute*: a method based on absolute values of the aggregated series; *Higuchi*: a method based on finding the fractal dimension; *Var. of Residuals*: a method based on residuals of regression, also known as Detrended Fluctuation Analysis (DFA); *R/S*: the original method by Hurst, based on the rescaled range statistic; *Periodogram*: a method based on the periodogram of the time series; *Modified Periodogram*: similar as the Periodogram method but with frequency axis divided into logarithmically equally spaced boxes and averaging the periodogram values inside the box (see details in Taqqu et al. 1995); *Local Whittle*: a semiparametric version of the Whittle estimator (see details in Robinson 1995).

Table 3.2. Estimation results for H using 200 independent realizations 8 192 long where τ is the standard deviation of the sample containing the estimated H's.

Estimation method		Nominal H			
		0.6	0.7	0.8	0.9
Maximum Likelihood	$\hat{H}$	0.599	0.700	0.799	0.899
	τ	0.008	0.007	0.008	0.007
	RMSE	0.008	0.007	0.008	0.007
Least Squares Standard Deviation	$\hat{H}$	0.599	0.699	0.799	0.892
	τ	0.011	0.011	0.015	0.015
	RMSE	0.011	0.012	0.015	0.017
Least Squares Variation	$\hat{H}$	0.599	0.700	0.800	0.895
	τ	0.009	0.008	0.011	0.014
	RMSE	0.009	0.008	0.011	0.015

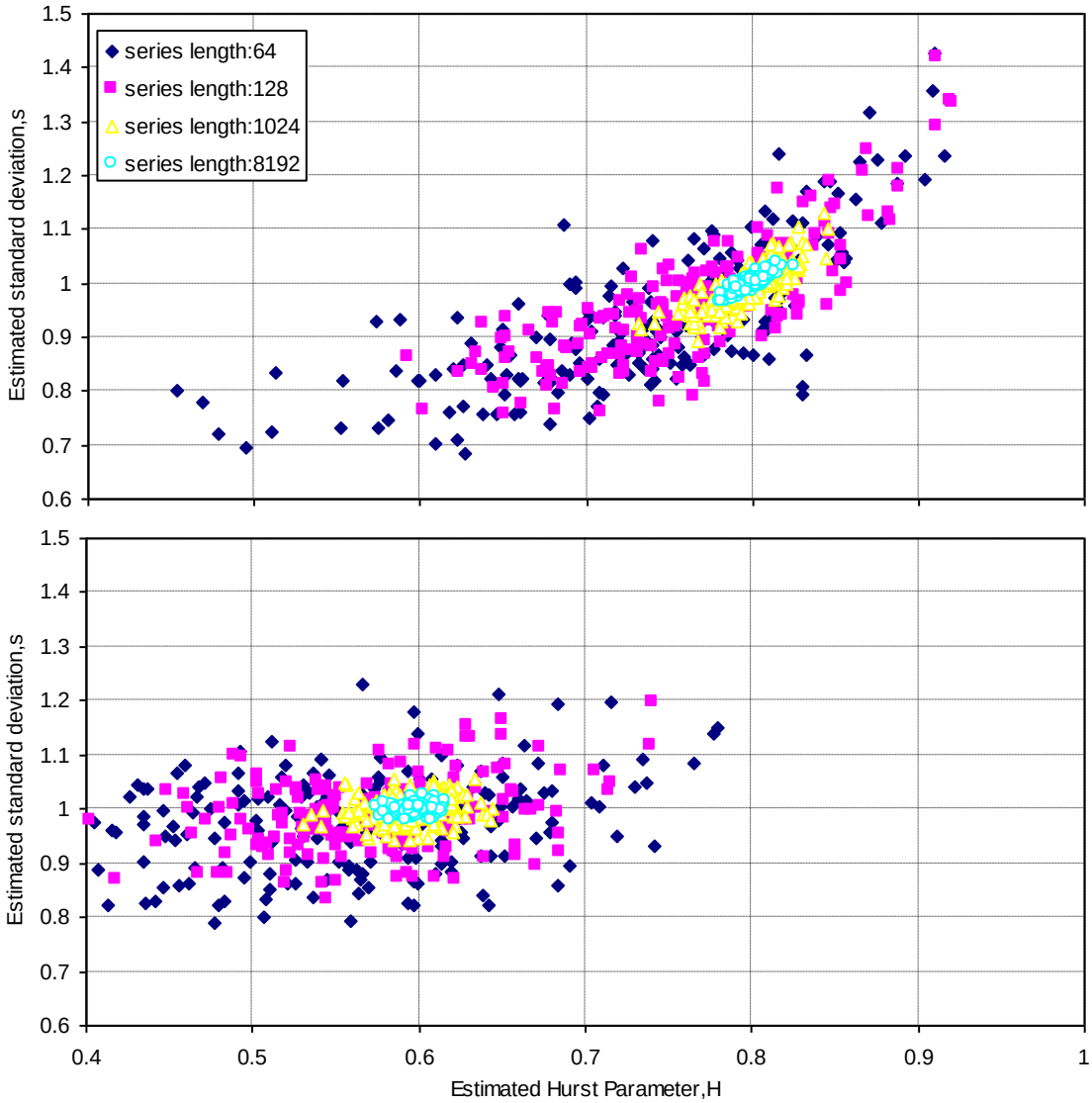


Figure 3.7. Estimated Hurst parameter H versus estimated standard deviation σ from the ML method from 200 ensembles of synthetic time series with various lengths. The upper diagram corresponds to true $H = 0.8$ and the lower diagram coresponds to true $H = 0.6$.

An additional advantage of these three estimators is that, in addition to H, they estimate σ which is essential for the model. As seen in Figure 3.7, and also proved in Section 3.7, $\hat{H}$ and $\hat{\sigma}$ are correlated and thus their maximum likelihood estimators cannot be calculated separately. Cox and Reid (1987) outline a number of statistical consequences of orthogonality. They state that the maximum likelihood estimate of H or σ when μ is given varies only slowly with μ. But this is not the case when examining σ versus H. As a consequence a non simultaneous estimator of σ and H may be suboptimal in terms of robustness compared to the ML, LSSD or LSV estimators which estimate H and σ simultaneously. From a more practical point of view, the importance of accounting for the dependence of the estimators, could be understood from the numerous publications that

calculate the standard deviation by the classical statistical estimator while at the same time find an $H > 0.5$, and sometimes very close to 1. Apparently, such estimates of standard deviation are heavily biased and this is a point which authors generally fail to note.

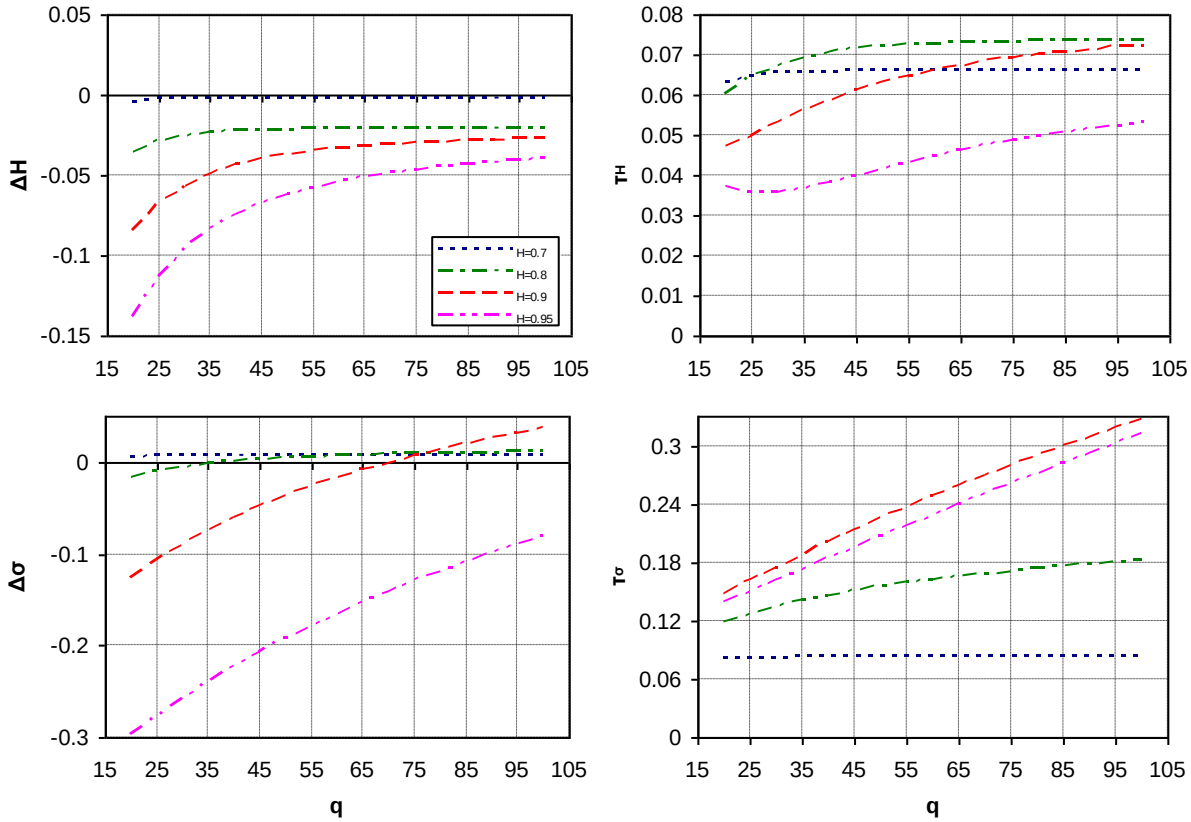


Figure 3.8. Mean of the estimated ΔH and $\Delta\sigma$ (left) and their corresponding standard deviations from 200 ensembles of synthetic time series 128 long (right) versus q, where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, τ_H and τ_σ are standard deviations and $p = 6$ for the LSV estimator. Definition of symbols used: $\tau_H := ((1/(200-1)) \sum_{k=1}^{200} (\Delta H_k)^2)^{1/2}$, $\tau_\sigma := ((1/(200-1)) \sum_{k=1}^{200} (\Delta\sigma)^2)^{1/2}$.

There are some problems with the choice of q or p in LSSD and LSV estimators. When choosing a large q we benefit from the fact that it decreases the variance of the σ estimator, but it causes an irregularity for high values of H, that cannot be controlled a priori. However we believe that the benefits from the presence of q are superior to the losses induced from its use, especially given that its presence does not affect the estimators for low values of H. For the choice of p the conflicting criteria of minimum bias and minimum variance of estimator should be considered. As a consequence, an a priori choice of p and q has a degree of subjectivity. In this study we chose $p = 6$ for LSV, $p = 2$ for LSSD and $q = 50$ for both methods, and the results were rather satisfactory. Additionally we chose $m = 10$, although Figure 3.10 allows to use lower m values. A choice of m below 10 does not influence the results.

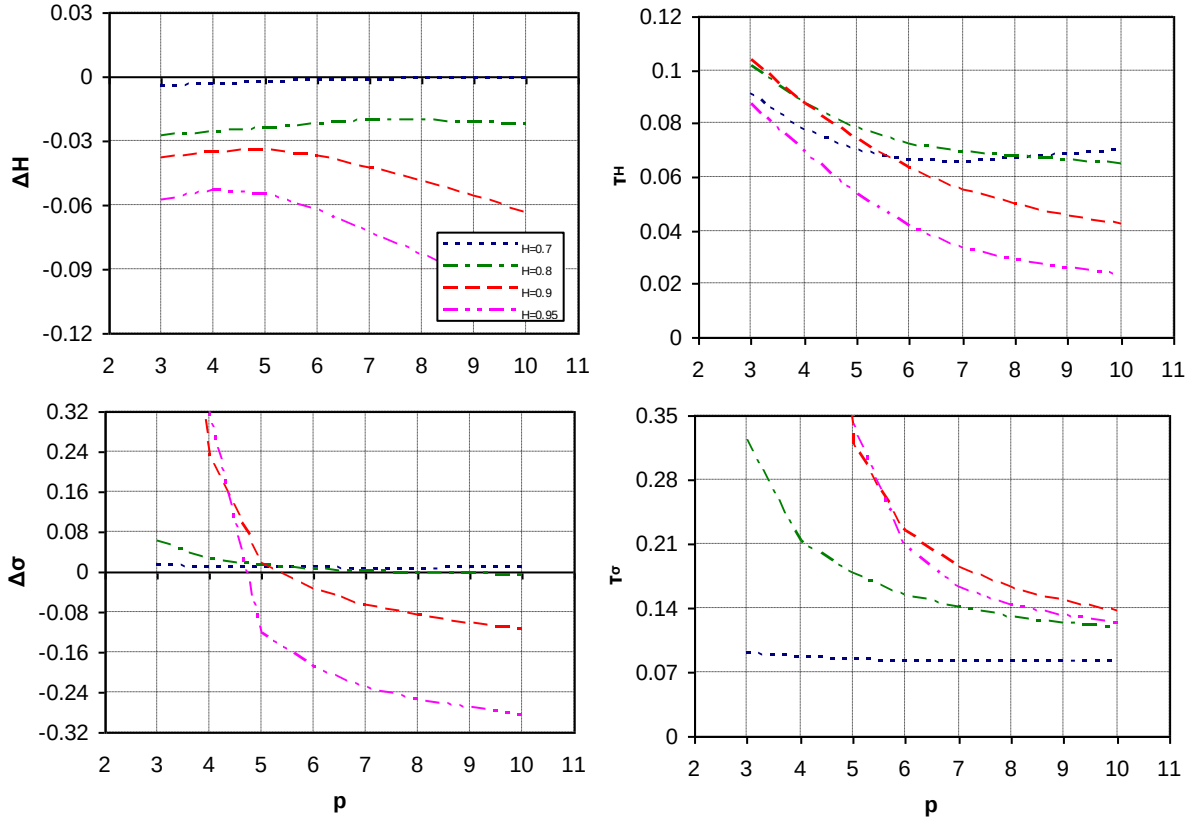


Figure 3.9. Mean of the estimated ΔH and $\Delta\sigma$ (left) and their corresponding standard deviations from 200 ensembles of synthetic time series 128 long (right) versus p, where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, τ_H and τ_σ are standard deviations and $q = 50$ for the LSV estimator. (See definition of symbols used in caption of Figure 3.8).

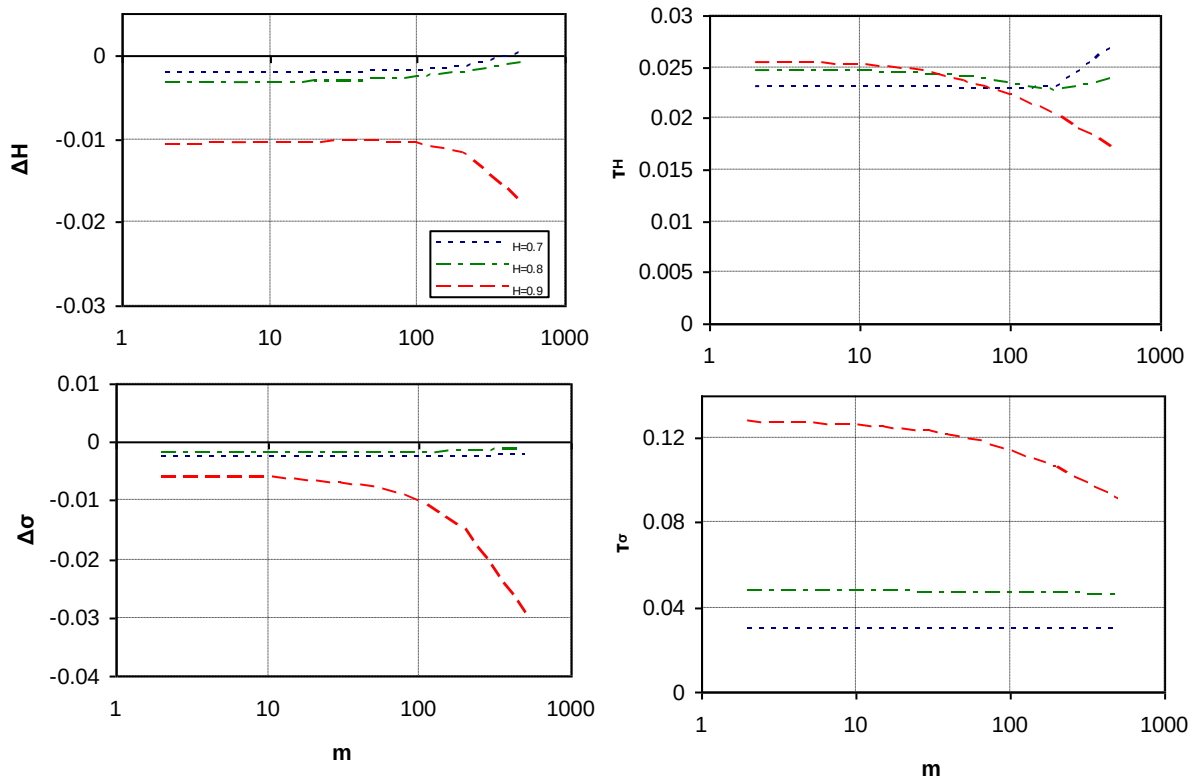


Figure 3.10. Mean of the estimated ΔH and $\Delta\sigma$ (left) and their corresponding standard deviations from 200 ensembles of synthetic time series 1 024 long (right) versus m, where $\Delta H = \hat{H} - H$, $\Delta\sigma = \hat{\sigma} - \sigma$, τ_H and τ_σ are standard deviations, $p = 6$ and $q = 50$ for the LSV estimator. (See definition of symbols used in caption of Figure 3.8).

Another strong point of these three estimators is that they are easy to understand, again because they rely on the structure of the HKp. They also enable some interesting theoretical analyses such as those presented here, namely the bracketing of H and the behaviour of the estimator for high values of H.

There is a problem with the implementation of the ML estimator, because it needs large computational times for large time series lengths (e.g. many thousands of data values). But in hydrology the available time series are usually short. Thus, we think that its use is preferable, when an estimation of the HKp parameters is required. When the time series length increases we can switch to the LSV or the LSSD method. Among the three estimators, the ML estimator is better when estimating H, followed by the LSV method. But when estimating σ the LSSD method is superior to the LSV method.

4. The predictive distribution of hydroclimatic variables

The HKp[1] entails high autocorrelations even for large lags, as well as high variability even at climatic scales. A problem that, thus, arises is how to incorporate the observed past hydroclimatic data in deriving the predictive distribution of hydroclimatic processes at climatic time scales. Here with the use of Bayesian techniques we create a framework to solve the aforementioned problem. We assume that there is no prior information for the parameters of the process and use a noninformative prior distribution. We apply this method with real-world data to derive the posterior distribution of the parameters and the posterior predictive distribution of various 30-year moving average climatic variables. The marginal distributions we examine are the normal and the truncated normal (for nonnegative variables). We also compare the results with two alternative models, one that assumes independence in time and one with Markovian dependence, and the results are dramatically different. The conclusion is that this framework is appropriate for the prediction of future hydroclimatic variables conditional on the observations.

4.1 Introduction

A lot of work has been done in predicting the future of hydroclimatic processes using Bayesian statistics. Berliner et al. (2000) applied a Markov model to a low-order dynamical system of tropical Pacific SST, using a hierarchical Bayesian dynamical modelling, which led to realistic error bounds on forecasts. Duan et al. (2007) illustrated how the Bayesian model averaging (BMA) scheme can be used to generate probabilistic hydrologic predictions from several competing individual predictions. Kumar and Maity (2008) used two different Bayesian dynamic modelling approaches, namely a constant model and a dynamic regression model (DRM) to forecast the volume of the Devil's lake. Maity and Kumar (2006) used a Bayesian dynamic linear model to predict the monthly Indian summer monsoon rainfall. Bakker and Hurk (2012) used a Bayesian model to predict multi-year geostrophic winds.

On the other hand, GCMs give deterministic projections of future hydroclimatic processes for some hypothesized scenarios e.g. for the increase of CO_2 concentration, etc. However, the uncertainty of these projections whose sources may be attributed to insufficient current understanding of climatic mechanisms, to inevitable weaknesses of numerical climatic and hydrologic models to represent processes and scales of interest, to complexity of processes

[1] Based on: Tyralis and Koutsoyiannis (2014)

and to unpredictability of causes (Koutsoyiannis et al. 2007), is not estimated by these models. Consequently, it is impossible to estimate whether any observed changes reflect the natural variability of the climatic processes or should be attributed to external forcings. Additionally, using deterministic projections and thus neglecting the uncertainty in future hydroclimatic conditions, may result in underestimation of possible range of the future hydroclimatic variation.

Koutsoyiannis et al. (2007) have done some work on the uncertainty assessment of future hydroclimatic predictions. They propose a stochastic framework for future climatic uncertainty, where climate is expressed by the 30-year time average of a natural process exhibiting HK behaviour. To this end, they combine analytical and Monte Carlo methods to determine uncertainty limits and they apply the framework developed to temperature, rainfall and runoff data from a catchment in Greece, for which measurements are available for about a century.

In the study by Koutsoyiannis et al. (2007), the climatic variability and the influence of parameter uncertainty are studied separately. As a result, a hydroclimatic prediction needs two confidence coefficients to be defined, one referring to the uncertainty of the climatic evolution and one to the uncertainty of model parameters. In this Chapter we unify the study of the two uncertainties so that a climatic prediction needs only one confidence coefficient to be defined. To this end, we solve the problem of climatic predictions of natural processes using Bayesian statistics, instead of the stochastic framework developed by Koutsoyiannis et al. (2007). For physical consistency with natural processes such as rainfall and runoff, whose values are nonnegative, we also examine the case where truncation of the negative part of the distributions is applied. No prior information for the parameters of processes is assumed, so that the prior distribution is noninformative. The posterior joint distribution is derived from a mixture for the case where truncation is not applied and a Gibbs sampler for the case where truncation is applied. We derive the posterior predictive distribution (Gelman et al. 2004 p.8) of the process in closed form given the posterior distribution of the parameters. We simulate a sample from the posterior predictive distribution and use it to make inference about the future evolution of the averaged process. We apply this procedure using the same data as in Koutsoyiannis et al. (2007), and specifically runoff (Case 1 or C1), rainfall (C2) and temperature (C3) data from catchments in Greece and temperature data from Berlin (C4, C6 with the last 90 years excluded from the dataset); in addition we used temperature data from Vienna (C5, C7 with the last 90 years excluded from the dataset). For the rainfall and runoff

data we use truncated distributions.

As per the temporal dependence of the processes, three alternative assumptions are made: (a) independence in time; (b) Markovian dependence modelled by first-order autoregressive (AR(1)) process; and (c) HK dependence (see Markonis and Koutsoyiannis 2013, for a justification of the latter). In the last Section we compare the results of the three models. Additional results such as the posterior distributions of the parameters and the asymptotic behaviour of the predictive distribution are also given.

While this Chapter uses the same case studies as those in Koutsoyiannis et al. (2007), the results are not directly comparable to each other. Here we give posterior predictive distributions of the climatic variables, whereas Koutsoyiannis et al. (2007) give confidence limits for specified quantiles of climatic variables. The posterior predictive distribution of the variables given here is exactly what we call climatic prediction, whereas we could say that the confidence limits of the quantiles, given by Koutsoyiannis et al. (2007), are intermediate or indirect results. The Bayesian methodology applied here aims at (stochastic) prediction (Robert 2007 p.7) and is direct, while its disadvantage compared to Koutsoyiannis et al. (2007) framework is the much heavier computational burden.

4.1.1 Definition of AR(1)

We assume that $\{\underline{x}_t\}$, $t = 1, 2, \ldots$ is a normal stationary stochastic process with parameters given by (1.1)-(1.4). We assume that $\{\underline{a}_t\}$ is a zero mean normal white noise process (WN), i.e. a sequence of independent random variables from a normal distribution with mean $\mathrm{E}[\underline{a}_t] = 0$ and variance $\mathrm{Var}[\underline{a}_t] = \sigma_a^2$. In the following discussion $\{\underline{a}_t\}$ is always referred to as WN. The following equation defines the first-order autoregressive process AR(1).

$$\underline{x}_t - \mu = \varphi_1(\underline{x}_{t-1} - \mu) + \underline{a}_t,\ |\varphi_1| < 1 \tag{4.1}$$

The ACF of the AR(1) is (Wei 2006 p.34)

$$\rho_k = \varphi_1^k,\ k = 0, 1, \ldots \tag{4.2}$$

4.2 Posterior distribution of the parameters of a stationary normal stochastic process

The distribution of the variable $\underline{\boldsymbol{x}}_{1:n} = (\underline{x}_1, \ldots, \underline{x}_n)^{\mathrm{T}}$ from a normal stationary stochastic process is

$$f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta}) = (2\pi)^{-n/2}\, |\sigma^2\, \boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \exp[(-1/2\sigma^2)\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)] \tag{4.3}$$

where $\boldsymbol{R}_{[1:n]\,[1:n]}$ is the autocorrelation matrix with elements $r_{ij} = \rho_{|i-j|}$, $i,j = 1,2, \ldots,n$. Details on the distributions used thereafter are given in Appendix A. The autocorrelation $\rho_{|i-j|}$ is assumed to be function of a parameter (scalar or vector) $\boldsymbol{\varphi}$, so that $\boldsymbol{\theta} := (\mu, \sigma^2, \boldsymbol{\varphi})$ is the parameter vector of the process. We note that if $\underline{\boldsymbol{x}}_n$ is white noise then $\rho_0 = 1$ and $\rho_k = 0$, $k = 1, 2, \ldots$; if it is AR(1) then ρ_k is given by (4.2) if it is HKp then ρ_k is given by (1.9).

We assume that $\boldsymbol{\varphi}$ is uniformly distributed a priori. We set as prior distribution for $\underline{\boldsymbol{\theta}}$ the noninformative distribution (see also Robert 2007 example 3.5.6)

$$\pi(\boldsymbol{\theta}) \propto 1/\sigma^2 \tag{4.4}$$

The posterior distribution of the parameters does not have a closed form. However it can be calculated from a mixture based on conditional distributions. Specifically, it is shown (see Section 4.4) that

$$\underline{\mu}|\sigma^2, \boldsymbol{\varphi}, \boldsymbol{x}_{1:n} \sim N[(\boldsymbol{x}_{1:n}^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n)/(\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n), \sigma^2/(\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n)] \tag{4.5}$$

$$\underline{\sigma}^2|\boldsymbol{\varphi}, \boldsymbol{x}_{1:n} \sim \text{Inv-gamma}\{(n-1)/2, [\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n \cdot \boldsymbol{x}_{1:n}^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{x}_{1:n} - (\boldsymbol{x}_{1:n}^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n)^2]/(2\, \boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n)\} \tag{4.6}$$

$$\pi(\boldsymbol{\varphi}|\boldsymbol{x}_{1:n}) \propto |\boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} [\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n \cdot \boldsymbol{x}_{1:n}^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{x}_{1:n} - (\boldsymbol{x}_{1:n}^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n)^2]^{-(n-1)/2} (\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n)^{n/2-1} \tag{4.7}$$

As real world problems often impose upper or lower bounds on the variables $\underline{x}_t$, we assume that the distribution of $\underline{\boldsymbol{x}}_{1:n}$ is two-sided truncated by bounds a and b, i.e.,

$$f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta}) \propto \exp[(-1/2\sigma^2) (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)]\, \mathrm{I}_{[a,b]^n}(x_1, \ldots, x_n) \tag{4.8}$$

where I denotes the indicator function, so that $\mathrm{I}_{[a,b]^n}(x_1, \ldots, x_n) = 1$ if $\boldsymbol{x}_n \in [a,b]^n$ and 0 otherwise.

We assume that the truncation set of μ is $[a,b]$, $a,b \in \mathrm{R} \cup \{-\infty,\infty\}$. The following Gibbs sampler is used to obtain a posterior sample from $\underline{\boldsymbol{\theta}} = (\underline{\mu}, \underline{\sigma}^2, \underline{\boldsymbol{\varphi}})$ (see Section 4.4).

$$\pi(\mu|\sigma^2, \boldsymbol{\varphi}, \boldsymbol{x}_{1:n}) \propto \exp\{-[\mu - (\boldsymbol{x}_{1:n}^{\mathrm{T}} \cdot \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n)/(\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n)]^2/(2\sigma^2/\boldsymbol{e}_n^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} \boldsymbol{e}_n)\}\, \mathrm{I}_{[a,b]}(\mu) \tag{4.9}$$

$$\underline{\sigma}^2|\mu, \boldsymbol{\varphi}, \boldsymbol{x}_{1:n} \sim \text{Inv-gamma}\{n/2, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)/2\} \tag{4.10}$$

$$\pi(\boldsymbol{\varphi}|\, \mu, \sigma^2, \boldsymbol{x}_{1:n}) \propto |\boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \exp[-(\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}} \boldsymbol{R}_{[1:n]\,[1:n]}^{-1} (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)/2\sigma^2] \tag{4.11}$$

4.3 Posterior predictive distributions

As we stated in Section 4.1, we seek to make an inference about the future evolution of a process given observations of its past. To this end, in this Section we derive the posterior predictive distributions of $\underline{\boldsymbol{x}}_{(n+1):(n+m)}|\boldsymbol{x}_{1:n}$ for the cases of the white noise, the AR(1) and the HKp, where $\underline{\boldsymbol{x}}_{(n+1):(n+m)} := (\underline{x}_{n+1},\ldots,\underline{x}_{n+m})^{\mathrm{T}}$.

4.3.1 White noise

We assume that $\underline{x}_t$, $t = 1, 2, \ldots$ is white noise, with $f(x_t|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp[-(x_t-\mu)^2/(2\sigma^2)]$. A noninformative prior distribution for $\underline{\boldsymbol{\theta}} = (\underline{\mu},\underline{\sigma}^2)$ is $\pi(\boldsymbol{\theta}) \propto 1/\sigma^2$. The posterior distributions of the parameters are given by (Gelman et al. 2004 p.75-77)

$$\underline{\mu}|\boldsymbol{x}_{1:n} \sim \mathrm{t}_{n-1}(\underline{x}_1^{(n)}, s_n^2/n) \tag{4.12}$$

$$\underline{\sigma}^2|\boldsymbol{x}_{1:n} \sim \text{Inv-gamma}((n-1)/2, ((n-1)\, s_n^2)/2) \tag{4.13}$$

Notice that (4.12) and (4.13) are derived from (4.5),(4.6),(4.7) for $\boldsymbol{R}_n = \boldsymbol{I}_n$ (the former after integrating out σ^2). The posterior predictive distribution is

$$\underline{x}_t|\boldsymbol{x}_{1:n} \sim \mathrm{t}_{n-1}(\underline{x}_1^{(n)},((n+1)/n)s_n^2),\ t = n+1, n+2, \ldots \tag{4.14}$$

where $\underline{x}_{n+1}$, $\underline{x}_{n+2},\ldots$ are mutually independent and $\mathrm{t}_\nu(\mu,\sigma^2)$ is the Student's distribution with ν degrees of freedom.

4.3.2 AR(1) and HKp

When there is dependence among the elements of $\underline{\boldsymbol{x}}_{1:(n+m)}$, the posterior predictive distribution of $\underline{\boldsymbol{x}}_{(n+1):(n+m)}$ given $\boldsymbol{\theta}$ and $\boldsymbol{x}_{1:n}$ is (Eaton 2007 p.116,117)

$$f(\boldsymbol{x}_{(n+1):(n+m)}|\boldsymbol{\theta},\boldsymbol{x}_{1:n}) = (2\pi\sigma^2)^{-m/2}\, |\boldsymbol{R}_{m|n}|^{-1/2} \exp[(-1/2\sigma^2)$$

$$(\boldsymbol{x}_{(n+1):(n+m)} - \boldsymbol{\mu}_{m|n})^{\mathrm{T}}\, \boldsymbol{R}_{m|n}^{-1}\, (\boldsymbol{x}_{(n+1):(n+m)} - \boldsymbol{\mu}_{m|n})] \tag{4.15}$$

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

$$\boldsymbol{\mu}_{m|n} = \mu\boldsymbol{e}_m + \boldsymbol{R}_{[(n+1):(n+m)]\,[1:n]}\, \boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\, (\boldsymbol{x}_{1:n} - \mu\boldsymbol{e}_n) \tag{4.16}$$

$$\boldsymbol{R}_{m|n} = \boldsymbol{R}_{[(n+1):(n+m)]\,[(n+1):(n+m)]} - \boldsymbol{R}_{[1:n]\,[(n+1):(n+m)]}^{\mathrm{T}}\, \boldsymbol{R}_{[1:n]\,[1:n]}^{-1}\, \boldsymbol{R}_{[1:n]\,[(n+1):(n+m)]} \tag{4.17}$$

where $\boldsymbol{R}_{[k:l]\,[m:n]}$ is the submatrix of $\boldsymbol{R}$ which contains the elements r_{ij}, $k \le i \le l$, $m \le j \le n$. The elements of the correlation matrices $\boldsymbol{R}_{1:n}$ and $\boldsymbol{R}_{1:(n+m)}$ are obtained from (4.2) for the case of the AR(1) and from 1.9) for the case of HKp. In the implementation of the AR(1) model we

assume that all three parameters μ, σ, φ_1 are unknown. For the HKp we examine two cases: (a) all three parameters μ, σ, H, are unknown, and (b) μ, σ, are unknown but H is considered to be known and equal to its maximum likelihood estimate (Chapter 3).

In the case that all three parameters of the AR(1) or HKp are unknown, we obtain a simulated sample of $\underline{\boldsymbol{\theta}}$ from (4.5),(4.6),(4.7) and use this sample to simulate $\underline{\boldsymbol{\mu}}_{m|n}$ and $\underline{\boldsymbol{R}}_{m|n}$ from (4.16) and (4.17) and generate a sample of $\underline{\boldsymbol{x}}_{(n+1):(n+m)}$ from (4.15). In the case where H is considered as known, we obtain a simulated sample of $\underline{\boldsymbol{\theta}} = (\underline{\mu}, \underline{\sigma}^2)$ from (4.5),(4.6) and use this sample to simulate $\underline{\boldsymbol{\mu}}_{m|n}$ and $\underline{\boldsymbol{R}}_{m|n}$ from (4.16) and (4.17) and generate a sample of $\underline{\boldsymbol{x}}_{(n+1):(n+m)}$ from (4.15).

4.3.3 Asymptotic behaviour of AR(1) and HKp

In most applications, it is useful to know the ultimate confidence regions as prediction horizon tends to infinity. This is expressed by the distribution of $\underline{\boldsymbol{x}}_{(n+m+1):(n+m+l)} := (\underline{x}_{n+m+1},\ldots,\underline{x}_{n+m+l})^{\mathrm{T}}$ as $m \to \infty$, conditional on $\boldsymbol{x}_{1:n}$. For given $\boldsymbol{\theta}$ this distribution is:

$$f(\boldsymbol{x}_{(n+m+1):(n+m+l)}|\boldsymbol{\theta},\boldsymbol{x}_{1:n}) = (2\pi\sigma^2)^{-l/2}|\boldsymbol{R}_{l|n}|^{-1/2} \cdot \exp[(-1/2\sigma^2)(\boldsymbol{x}_{(n+m+1):(n+m+l)} - \boldsymbol{\mu}_{l|n})^{\mathrm{T}}\boldsymbol{R}_{l|n}^{-1}(\boldsymbol{x}_{(n+m+1):(n+m+l)} - \boldsymbol{\mu}_{l|n})] \quad (4.18)$$

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

$$\boldsymbol{\mu}_{l|n} = \mu\boldsymbol{e}_l + \boldsymbol{R}_{[(n+m+1):(n+m+l)]\,[1:n]}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,(\boldsymbol{x}_{1:n} - \mu\boldsymbol{e}_n) \quad (4.19)$$

$$\boldsymbol{R}_{l|n} = \boldsymbol{R}_{[(n+m+1):(n+m+l)]\,[(n+m+1):(n+m+l)]} - \boldsymbol{R}^{\mathrm{T}}_{[1:n]\,[(n+m+1):(n+m+l)]}\,\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\,\boldsymbol{R}_{[1:n]\,[(n+m+1):(n+m+l)]} \quad (4.20)$$

We observe that, as $m \to \infty$, $\boldsymbol{R}_{[1:n]\,[(n+m+1):(n+m+l)]}$ and $\boldsymbol{R}_{[(n+m+1):(n+m+l)]\,[1:n]}$ become zero matrices and $\boldsymbol{R}_{[(n+m+1):(n+m+l)]\,[(n+m+1):(n+m+l)]} = \boldsymbol{R}_{[1:l]\,[1:l]}$. This implies that:

$$\boldsymbol{\mu}_{l|n} = \mu\boldsymbol{e}_l \quad (4.21)$$

$$\boldsymbol{R}_{l|n} = \boldsymbol{R}_{[1:l]\,[1:l]} \quad (4.22)$$

where $\boldsymbol{R}_l$ is again obtained from (4.2) for the case of the AR(1) and from (1.9) for the case of HKp.

Accordingly, the application can proceed as follows. We obtain a simulated sample of $\underline{\boldsymbol{\theta}}$ from (4.5),(4.6),(4.7) and use this sample to simulate $\underline{\boldsymbol{\mu}}_{l|n}$ and $\underline{\boldsymbol{R}}_{l|n}$ from (4.21) and (4.22) and generate a sample of $\underline{\boldsymbol{x}}_{(n+m+1):(n+m+l)}$ from (4.18) for a large m.

4.3.4 Truncated white noise, AR(1) and HKp

To examine real world problems which often impose upper or lower bounds on the variables $\underline{x}_t$, we assume that the distribution of $\underline{\boldsymbol{x}}_{1:n}$ is two-sided truncated, and is given by (4.8). We obtain a posterior sample of $\underline{\boldsymbol{\theta}}$ using the Gibbs sampler defined by (4.9), (4.10), (4.11). When $\boldsymbol{\varphi}$ is known, we obtain a posterior sample of $(\underline{\mu},\underline{\sigma}^2)$ using the Gibbs sampler defined by (4.9) and (4.10). Then $\underline{\boldsymbol{x}}_{1:m}|\boldsymbol{\theta}$ follows a truncated normal multivariate distribution and according to Horrace (2005) the conditional multivariate distributions of $\underline{\boldsymbol{x}}_{(n+1):(n+m)}|\boldsymbol{\theta},\boldsymbol{x}_{1:n}$ are again truncated normal. As a result (4.15) still holds after slight modifications and (4.16), (4.17) are valid. The posterior predictive distribution of $\underline{\boldsymbol{x}}_{(n+1):(n+m)}|\boldsymbol{\theta},\boldsymbol{x}_{1:n}$ is then a multivariate truncated normal distribution:

$$f(\boldsymbol{x}_{(n+1):(n+m)}|\boldsymbol{\theta},\boldsymbol{x}_{1:n}) \propto \exp[(-1/2\sigma^2)(\boldsymbol{x}_{(n+1):(n+m)}-\boldsymbol{\mu}_{m|n})^{\mathrm{T}}\boldsymbol{R}_{m|n}^{-1}(\boldsymbol{x}_{(n+1):(n+m)}-\boldsymbol{\mu}_{m|n})]\mathrm{I}_{[a,b]^m}(\boldsymbol{x}_{n+1,n+m}) \qquad (4.23)$$

Now for the case of white noise, (4.12), (4.13) and (4.14) are not valid. But from (4.16), (4.17) and for $\rho_0 = 1$ and $\rho_k = 0$, $k = 1, 2, \ldots$, we obtain that $\boldsymbol{\mu}_{m|n} = \mu\boldsymbol{e}_m$ and $\boldsymbol{R}_{m|n} = \boldsymbol{R}_{[1:m]\,[1:m]}$.

When looking for the asymptotic behaviour of the process, (4.18) still holds after slight modifications, according to Horrace (2005). As a result, the distribution of $\underline{\boldsymbol{x}}_{(n+m+1):(n+m+l)}|\boldsymbol{\theta},\boldsymbol{x}_{1:n}$ is truncated multivariate normal, while (4.21) and (4.22) remain valid:

$$f(\boldsymbol{x}_{(n+m+1):(n+m+l)}|\boldsymbol{\theta},\boldsymbol{x}_{1:n}) \propto \exp[(-1/2\sigma^2)(\boldsymbol{x}_{(n+m+1):(n+m+l)} - \boldsymbol{\mu}_{l|n})^{\mathrm{T}} \cdot \boldsymbol{R}_{l|n}^{-1}(\boldsymbol{x}_{(n+m+1):(n+m+l)} - \boldsymbol{\mu}_{l|n})]\ \mathrm{I}_{[a,b]^l}(\boldsymbol{x}_{(n+m+1):(n+m+l)}) \qquad (4.24)$$

4.3.5 Asymptotic convergence of MCMC

To simulate from (4.7) we use a random walk Metropolis-Hastings algorithm with a normal instrumental (or proposal) distribution (Robert and Casella 2004 p.271). We implement the algorithm using the function MCMCmetrop1R of the R package 'MCMCpack' (Martin et al. 2011). The variable 'burnin' in this package is given the value 0, whereas the other variables keep their default values.

There are a lot of methods to decide whether convergence can be assumed to hold for the generated sample (see Gamerman and Lopes 2006 p.157-169; Robert and Casella 2004 p.272-276). We use the methods of Heidelberger and Welch (1983) and Raftery and Lewis (1992). These methods are described by Smith (2007), whose notation we use here. We use the R package 'coda' (Plummer et al. 2006) to implement these methods. We assume that we have obtained a sample $\psi_1, \psi_2, \ldots$ of a scalar variable φ using the MCMC algorithm.

The diagnostic of Heidelberger's method provides an estimate of the number of samples that should be discarded as a burn-in sequence and a formal test for non-convergence. The null hypothesis of convergence to a stationary chain is based on Brownian bridge theory and uses the Cramer-von-Mises test statistic $\int_0^1 B_n(t)^2 dt$, where

$$B_n(t) = (T_{\lfloor nt \rfloor} - \lfloor nt \rfloor \bar{\psi})/\sqrt{nS(0)} \tag{4.25}$$

$$T_k = \sum_{j=1}^{k} \psi_j,\ k = 1, 2, \ldots \text{ and } T_0 = 0 \tag{4.26}$$

where $\lfloor x \rfloor$ denotes the floor of x (the greatest integer not greater than x) and $S(0)$ is the spectral density evaluated at frequency zero. In calculating the test statistic, the spectral density is estimated from the second half of the original chain. If the null hypothesis is rejected, then the first $0.1n$ of the samples are discarded and the test is reapplied to the resulting chain. This process is repeated until the test is either non-significant or 50% of the samples have been discarded, at which point the chain is declared to be non-stationary. For more details see Smith (2007).

The methods of Raftery and Lewis are designed to estimate the number of MCMC samples needed when quantiles are the posterior summaries of interest. Their diagnostic is applicable for the univariate analysis of a single parameter and chain. For instance, let us consider the estimation of the following posterior probability of a model parameter θ:

$$P(f(\theta) < a \mid \mathbf{x}) = q \tag{4.27}$$

where $\boldsymbol{x}$ denotes the observed data. Raftery and Lewis sought to determine the number of MCMC samples to generate and the number of samples to discard in order to estimate q to within $\pm r$ with probability s. In practice, users specify the values of q, r and s to be used in applying the diagnostic (For more details see Smith, 2007).

To simulate from (4.11) we use an accept-reject algorithm (Robert and Casella 2004 p.51-53) with a uniform instrumental density. Simulation from (4.9) and (4.10) is trivial. We assess the convergence of the chain simulated from (4.9), (4.10), (4.11) using the method of Gelman and Rubin (1992; see also Gelman 1996; Gamerman and Lopes 2006 p.166-168). An indicator of convergence is formed by the estimator of a potential scale reduction (PSR) that is always larger than 1. Convergence can be evaluated by the proximity of PSR to 1. Gelman

(1996) suggested accepting convergence when the value of PSR is below 1.2.

4.4 Mathematical proofs

In Section 4.4 the proofs of (4.5),(4.6),(4.7),(4.9),(4.10),(4.11) are given. It is easily shown that

$$(\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n) = \boldsymbol{e}_n^{\mathrm{T}}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n\, \mu^2 - 2\, \boldsymbol{x}_{1:n}^{\mathrm{T}}\, \boldsymbol{R}_n^{-1}\, \boldsymbol{e}_n\, \mu + \boldsymbol{x}_{1:n}^{\mathrm{T}}\, \boldsymbol{R}_n^{-1}\, \boldsymbol{x}_{1:n} \qquad (4.28)$$

After completing the squares the above expression becomes:

$$\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, \boldsymbol{e}_n\mu^2 - 2\boldsymbol{x}_{1:n}^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{e}_n\mu + \boldsymbol{x}_{1:n}^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{x}_n = \boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{e}_n[\mu - (\boldsymbol{x}_{1:n}^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{e}_n) / (\boldsymbol{e}_n^{\mathrm{T}}$$
$$\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{e}_n)]^2 + [\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{e}_n\boldsymbol{x}_{1:n}^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{x}_n - (\boldsymbol{x}_{1:n}^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{e}_n)^2] / (\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{e}_n) \qquad (4.29)$$

From (4.3) and (4.4) we obtain the following:

$$\pi(\boldsymbol{\theta})\, f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta}) \propto \sigma^{-(n+2)}\, |\boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \exp[(-1/2\sigma^2)\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)] \qquad (4.30)$$

From (4.28),(4.29) and (4.30) we obtain (4.5). After integration of (4.30) we obtain (4.31) which proves (4.6):

$$\pi(\sigma^2|\boldsymbol{\varphi},\boldsymbol{x}_{1:n}) \propto (\sigma^2)^{-(n+1)/2}\, |\boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \exp[(-1/2\sigma^2)[\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{e}_n\boldsymbol{x}_{1:n}^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{x}_n - (\boldsymbol{x}_{1:n}^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}$$
$$\boldsymbol{e}_n)^2]/(\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\boldsymbol{e}_n)] \qquad (4.31)$$

After integration of (4.30) we obtain (4.32), which proves (4.7) after integration:

$$\pi(\boldsymbol{\varphi}|\boldsymbol{x}_{1:n}) \propto \iint \sigma^{-(n+2)}\, |\boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \exp[(-1/2\sigma^2)\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)]\, \mathrm{d}\mu\, \mathrm{d}\sigma^2 \qquad (4.32)$$

See also Falconer and Fernadez (2007) for some results.

Now for the case where truncation is applied we obtain from (4.4) and (4.8):

$$\pi(\boldsymbol{\theta})\, f(\boldsymbol{x}_{1:n}|\boldsymbol{\theta}) \propto \sigma^{-(n+2)}\, |\boldsymbol{R}_{[1:n]\,[1:n]}|^{-1/2} \cdot$$
$$\exp[(-1/2\sigma^2)\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)^{\mathrm{T}}\, \boldsymbol{R}^{-1}_{[1:n]\,[1:n]}\, (\boldsymbol{x}_{1:n} - \mu\, \boldsymbol{e}_n)]\, \mathrm{I}_{[a,b]^n}(x_1, \ldots, x_n) \qquad (4.33)$$

Conditional on $\mu \in [a,b]$, $a,b \in \mathrm{R} \cup \{-\infty,\infty\}$the derivation of (4.9), (4.10) and (4.11) from (4.33) is then trivial.

4.5 Case studies

In this Section we apply the methodology developed in the previous Sections to five historical datasets; three of them obtained from the Boeoticos Kephisos River basin, one from Berlin and one from Vienna. The choice of these datasets was dictated by the fact that they have been also studied in other works with similar objectives, i.e. Koutsoyiannis et al. (2007) and

Koutsoyiannis (2011), so that the interested reader can make some comparisons. We present the results of the application of the methodology to the aforementioned datasets.

4.5.1 Historical datasets

The first case study is performed on an important catchment in Greece, which is part of the water supply system of Athens and has a history, as regards hydraulic infrastructure and management that extends backward at least 3500 years. This is the closed (i.e. without outlet to the sea) basin of the Boeoticos Kephisos River (Figure 4.1), with an area of 1955.6 km^2, mostly formed over a karstic subsurface. Owing to its importance for irrigation and water supply, data availability for the catchment extends for about 100 years (the longest dataset in Greece) and modelling attempts with good performance have already been carried out on the hydrosystem (Rozos et al. 2004).

Figure 4.1. The Boeoticos Kephisos River basin.

The long-term dataset for the basin extends from 1908 to 2003 and comprises a flow record at the river outlet at the Karditsa station (C1), rainfall observations in the raingage Aliartos (C2) and a temperature record at the same station (C3); the station locations are shown in Figure 4.1. Further details on the construction of these datasets are given by Koutsoyiannis et al. (2007). The relatively long records have already made it possible to identify the scaling behaviour of rainfall and runoff in this basin (Koutsoyiannis 2003), and make the catchment ideal for a case study of uncertainty assessment.

The two other datasets which we use are the mean annual temperature record of Berlin/Templehof and Vienna, two of the longest series of instrumental meteorological

observations. For further details on the Berlin mean annual temperature dataset see Koutsoyiannis et al. (2007) and for the Vienna mean annual temperature dataset see Koutsoyiannis (2011). We examine two cases. In the first case we assume that the update of the prior information is done (C4, C5), using the whole dataset. In the second case the update is done excluding the last 90 years of the datasets (C6, C7).

4.5.2 Application of the method

We classified the data into three classes, the first containing the data from the Boeoticos Kephisos River basin (C1-C3), the second containing the data from Berlin and Vienna (C4, C6) and the third containing again the data from Berlin and Vienna (C5, C7) but excluding the last 90 years. In the third case the posterior results were compared to the actual 90 last years.

Table 4.1. Summarized results and maximum likelihood estimates for the cases of WN, AR(1) and HKp at Boeoticos Kephisos River basin.

	Boeoticos basin		
	Runoff (mm)	Rainfall (mm)	Temperature (°C)
Start year	1908	1908	1898
End year	2003	2003	2003
Size, n	96	96	106
WN			
$\hat{\mu}$	197.63	658.36	16.96
$\hat{\sigma}$	81.25	155.82	0.69
AR(1)			
$\hat{\mu}$	197.65	658.22	16.96
$\hat{\sigma}$	81.22	155.81	0.69
$\hat{\varphi}_1$	0.34	0.10	0.31
HK			
$\hat{\mu}$	195.11	657.38	16.97
$\hat{\sigma}$	80.47	155.00	0.70
$\hat{H}$	0.71	0.60	0.71

First we calculated the maximum likelihood estimates of the parameters for all the examined cases (WN, AR(1), HKp). The results are given in Table 4.1 and Table 4.2. Truncated models were used for C1 and C2 datasets due to the relatively high estimated σ which otherwise would result in negative values. Instead, when we examined the temperature datasets (C3-C7), simulated values near the absolute zero never appeared, indicating a good behaviour of the non-truncated model.

The procedure for the temperature datasets is described below. We used (4.12) and (4.13) to generate a posterior sample from $\underline{\mu}$ and $\underline{\sigma}^2$ for the WN case. To simulate from (4.7) for the φ_1 and $\underline{H}$ posterior distribution of the AR(1) and HK cases correspondingly, we used a

random walk Metropolis-Hastings algorithm. We simulated a single chain with 3 000 000 MCMC samples. The Metropolis acceptance rates are given in Table 4.3. To decide whether convergence has been achieved, we used the Heidelberger and Welch method (1983). We tested four cases, the first case containing all the 3 000 000 samples, the second containing the last 2 000 000 samples and so forth. The results are presented in Table 4.4 and Table 4.5, from where we conclude that stationary chain hypothesis holds in every case. We also used the methods of Raftery and Lewis (1992), to estimate the number of MCMC samples needed when quantiles are the posterior summaries of interest. The minimum number of samples and the burn-in period for the simulation is given in Table 4.6 and Table 4.7, where q = 0.025, 0.500, 0.975 are the quantiles to be estimated, r = 0.005 is the desired margin of error of the estimate and s = 0.95 is the probability of obtaining an estimate in the interval $(q-r, q+r)$. We decided to use the last 2 000 000 samples of the chains, to obtain the histograms of the posterior distributions of the parameters $\underline{\varphi}_1$ and $\underline{H}$. The simulation of $\underline{\mu}$, $\underline{\sigma}^2$ from (4.5) and (4.6) is then trivial. Summarized results for the parameters of the AR(1) and HK cases respectively are shown in Table 4.8 and Table 4.9.

Table 4.2. Summarized results and maximum likelihood estimates for the cases of WN, AR(1) and HKp at Berlin and Vienna.

	Berlin	Vienna	Berlin	Vienna
	Temperature (°C)	Temperature (°C)	Temperature (°C)	Temperature (°C)
Start year	1756	1775	1756	1775
End year	2009	2009	1919	1919
Size, n	254	235	164	145
WN				
$\hat{\mu}$	9.17	9.58	9.04	9.36
$\hat{\sigma}$	0.91	0.87	0.92	0.84
AR(1)				
$\hat{\mu}$	9.18	9.58	9.05	9.36
$\hat{\sigma}$	0.92	0.87	0.92	0.84
$\hat{\varphi}_1$	0.37	0.30	0.30	0.11
HK				
$\hat{\mu}$	9.27	9.64	9.10	9.37
$\hat{\sigma}$	0.91	0.86	0.92	0.84
$\hat{H}$	0.73	0.70	0.70	0.59

From the simulated samples we obtained the posterior probability plots of $\underline{\mu}$, $\underline{\sigma}$, $\underline{H}$, $\underline{\varphi}_1$ for the AR(1) and HK cases (Figures 4.2-4.8). The last 100 000 simulated samples of the parameters, described in the previous paragraph were used to obtain samples from the required posterior predictive probabilities. The samples from the posterior predictive probability of $\underline{x}_t|\mathbf{x}_{1:n}$, $t = n+1, n+2,..., n+90$ were used to obtain samples for the variable of interest $\underline{x}_{t(30)}$ given by (4.34).

$$\underline{x}_{t(30)} := (1/30)\Big(\sum_{l=t-29}^{n} x_l + \sum_{l=n+1}^{t} \underline{x}_l\Big),\ t = n+1, \ldots, n+29 \text{ and}$$

$$\underline{x}_{t(30)} := (1/30)\sum_{l=t-29}^{t} \underline{x}_l,\ t = n+30, n+31, \ldots \tag{4.34}$$

Table 4.3. Metropolis acceptance rate for the MCMC simulation of φ_1 and H, respectively, at Boeoticos Kephisos River basin.

	Aliartos temperature	Berlin temperature (1756-2009)	Vienna temperature (1775-2009)	Berlin temperature (1756-1919)	Vienna temperature (1775-1919)
φ_1	0.70731	0.70603	0.70612	0.70649	0.70654
H	0.706037	0.70551	0.70599	0.70601	0.70638

Table 4.4. Heidelberger and Welch test, for significance level 0.05, at Boeoticos Kephisos River basin.

	Aliartos temperature							
Parameter	φ_1				H			
Stationarity test	passed	passed	passed	passed	passed	passed	passed	passed
Start iteration	1	1	1	1	1	1	1	1
p-value	0.427	0.745	0.46	0242	0.869	0.567	0.338	0.618

Table 4.5. Heidelberger and Welch test, for significance level 0.05, at Berlin and Vienna.

	Berlin temperature (1756-2009)				Vienna temperature (1775-2009)			
Data start	1	1000000	2000000	2900000	1	1000000	2000000	2900000
Parameter	φ_1				φ_1			
Stationarity test	passed	passed	passed	passed	passed	passed	passed	passed
Start iteration	1	1	1	1	1	1	1	1
p-value	0.943	0.738	0.342	0.448	0.928	0.696	0.366	0.0761
Parameter	H				H			
Stationarity test	passed	passed	passed	passed	passed	passed	passed	passed
Start iteration	1	1	1	1	1	1	1	1
p-value	0.837	0.466	0.279	0.691	0.789	0.501	0.296	0.84
	Berlin temperature (1756-1919)				Vienna temperature (1775-1919)			
Parameter	φ_1				φ_1			
Stationarity test	passed	passed	passed	passed	passed	passed	passed	passed
Start iteration	1	1	1	1	1	1	1	1
p-value	0.94	0.589	0.376	0.425	0.777	0.55	0.308	0.592
Parameter	H				H			
Stationarity test	passed	passed	passed	passed	passed	passed	passed	passed
Start iteration	1	1	1	1	1	1	1	1
p-value	0.833	0.606	0.339	0.923	0.885	0.83	0.373	0.323

We examined the cases of WN, AR(1), asymptotic behaviour of AR(1), HK where H is considered to be known and has the value of the maximum likelihood estimate, HK when H is not known, and its asymptotic behaviour. Figures 4.9-4.11 show the 0.025, 0.500 and 0.975 quantiles of the posterior predictive distributions of $\underline{x}_{t(30)}|\boldsymbol{x}_{1:n}$, $t = n+1, n+2,..., n+90$.

The procedure for C1 and C2 is described below. We simulated from (4.9), (4.10) and (4.11) to obtain a posterior sample from $\underline{\mu}$, $\underline{\sigma}^2$ and $\underline{\varphi}$ for all cases. We simulated 10 chains with each one having 300 000 MCMC samples. To decide whether convergence has been

achieved, we used the Gelman and Rubin (1992) rule. In all cases PSR ≈ 1 which shows that the chains converged to the target distribution. We decided to use the last 200 000 samples of each chain, to obtain the histograms of the posterior distributions of the parameters $\underline{\varphi}_1$ and $\underline{H}$. Summarized results for the parameters of the AR(1) and HK cases respectively are shown in Table 4.8.

Table 4.6. Raftery and Lewis test for the case of Boeoticos Kephisos River basin.

		Aliartos temperature							
	q	Burn-in	Total	Lower bound	Dependence factor		Burn-in	Total	Lower bound	Dependence factor
$\underline{\varphi}_1$	0.025	21	31794	3746	8.49	$\underline{H}$	18	35784	4899	7.3
	0.500	24	356752	38415	9.29		24	464024	50239	9.24
	0.975	28	32298	3746	8.62		28	42161	4899	8.61

Note: q is the quantile to be estimated, $r = 0.005$ is the desired margin of error of the estimate, $s = 0.95$ the probability of obtaining an estimate in the interval $(q-r, q+r)$, eps = 0.001 is the precision required for estimating time to convergence.

Table 4.7. Raftery and Lewis test for the cases of Berlin and Vienna.

		Berlin temperature (1756-2009)				Vienna temperature (1775-2009)			
	q	Burn-in	Total	Lower bound	Dependence factor	Burn-in	Total	Lower bound	Dependence factor
$\underline{\varphi}_1$	0.025	21	31416	3746	8.39	21	31612	3746	8.44
	0.500	24	356512	38415	9.28	21	322441	38415	8.39
	0.975	21	31731	3746	8.47	21	31745	3746	8.47
$\underline{H}$	0.025	18	27288	3746	7.28	18	35670	4899	7.28
	0.500	21	322777	38415	8.4	21	422975	50239	8.42
	0.975	28	32732	3746	8.74	28	42882	4899	8.75
		Berlin temperature (1756-1919)				Vienna temperature (1775-1919)			
$\underline{\varphi}_1$	0.025	21	31780	3746	8.48	21	31780	3746	8.48
	0.500	24	356656	38415	9.28	21	323631	38415	8.42
	0.975	21	32193	3746	8.59	21	32137	3746	8.58
$\underline{H}$	0.025	18	27330	3746	7.3	18	27072	3746	7.23
	0.500	21	323330	38415	8.42	21	324177	38415	8.44
	0.975	18	32991	3746	8.81	27	39690	3746	10.6

Note: q is the quantile to be estimated, $r = 0.005$ is the desired margin of error of the estimate, $s = 0.95$ the probability of obtaining an estimate in the interval $(q-r, q+r)$, eps = 0.001 is the precision required for estimating time to convergence.

From the simulated samples we obtained the posterior probability plots of $\underline{\mu}$, $\underline{\sigma}$, $\underline{H}$, $\underline{\varphi}_1$ for the AR(1) and HK cases (Figure 4.2 and Figure 4.3). The last 10 000 simulated samples of the parameters of each chain, described in the previous paragraph are used to obtain samples from the required posterior predictive probabilities. The samples from the posterior predictive probability of $\underline{x}_t|\boldsymbol{x}_{1:n}$, $t = n+1, n+2,..., n+90$ are used to obtain samples for the variable of interest $\underline{x}_t^{(30)}$ given by (4.34). We examined the cases of WN, AR(1), asymptotic behaviour of AR(1), HK where H is considered to be known and has the value of the maximum likelihood estimate, HK with unknown H and its asymptotic behaviour. Figure 4.9 shows the 0.025, 0.500 and 0.975 quantiles of the posterior predictive distributions of $\underline{x}_{t(30)}|\boldsymbol{x}_{1:n}$, $t = n+1, n+2,...,$

n+90.

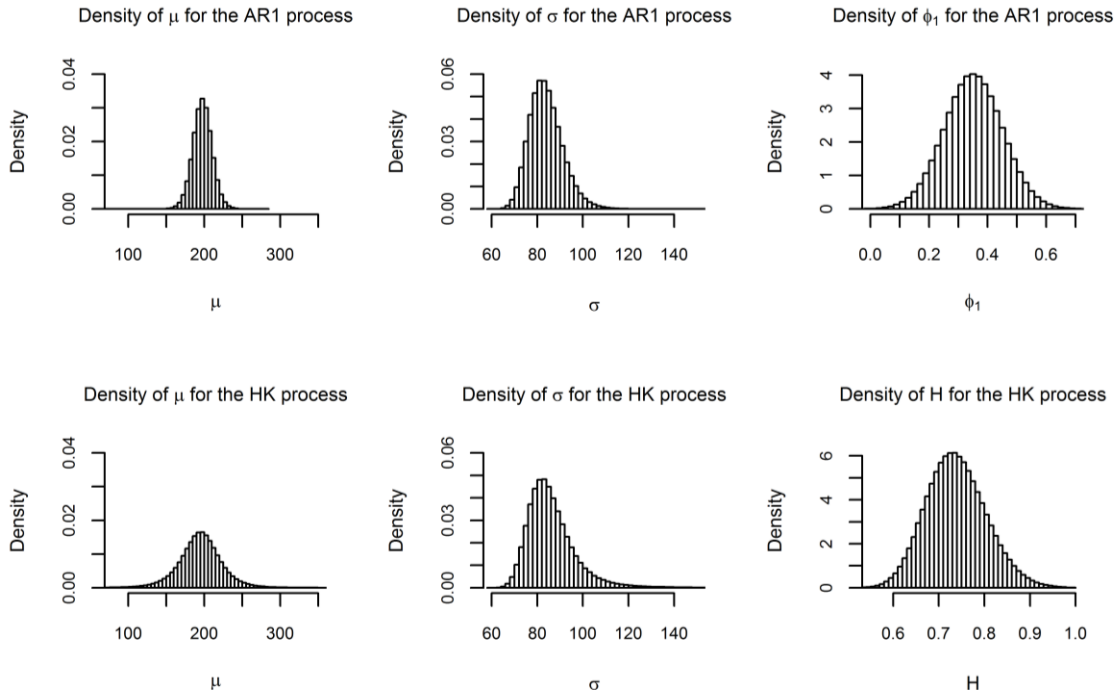


Figure 4.2. Posterior probability distributions of $\underline{\mu}$, $\underline{\sigma}$, $\underline{H}$, $\underline{\varphi}_1$ for the cases of AR(1) and HK processes, for the runoff of Boeoticos Kephisos.

4.5.3 Results

A first important result of the proposed framework is that it provides good estimates of the model parameters without introducing any assumptions (i.e., using noninformative priors). While common statistical methods give point estimates of parameters, the Bayesian framework provides also interval estimates based on their posterior distributions. The estimated values of μ are given in Table 4.10. It turns out that irrespective of the method used (MLE or posterior medians) they are almost equal. When examining temperatures, HKp resulted in the largest $\hat{\mu}$ and AR(1) in the second largest. In C4 and C6, $\hat{\mu}$ was larger than in C5 and C7 respectively. From the density diagrams of the posterior distributions (Figures 4.2-4.8) it seems that the posterior distribution of $\underline{\mu}$ is wider when HKp is used. The posterior distribution of $\underline{\sigma}$ is also wider on the right (see the values of the 0.975 quantiles in Table 4.8 and Table 4.9) for the HKp. However the estimated values of σ are almost equal for the three used models (Table 4.1 and Table 4.2). The estimated φ_1 and H are given in Table 4.1 and Table 4.2. Their estimated values for C5 are considerably higher compared to C7, but their posterior distributions are narrower (Table 4.9), probably because of the bigger sample size in the former case. Their posterior distributions are also narrower for C4 compared to C6.

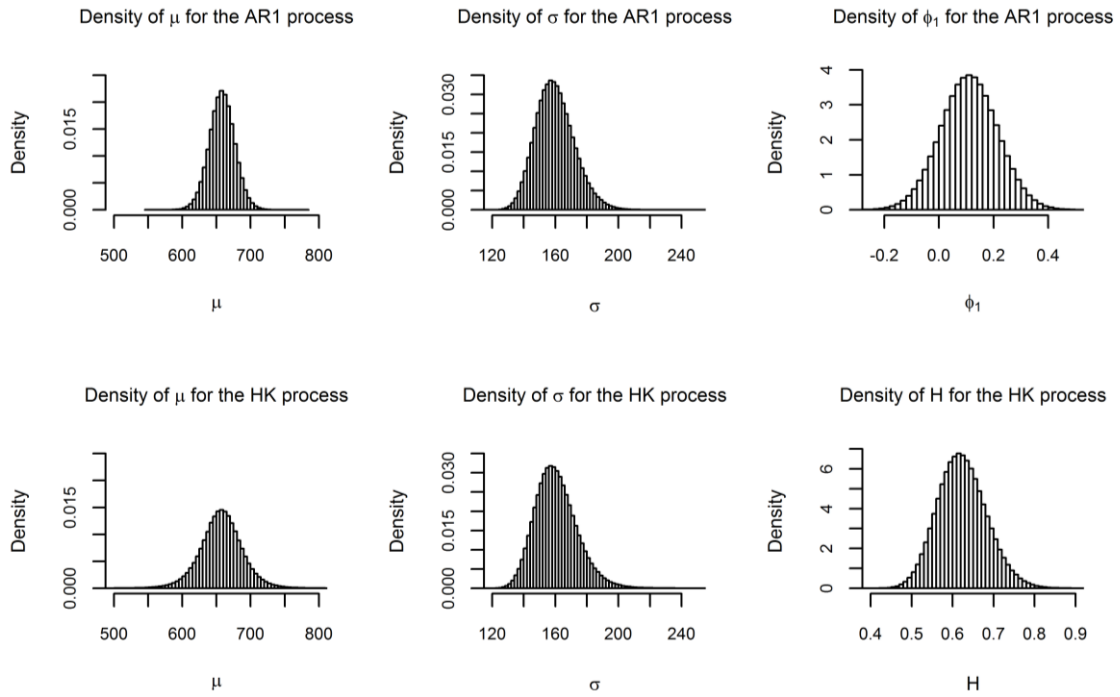


Figure 4.3. Posterior probability distributions of $\underline{\mu}$, $\underline{\sigma}$, $\underline{H}$, $\underline{\varphi}_1$ for the cases of AR(1) and HK processes, for the rainfall at Aliartos.

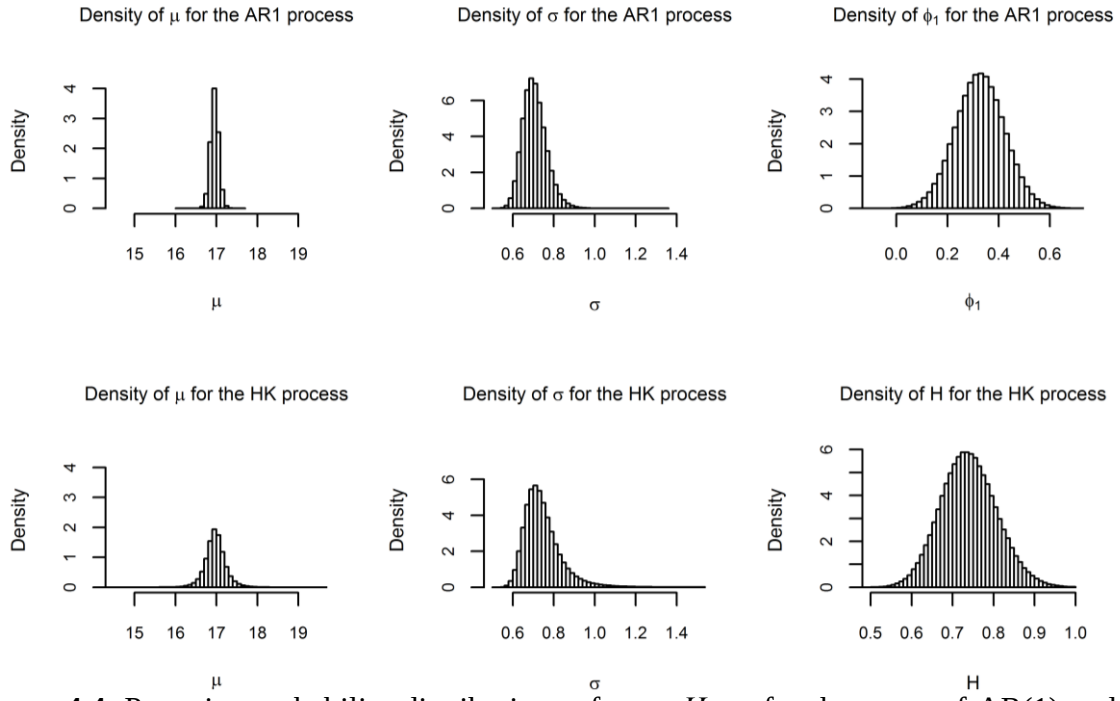


Figure 4.4. Posterior probability distributions of $\underline{\mu}$, $\underline{\sigma}$, $\underline{H}$, $\underline{\varphi}_1$ for the cases of AR(1) and HK processes, for the temperature at Aliartos.

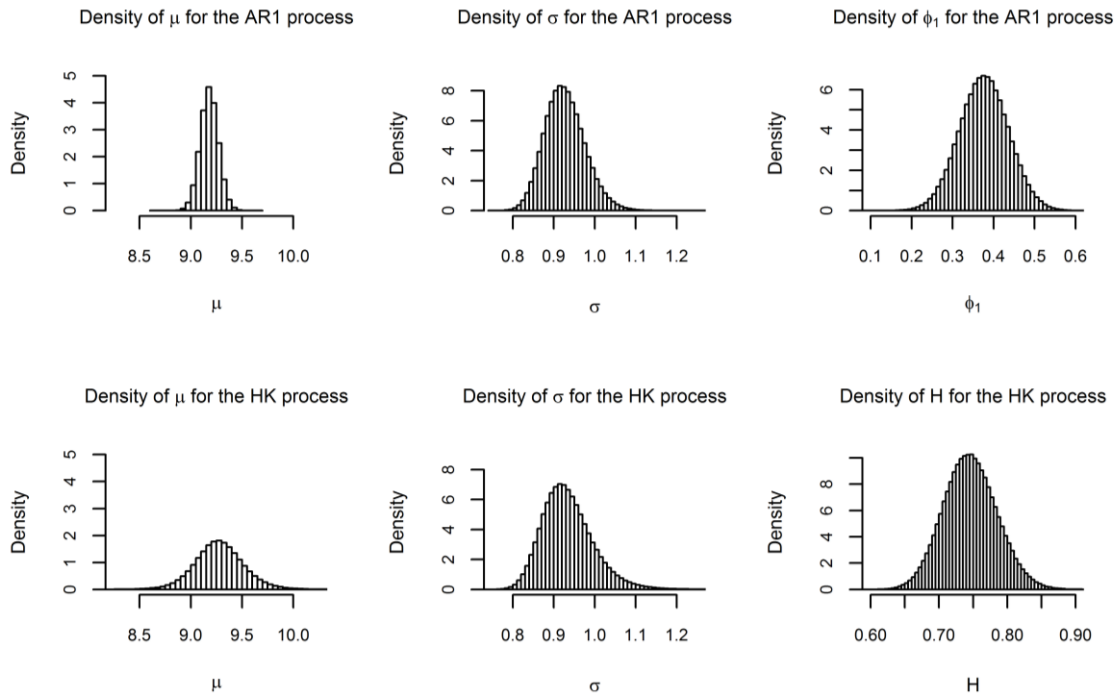


Figure 4.5. Posterior probability distributions of $\underline{\mu}$, $\underline{\sigma}$, $\underline{H}$, $\underline{\varphi}_1$ for the cases of AR(1) and HK processes, for the temperature at Berlin/Tempelhof. In this case the parameters are estimated from years 1756-2009.

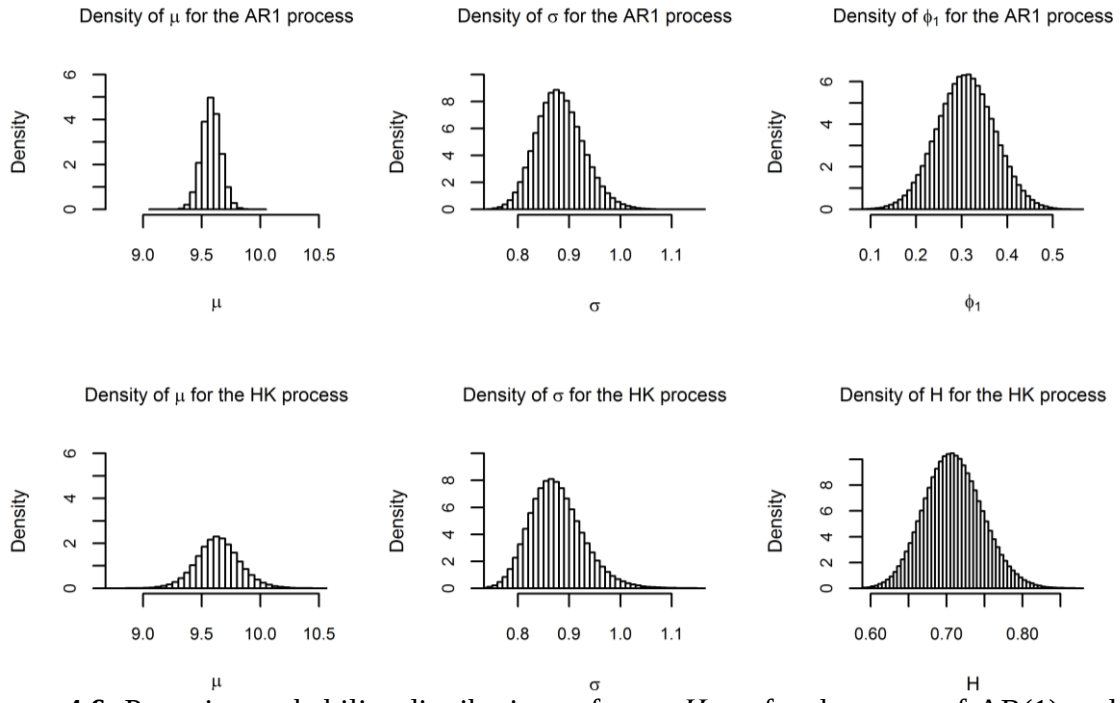


Figure 4.6. Posterior probability distributions of $\underline{\mu}$, $\underline{\sigma}$, $\underline{H}$, $\underline{\varphi}_1$ for the cases of AR(1) and HK processes, for the temperature at Vienna. In this case the parameters are estimated from years 1775-2009.

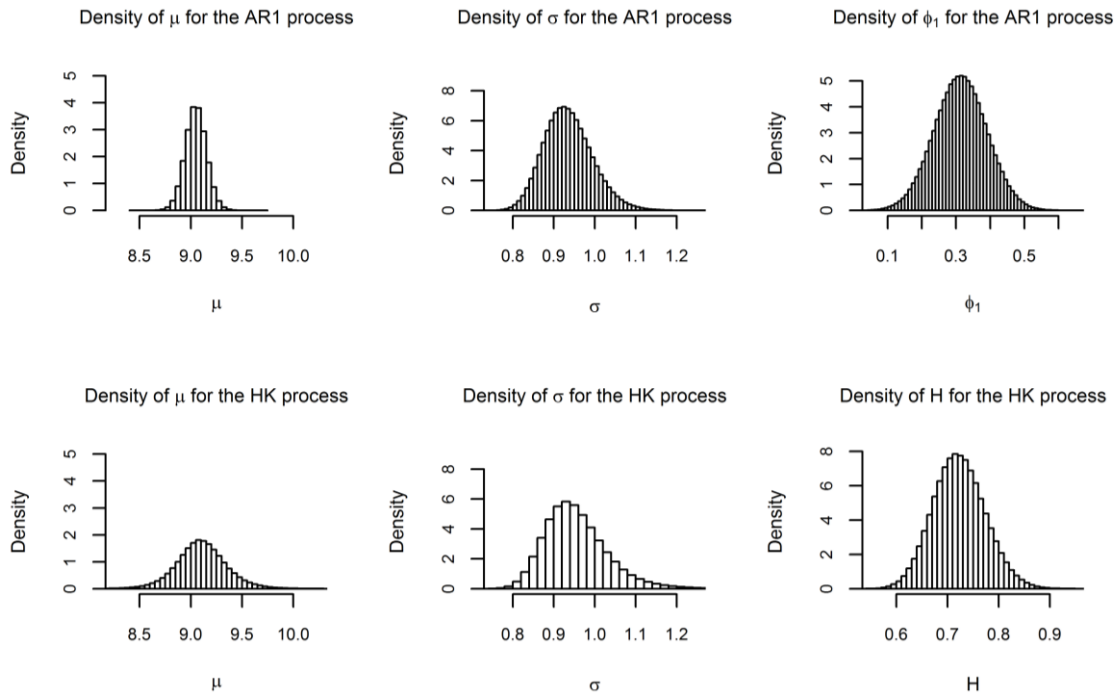


Figure 4.7. Posterior probability distributions of $\underline{\mu}$, $\underline{\sigma}$, $\underline{H}$, $\underline{\varphi}_1$ for the cases of AR(1) and HK processes, for the temperature at Berlin/Tempelhof. In this case the parameters are estimated from years 1756-1919.

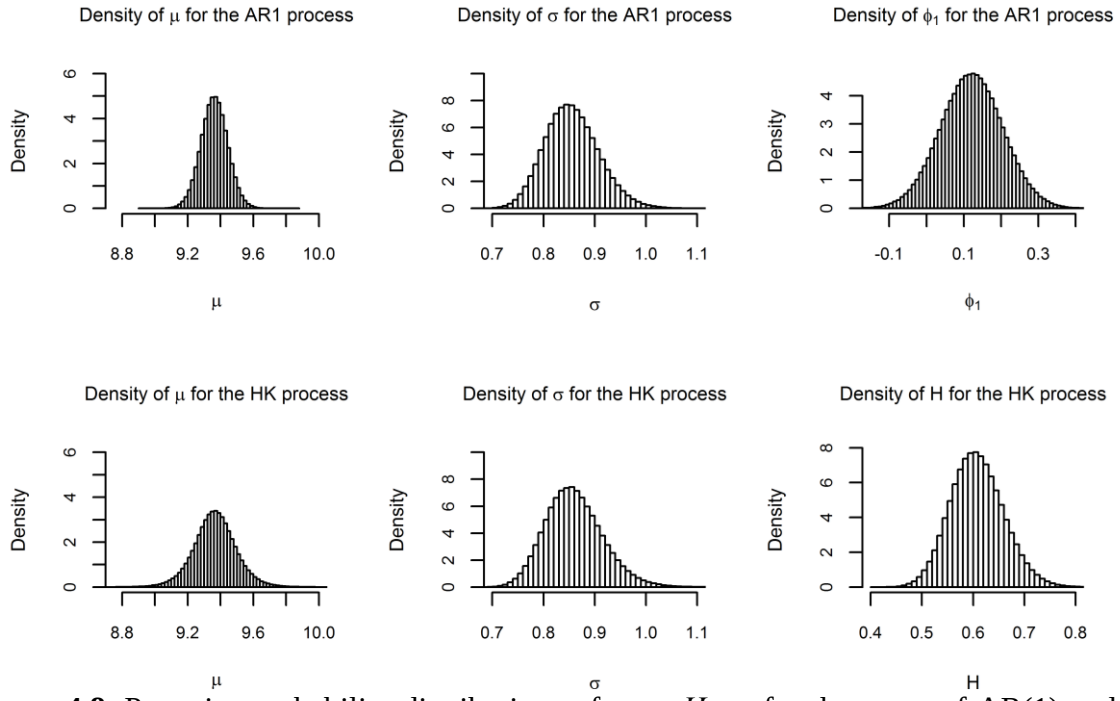


Figure 4.8. Posterior probability distributions of $\underline{\mu}$, $\underline{\sigma}$, $\underline{H}$, $\underline{\varphi}_1$ for the cases of AR(1) and HK processes, for the temperature at Vienna. In this case the parameters are estimated from years 1775-1919.

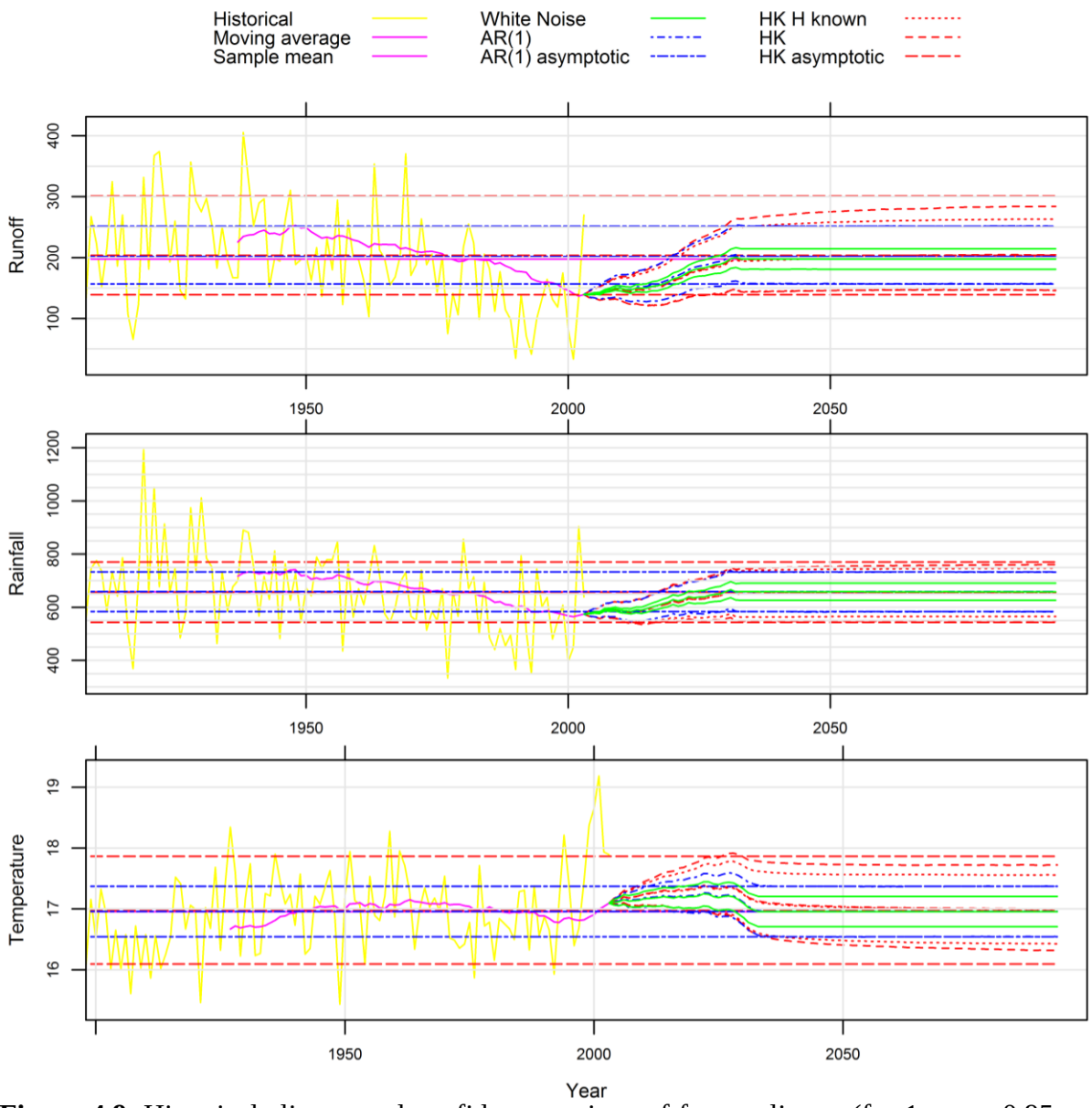


Figure 4.9. Historical climate and confidence regions of future climate (for $1 - a = 0.95$ and climatic time scale of 30 years) for (upper) runoff of Boeoticos Kephisos, (middle) rainfall at Aliartos, and (lower) temperature at Aliartos.

The second result of the framework is the predictive distribution of the future evolution of the process of interest. The posterior predictive 0.95-confidence regions for the 30-year moving averages are given in Figures 4.9-4.11. For C1 the confidence region is not symmetric with respect to the estimated mean, owing to the lower truncation bound alongside with the relatively big $\hat{\sigma}$. In contrast, there is a symmetry for C2 owing to the relatively small $\hat{\sigma}$, which justifies our decision to use models without truncation in those cases where $\hat{\sigma}$ is even smaller (compared to mean). For all cases, the widest confidence regions correspond to the HKp (due to the existence of persistence), followed by the AR(1), while the narrowest confidence regions appear for the WN. Of course the confidence regions for unknown H are wider than in

the case where *H* was considered to be known and equal to its maximum likelihood estimate. In C5 and C7 the HKp seems to be the best model, because it captures better than the others the observed values of the climate variable for the last 90 years based on the observed values of the previous years. In C7 it seems that the HKp did not capture the increase of temperature in last decades. But when we examine the full dataset (C5), the behaviour in last 90 years does not appear extraordinary. For the asymptotic values in the HKp, the 0.95-confidence region ranges at intervals of the order of 150 mm (C1), 220 mm (C2), 1.6°C (C3), 1.9°C (C4), 1.4°C (C5) for the 30-year moving average. The corresponding values for the case of the WN of the order of 50 mm (C1), 75 mm (C2), 0.5°C (C3), 0.6°C (C4), 0.6°C (C5) are considerably smaller compared to the case of the HKp.

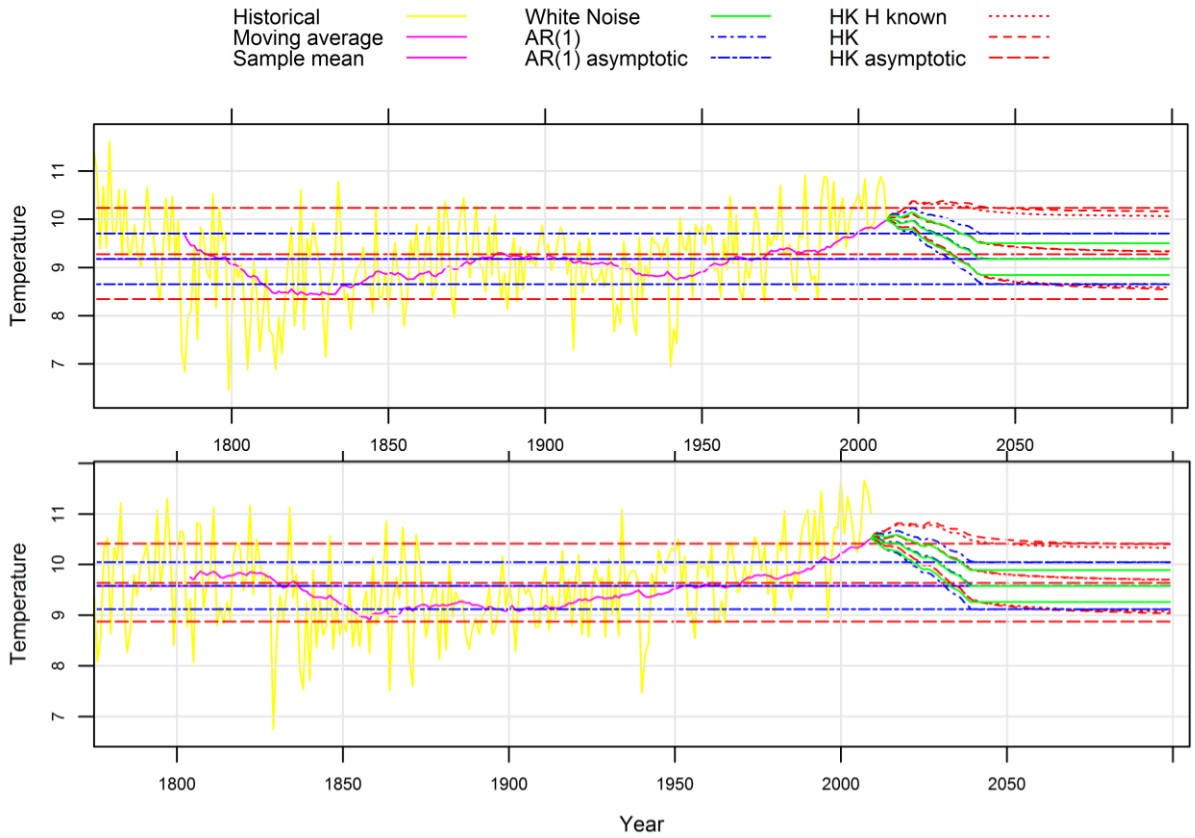


Figure 4.10. Historical climate and confidence regions of future climate (for $1 - a = 0.95$ and climatic time scale of 30 years) for (upper) temperature at Berlin, and (lower) temperature at Vienna.

4.6 Summary

We developed a Bayesian statistical methodology to make hydroclimatic prognosis in terms of estimating future confidence regions on the basis of a stationary normal stochastic process. We applied this methodology to five cases, namely the runoff (C1), the rainfall (C2) and the temperature (C3) at Boeoticos Kephisos river basin in Greece, as well as the temperature at

Berlin (C4, C6) and the temperature at Vienna (C5, C7). The Bayesian statistical model consisted of a stationary normal process (or truncated stationary normal process for the runoff and rainfall cases) with a noninformative prior distribution. Three kinds of stationary normal processes were examined, namely WN, AR(1) and HKp. We derived the posterior distributions of the parameters of the models, the posterior predictive distributions of the variables of the process and the posterior predictive distribution of the 30-year moving average which was the climatic variable of interest. The methodology can also be applied to other structures of the ACF.

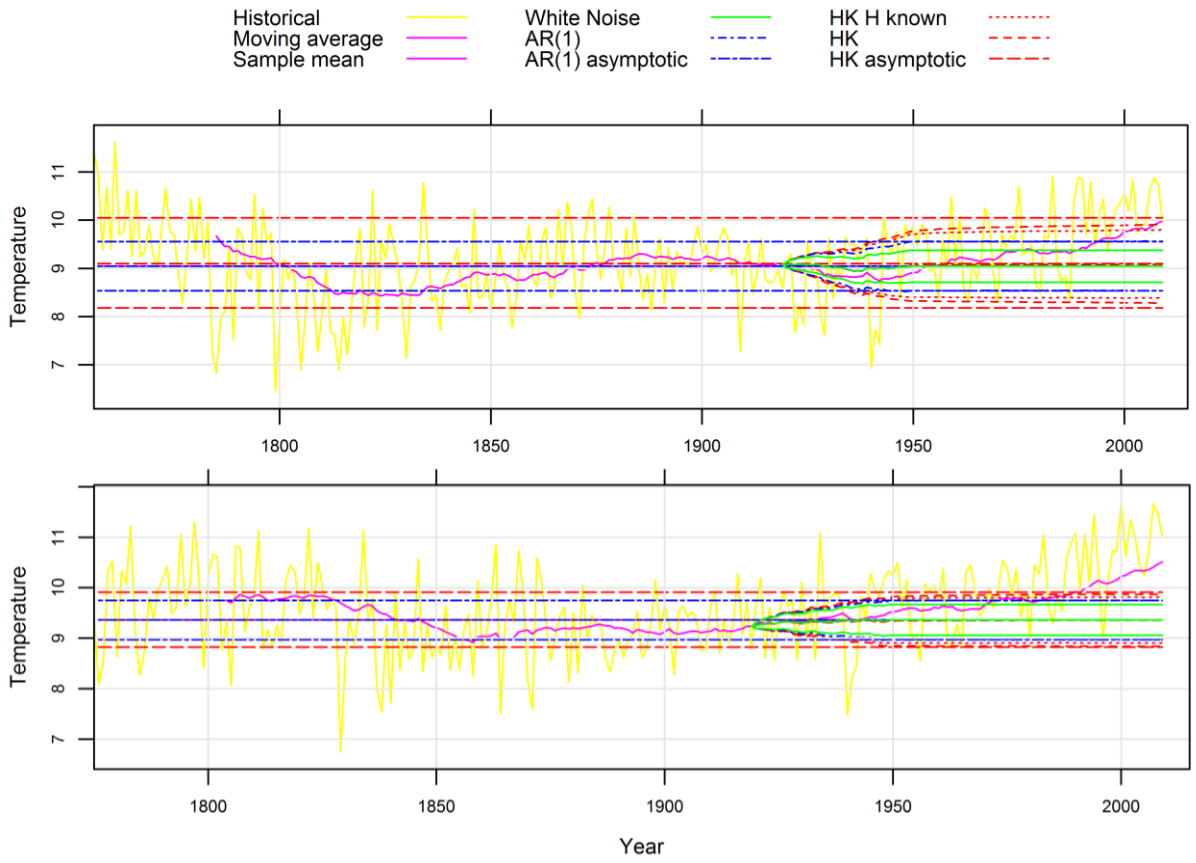


Figure 4.11. Historical climate and confidence regions of climate (for $1 - a = 0.95$ and climatic time scale of 30 years) for (upper) temperature at Berlin/Tempelhof after the year 1920 and (lower) temperature at Vienna after the year 1920.

A first important conclusion is that for all the examined cases and for all the examined processes their estimated means are almost equal as expected. However the posterior distributions of the means are wider when using the HKp, due to the persistence of the process, and even wider when all parameters of the process are assumed to be unknown. This results in wider confidence regions for future climatic variables of the processes. Moreover the confidence regions of truncated future variables are asymmetric. This asymmetry depends on the variance of the examined process. However the posterior distributions of the means of

all processes were less asymmetric.

Table 4.8. Summary results for the parameters of the AR(1) and HK cases at Boeoticos Kephisos River basin.

			Quantiles				
Case	Mean	Standard Deviation	2.5%	25%	50%	75%	97.5%
Boeoticos runoff							
AR(1)							
μ	197.7	12.69	172.5	189.4	197.7	205.9	222.8
σ	83.93	7.41	71.50	78.78	83.23	88.29	100.45
φ_1	0.35	0.10	0.16	0.28	0.35	0.42	0.55
HK							
μ	194.85	31.30	132	178.1	195	211.6	256.1
σ	86.51	12.35	71.19	79.15	84.40	91.06	114.22
H	0.74	0.07	0.62	0.69	0.74	0.78	0.88
Aliartos rainfall							
AR(1)							
μ	658.18	18.57	621.5	646	658.2	670.4	694.7
σ	159.9	12.24	138.3	151.3	159.1	167.5	186.2
φ_1	0.11	0.10	−0.09	0.04	0.11	0.18	0.32
HK							
μ	657.09	31.98	592.5	638.4	657.3	676.1	720.4
σ	160.7	13.45	137.9	151.4	159.5	168.6	190.3
H	0.62	0.06	0.51	0.58	0.62	0.66	0.75
Aliartos temperature							
AR(1)							
μ	16.96	0.10	16.76	16.89	16.96	17.02	17.15
σ	0.71	0.06	0.61	0.67	0.70	0.74	0.84
φ_1	0.33	0.10	0.14	0.26	0.33	0.39	0.52
HK							
μ	16.97	0.29	16.44	16.83	16.97	17.11	17.52
σ	0.75	0.13	0.62	0.68	0.73	0.79	0.99
H	0.74	0.07	0.61	0.69	0.74	0.79	0.88

Table 4.9. Summary results for the parameters of the AR(1) and HK cases respectively at Berlin and Vienna.

Case	Mean	Standard Deviation	Quantiles 2.5%	25%	50%	75%	97.5%
Berlin temperature (1756-2009)							
AR(1)							
$\underline{\mu}$	9.18	0.09	9.01	9.12	9.18	9.24	9.35
$\underline{\sigma}$	0.93	0.05	0.84	0.89	0.92	0.96	1.03
φ_1	0.38	0.06	0.26	0.34	0.38	0.42	0.49
HK							
$\underline{\mu}$	9.28	0.25	8.80	9.13	9.27	9.43	9.79
$\underline{\sigma}$	0.94	0.06	0.83	0.89	0.93	0.97	1.08
$\underline{H}$	0.75	0.03	0.67	0.72	0.75	0.77	0.83
Vienna temperature (1775-2009)							
AR(1)							
$\underline{\mu}$	9.58	0.08	9.42	9.53	9.58	9.63	9.74
$\underline{\sigma}$	0.88	0.05	0.80	0.85	0.88	0.91	0.98
φ_1	0.31	0.06	0.19	0.27	0.31	0.35	0.43
HK							
$\underline{\mu}$	9.64	0.19	9.27	9.52	9.64	9.76	10.03
$\underline{\sigma}$	0.88	0.05	0.79	0.84	0.87	0.91	0.99
$\underline{H}$	0.71	0.04	0.64	0.68	0.71	0.73	0.79
Berlin temperature (1756-1919)							
AR(1)							
$\underline{\mu}$	9.05	0.10	8.85	8.98	9.05	9.12	9.25
$\underline{\sigma}$	0.94	0.06	0.83	0.89	0.93	0.97	1.06
φ_1	0.31	0.08	0.16	0.26	0.31	0.37	0.46
HK							
$\underline{\mu}$	9.11	0.26	8.60	8.95	9.10	9.26	9.64
$\underline{\sigma}$	0.96	0.08	0.83	0.90	0.95	1.00	1.14
$\underline{H}$	0.72	0.05	0.63	0.69	0.72	0.76	0.83
Vienna temperature (1775-1919)							
AR(1)							
$\underline{\mu}$	9.36	0.08	9.20	9.31	9.36	9.42	9.52
$\underline{\sigma}$	0.86	0.05	0.76	0.82	0.85	0.89	0.97
φ_1	0.12	0.08	−0.04	0.06	0.12	0.18	0.29
HK							
$\underline{\mu}$	9.37	0.13	9.10	9.29	9.37	9.45	9.63
$\underline{\sigma}$	0.86	0.06	0.76	0.82	0.86	0.89	0.98
$\underline{H}$	0.61	0.05	0.51	0.57	0.61	0.64	0.72

Table 4.10. Estimates of μ using various methods.

	Maximum likelihood estimate			50% quantile	
Examined case	WN	AR(1)	HKp	AR(1)	HK
Boeoticos runoff	197.63	197.65	195.11	197.7	195
Aliartos rainfall	658.36	658.22	657.38	658.2	657.3
Aliartos temperature	16.96	16.96	16.97	16.96	16.97
Berlin temperature (1756-2009)	9.17	9.18	9.27	9.18	9.28
Vienna temperature (1775-2009)	9.58	9.58	9.64	9.58	9.64
Berlin temperature (1756-1919)	9.04	9.05	9.10	9.05	9.11
Vienna temperature (1775-1919)	9.36	9.36	9.37	9.36	9.37

Another important conclusion is that the use of short-range dependence stochastic processes is not suitable to model geophysical processes, because they underestimate uncertainty. However stationary persistent stochastic processes are suitable to achieve this purpose. In the examined cases they performed well and were able to explain the fluctuations

of the process.

One may claim that, when climate is to be predicted, an assumption of stationarity is not an appropriate one as currently several climate models project a changing future climate. Nonetheless, an assessment of future climate variability and uncertainty based on the stationarity hypothesis is a necessary step in establishing a stochastic method, whose generalization at a second step would enable incorporating nonstationary components. In addition, without knowing the variability under stationary conditions, it would not be possible to quantify the credibility of climate models and even their usefulness. Work on the generalization of the methodology to incorporate deterministic predictions by climate models is presented in Chapter 5.

5. On the prediction of persistent processes using the output of deterministic models

A problem frequently met[1] in the engineering hydrology community is the prediction of future hydrologic variables conditional on their historical observations and the hindcasts and forecasts of a deterministic model. Various methods have been developed to deal with this task under the independence or the Markovian dependence assumption of the variables. On the other hand it is a common practice for climatologists to use the output of GCMs for the prediction of climatic variables despite their imperfections and their inability to quantify the uncertainty of the predictions. In this Chapter we extend the aforementioned hydrological frameworks to include cases where persistent dependence appears. The framework is applied to climate time series and the output of GCMs. Predictions of the climate variables are derived, with their uncertainty. We conclude that the influence of the GCMs to the reduction of the uncertainty is negligible.

5.1 Introduction

Recently various studies regarding the prediction of future hydrologic variables based on stochastic models have been carried out. To mention some of them, Koutsoyiannis et al. (2008b) proposed a stochastic model for the prediction of the Nile flow a month ahead. On larger time scales Koutsoyiannis et al. (2007) proposed a stochastic framework to calculate future climatic uncertainties conditional on historic observations, while the same problem was tackled in a Bayesian framework in Chapter 4. Stochastic models are frequently used also by engineering hydrologists for the prediction of hydrologic variables, whereas the climatologists focus on deterministic models (GCMs) (Koutsoyiannis et al. 2007). While it is true that deterministic models incorporate knowledge of the climatic mechanisms expressed through deterministic equations, they are not appropriate to quantify the uncertainty of prediction.

The task of exploiting the output of deterministic models to improve the output of stochastic models has been studied as well by hydrologists, e.g. Montanari and Grossi (2008), Zhao et al. (2011), Smith et al. (2012) and others. Krzysztofowicz (1987a,b; 1999a,b; 2001; 2002), Krzysztofowicz and Maranzano (2004), and Krzysztofowicz and Evans (2008) proposed a stochastic framework, namely the Bayesian Forecasting System (BFS) for producing a probabilistic forecast of a hydrologic predictand via any deterministic catchment

[1] Based on: Tyralis and Koutsoyiannis (2015)

model. Wang et al. (2009) and Pokhrel et al. (2013) proposed a Bayesian joint probability (BJP) modelling approach for seasonal forecasting of streamflows at multiple sites. The BFS and the BJP can be applied to any hydrological processes, irrespective of their autocorrelation structure. However to the authors' knowledge they have been applied only to white noise or Markovian stochastic processes.

Koutsoyiannis (2000) examined the handling of HK behaviour in multivariate stochastic simulation of hydrological processes. A multivariate extension of the HKp was proposed by Lavancier et al. (2009) and its properties were studied extensively by Amblard and Coeurjolly (2011), Amblard et al. (2012), Coeurjolly et al. (2010; 2013).

In this Chapter we modify the BJP proposed by Wang et al. (2009) to make prediction of hydrologic processes exhibiting HK behaviour conditional on their historical observations and the output of some deterministic models. The output's time period spans from the historical observations period to future projection in an arbitrary time. To this end we model the two time series (the observed data and the output of the deterministic model) using a well-balanced bivariate HKp (see definition below). A maximum likelihood estimator of the parameters of the model is proposed and the estimated values of the parameters are used to make inference for the distribution of the processes under study. In the proposed framework, the knowledge of the exact dynamics of the deterministic model is not a requirement, similarly to the BFS. However the structure of the proposed approach differs, in that the BFS relies on an assumption of conditional independence between the variables of the stochastic process and the deterministic model. Hence the distribution of the deterministic model is determined from the stochastic process. In our model the distribution of the variables of the deterministic model is considered known and the correlation between the variables is examined.

The framework is applied to global averaged temperature and precipitation datasets, which are assumed to exhibit HK behaviour. The deterministic models are GCMs. It is shown that the information added by the deterministic models is negligible, particularly for precipitation. This was expected according to Koutsoyiannis et al. (2008a) and Anagnostopoulos et al. (2010) who compared the output of various GCMs to temperature and precipitation observations and showed that the spatially integrated projections were poor.

5.1.1 Definition of the well-balanced bivariate HKp

We assume that $\{\underline{x}_{1t}\}$ and $\{\underline{x}_{2t}\}$, $t = 1, 2, \ldots$ are two HKp's with parameters (μ_1, σ_1, H_1) and (μ_2, σ_2, H_2) respectively. Then the normal bivariate process $\{\underline{\boldsymbol{x}}_t = (\underline{x}_{1t},\underline{x}_{2t})\}$, $t = 1, 2, \ldots$ is a well-balanced (i.e. a time-reversible) HKp if (Amblard et al. 2012)

$$w_{ij}(k) := \rho_{i,j}\,|k|^{H_i+H_j},\ \rho_{i,i} = 1,\ \rho_{i,j} = \rho_{j,i} = \rho,\ \{i,j\} \in \{\{1,2\},\{1,2\}\} \tag{5.1}$$

$$\gamma_{ij}(k) := \mathrm{Cov}[\underline{x}_{it},\ \underline{x}_{j\,t+k}] = (1/2)\,\sigma_i\,\sigma_j\ \left(\, w_{ij}(k-1) - 2\,w_{ij}(k) + w_{ij}(k+1)\,\right) \tag{5.2}$$

under the following restriction

$$\rho^2 \le \frac{\Gamma(2H_1+1)\,\Gamma(2H_2+1)\,\sin(\pi H_1)\,\sin(\pi H_2)}{\Gamma^2(H_1+H_2+1)\,\sin^2(\pi(H_1+H_2)/2)} \tag{5.3}$$

Note that for $i = j$, (5.2) is equivalent to (1.9).

5.2 Maximum likelihood estimator for the parameters of the bivariate HKp

The problem of finding and assessing the maximum likelihood estimator for the parameters of the HKp was studied in Chapter 3. The solution of this problem for the bivariate HKp is more complicated. We assume that there is a record of n observations $\boldsymbol{x}_{1\,1:n} := (x_{11},\ldots,x_{1n})^{\mathrm{T}}$ and $\boldsymbol{x}_{2\,1:n} := (x_{21},\ldots,x_{2n})^{\mathrm{T}}$. The parameters of the bivariate HKp are $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, H_1, H_2, \rho)$. We use the terminology of Wei (2006 p.382-427). Hence we have the mean vector

$$\mathrm{E}[\underline{\boldsymbol{x}}_t] = (\mu_1\ ,\ \mu_2)^{\mathrm{T}} \tag{5.4}$$

and the lag-k covariance matrix $\boldsymbol{\Gamma}(k)$, which as a function of k is called the covariance matrix function for the process $\underline{\boldsymbol{x}}_t$.

$$\boldsymbol{\Gamma}(k) := \mathrm{Cov}[\underline{\boldsymbol{x}}_t,\ \underline{\boldsymbol{x}}_{t+k}] = \begin{bmatrix} \gamma_{11}(k) & \gamma_{21}(k) \\ \gamma_{21}(k) & \gamma_{22}(k) \end{bmatrix} \tag{5.5}$$

The covariance matrix of the multivariate normal variable $\underline{\boldsymbol{x}}_{1:n} := (\underline{\boldsymbol{x}}_1^{\mathrm{T}},\underline{\boldsymbol{x}}_2^{\mathrm{T}},\ldots,\underline{\boldsymbol{x}}_n^{\mathrm{T}})^{\mathrm{T}}$ is

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}(0) & \boldsymbol{\Gamma}(1) & \ldots & \boldsymbol{\Gamma}(n-1) \\ \boldsymbol{\Gamma}(1) & \boldsymbol{\Gamma}(0) & \ldots & \boldsymbol{\Gamma}(n-2) \\ \ldots & \ldots & \ldots & \ldots \\ \boldsymbol{\Gamma}(n-1) & \boldsymbol{\Gamma}(n-2) & \ldots & \boldsymbol{\Gamma}(0) \end{bmatrix} \tag{5.6}$$

Rearranging the elements of $\underline{\boldsymbol{x}}_{1:n}$ we define the vector $\underline{\boldsymbol{w}}_{1:n} := (\underline{\boldsymbol{x}}_{1\,1:n}^{\mathrm{T}}\ ,\ \underline{\boldsymbol{x}}_{2\,1:n}^{\mathrm{T}})^{\mathrm{T}}$ with covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix} \tag{5.7}$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are the covariance matrices of $\boldsymbol{x}_{1\ 1:n}$ and $\boldsymbol{x}_{2\ 1:n}$ and $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21}$ are their cross-covariance matrices.

$$\boldsymbol{\Sigma}_1 := \sigma_1^2\, \boldsymbol{R}_1,\ \boldsymbol{R}_1(i,j) = \boldsymbol{R}_1(j,i) := \rho_{1(j-i)} \text{ and } \boldsymbol{\Sigma}_2 := \sigma_2^2\, \boldsymbol{R}_2,\ \boldsymbol{R}_2(i,j) = \boldsymbol{R}_2(j,i) := \rho_{2(j-i)} \tag{5.8}$$

$$\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12} := \rho\ \sigma_1\ \sigma_2\ \boldsymbol{R}_{21},\ \boldsymbol{R}_{21}(i,j) = \boldsymbol{R}_{21}(j,i) = \boldsymbol{R}_{21}(j-i) := \rho_{21}(j-i) \tag{5.9}$$

$$\rho_{21}(j-i) := \gamma_{21}(j-i) / (\rho\ \sigma_1\ \sigma_2) = (1/2)\ (\ |j-i-1|^{H_1+H_2} - 2\ |j-i\ |^{H_1+H_2} + |j-i+1|^{H_1+H_2}\ ) \tag{5.10}$$

The $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, $\boldsymbol{R}_{21}$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_{21}$ are symmetric Toeplitz positive definite matrices (Golub and Van Loan 1996 p.193). The Schur complements (Horn and Zhang 2005 p.18) of the matrices $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_1$ are

$$\boldsymbol{S}_1 = \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{21}\, \boldsymbol{\Sigma}_2^{-1}\, \boldsymbol{\Sigma}_{21} = \sigma_1^2\, (\boldsymbol{R}_1 - \rho^2\, \boldsymbol{R}_{21}\, \boldsymbol{R}_2^{-1}\, \boldsymbol{R}_{21}) \tag{5.11}$$

$$\boldsymbol{S}_2 = \boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_{21}\, \boldsymbol{\Sigma}_1^{-1}\, \boldsymbol{\Sigma}_{21} = \sigma_2^2\, (\boldsymbol{R}_2 - \rho^2\, \boldsymbol{R}_{21}\, \boldsymbol{R}_1^{-1}\, \boldsymbol{R}_{21}) \tag{5.12}$$

and they are symmetric as well. It is proved after substituting (5.8) and (5.9) in (5.13) that

$$-\boldsymbol{\Sigma}_1^{-1}\, \boldsymbol{\Sigma}_{21}\, \boldsymbol{S}_2^{-1} = -\frac{\rho}{\sigma_1\ \sigma_2}\, \boldsymbol{R}_1^{-1}\, \boldsymbol{R}_{21}\, (\boldsymbol{R}_2 - \rho^2\, \boldsymbol{R}_{21}\, \boldsymbol{R}_1^{-1}\, \boldsymbol{R}_{21})]^{-1} \tag{5.13}$$

Additionally

$$-\boldsymbol{S}_2^{-1}\, \boldsymbol{\Sigma}_{21}\, \boldsymbol{\Sigma}_1^{-1} = (-\boldsymbol{\Sigma}_1^{-1}\, \boldsymbol{\Sigma}_{21}\, \boldsymbol{S}_2^{-1})^{\mathrm{T}} \tag{5.14}$$

because $\boldsymbol{S}_2$, $\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_1$ are symmetric matrices, hence the inverse of $\boldsymbol{\Sigma}$ is (Horn and Zhang 2005 p.19)

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{S}_1^{-1} & -\boldsymbol{\Sigma}_1^{-1}\, \boldsymbol{\Sigma}_{21}\, \boldsymbol{S}_2^{-1} \\ (-\boldsymbol{\Sigma}_1^{-1}\, \boldsymbol{\Sigma}_{21}\, \boldsymbol{S}_2^{-1})^{\mathrm{T}} & \boldsymbol{S}_2^{-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}_1^{-1} & -\boldsymbol{S}_1^{-1}\, \boldsymbol{\Sigma}_{21}\, \boldsymbol{\Sigma}_2^{-1} \\ (-\boldsymbol{S}_1^{-1}\, \boldsymbol{\Sigma}_{21}\, \boldsymbol{\Sigma}_2^{-1})^{\mathrm{T}} & \boldsymbol{S}_2^{-1} \end{bmatrix} \tag{5.15}$$

Now we define the vectors

$$\boldsymbol{e}_n = (1,1,\ldots,1)^{\mathrm{T}} \tag{5.16}$$

$$\boldsymbol{\mu} = (\mu_1 \boldsymbol{e}_n^{\mathrm{T}}\ ,\ \mu_2 \boldsymbol{e}_n^{\mathrm{T}}\ )^{\mathrm{T}} \tag{5.17}$$

The probability distribution function of $\underline{\boldsymbol{w}}_{1:n}$ is (Eaton 2007 p.122)

$$f(\boldsymbol{w}_{1:n} | \boldsymbol{\mu}\ ,\ \boldsymbol{\Sigma}) = (2\pi)^{-n}\ |\boldsymbol{\Sigma}|^{-1/2} \exp(-(1/2)\ (\boldsymbol{w}_{1:n} - \boldsymbol{\mu})^{\mathrm{T}}\ \boldsymbol{\Sigma}^{-1}\ (\boldsymbol{w}_{1:n} - \boldsymbol{\mu})) \tag{5.18}$$

The maximum likelihood estimates $\hat{\mu}_1$ and $\hat{\mu}_2$ are given in Section 5.3 and depend on the other parameters of the bivariate HKp. However when substituting them in (5.18) its maximization becomes complicated. From now on we assume that μ_1, μ_2 are known or estimated from the corresponding sample means, e.g. see the estimation techniques proposed

by Amblard and Coeurjolly (2011). Substituting $\boldsymbol{\Sigma}^{-1}$ from (5.15) in (5.18) and taking the partial derivatives of the log-likelihood function (5.18) with respect to σ_1 and σ_2 we obtain

$$\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_1} = \begin{bmatrix} -\frac{2}{\sigma_1^3}(\boldsymbol{R}_1 - \rho^2 \boldsymbol{R}_{21}\boldsymbol{R}_2^{-1}\boldsymbol{R}_{21})^{-1} & \frac{\rho_{12}}{\sigma_1^2\sigma_2}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21}(\boldsymbol{R}_2 - \rho^2 \boldsymbol{R}_{21}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21})]^{-1} \\ \frac{\rho_{12}}{\sigma_1^2\sigma_2}(\boldsymbol{R}_2 - \rho^2 \boldsymbol{R}_{21}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21})]^{-1}\boldsymbol{R}_{21}\boldsymbol{R}_1^{-1} & \boldsymbol{0} \end{bmatrix} \quad (5.19)$$

$$\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_2} = \begin{bmatrix} \boldsymbol{0} & \frac{\rho_{12}}{\sigma_1\sigma_2^2}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21}(\boldsymbol{R}_2 - \rho^2 \boldsymbol{R}_{21}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21})]^{-1} \\ \frac{\rho_{12}}{\sigma_1\sigma_2^2}(\boldsymbol{R}_2 - \rho^2 \boldsymbol{R}_{21}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21})]^{-1}\boldsymbol{R}_{21}\boldsymbol{R}_1^{-1} & -\frac{2}{\sigma_2^3}(\boldsymbol{R}_2 - \rho^2 \boldsymbol{R}_{21}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21})^{-1} \end{bmatrix} \quad (5.20)$$

The determinant of $\boldsymbol{\Sigma}$ is (Horn and Zhang 2005 p.19)

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_2|\,|\boldsymbol{S}_1| = |\boldsymbol{\Sigma}_1|\,|\boldsymbol{S}_2| = |\sigma_2^2\boldsymbol{R}_2|\,|\sigma_1^2(\boldsymbol{R}_1 - \rho^2\boldsymbol{R}_{21}\boldsymbol{R}_2^{-1}\boldsymbol{R}_{21})| = |\sigma_1^2\boldsymbol{R}_1|\,|\sigma_2^2(\boldsymbol{R}_2 - \rho^2\boldsymbol{R}_{21}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21})| \quad (5.21)$$

and

$$\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \sigma_1} = \frac{2n}{\sigma_1}, \frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \sigma_2} = \frac{2n}{\sigma_2} \quad (5.22)$$

Solving the system

$$\frac{\partial \ln(f(\boldsymbol{w}_{1:n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}))}{\partial \sigma_1} = 0 \text{ and } \frac{\partial \ln(f(\boldsymbol{w}_{1:n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}))}{\partial \sigma_2} = 0 \quad (5.23)$$

for σ_1 and σ_2 we obtain

$$\hat{\sigma}_1 = ((a_1 a_3^{1/2} - \rho\, a_2 a_1^{1/2})/(\mathrm{n}\, a_3^{1/2}))^{1/2}, \hat{\sigma}_2 = ((a_3 a_1^{1/2} - \rho\, a_2 a_3^{1/2})/(n\, a_1^{1/2}))^{1/2} \quad (5.24)$$

where

$$a_1 := \boldsymbol{y}_{1\,1:n}^{\mathrm{T}}(\boldsymbol{R}_1 - \rho^2 \boldsymbol{R}_{21}\boldsymbol{R}_2^{-1}\boldsymbol{R}_{21})^{-1}\boldsymbol{y}_{1\,1:n},$$

$$a_2 := \boldsymbol{y}_{2\,1:n}^{\mathrm{T}}(\boldsymbol{R}_2 - \rho^2 \boldsymbol{R}_{21}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21})^{-1}\boldsymbol{R}_{21}\boldsymbol{R}_1^{-1}\boldsymbol{y}_{1\,1:n}, a_3 := \boldsymbol{y}_{2\,1:n}^{\mathrm{T}}(\boldsymbol{R}_2 - \rho^2 \boldsymbol{R}_{21}\boldsymbol{R}_1^{-1}\boldsymbol{R}_{21})^{-1}\boldsymbol{y}_{2\,1:n} \quad (5.25)$$

and

$$\boldsymbol{y}_{1\,1:n} = (\boldsymbol{x}_{1\,1:n}^{\mathrm{T}} - \mu_1\boldsymbol{e}_n^{\mathrm{T}})^{\mathrm{T}}, \boldsymbol{y}_{2\,1:n} = (\boldsymbol{x}_{2\,1:n}^{\mathrm{T}} - \mu_2\boldsymbol{e}_n^{\mathrm{T}})^{\mathrm{T}} \quad (5.26)$$

Now substituting (5.24) in (5.18) and maximizing the log-likelihood of the three parameters we obtain $\hat{H}_1, \hat{H}_2, \hat{\rho}$. After substituting these values in (5.24) we obtain $\hat{\sigma}_1$ and $\hat{\sigma}_2$.

We assume now that there is a record of observations $\boldsymbol{x}_{1\,1:(n+m)} := (x_{11},\ldots,x_{1\,(n+m)})^{\mathrm{T}}$ and $\boldsymbol{x}_{2\,1:n} := (x_{21},\ldots,x_{2n})^{\mathrm{T}}$. Following the same procedure it is shown that $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^{\mathrm{T}}$ and

$$a_1 := \boldsymbol{y}^{\mathrm{T}}_{1\,1:(n+m)} \left(\boldsymbol{R}_1 - \rho^2\, \boldsymbol{R}^{\mathrm{T}}_{21}\, \boldsymbol{R}^{-1}_2\, \boldsymbol{R}_{21}\right)^{-1} \boldsymbol{y}_{1\,1:(n+m)},$$

$$a_2 := \boldsymbol{y}^{\mathrm{T}}_{2\,1:n} \left(\boldsymbol{R}_2 - \rho^2 \boldsymbol{R}_{21} \boldsymbol{R}^{-1}_1 \boldsymbol{R}^{\mathrm{T}}_{21}\right)^{-1} \boldsymbol{R}_{21} \boldsymbol{R}^{-1}_1 \boldsymbol{y}_{1\,1:(n+m)},\ a_3 := \boldsymbol{y}^{\mathrm{T}}_{2\,1:n} \left(\boldsymbol{R}_2 - \rho^2 \boldsymbol{R}_{21} \boldsymbol{R}^{-1}_1 \boldsymbol{R}^{\mathrm{T}}_{21}\right)^{-1} \boldsymbol{y}_{2\,1:n} \tag{5.27}$$

5.3 Maximum likelihood estimators of the means of the bivariate HKp

We mentioned in Section 5.2 that the maximum likelihood estimates $\hat{\mu}_1$ and $\hat{\mu}_2$ depend on the other parameters of the bivariate HKp. To obtain them we substitute (5.15) in (5.28):

$$(\boldsymbol{w}_{1:n} - \boldsymbol{\mu})^{\mathrm{T}}\, \boldsymbol{\Sigma}^{-1}\, (\boldsymbol{w}_{1:n} - \boldsymbol{\mu}) = (1/2)\, (\mu_1\,,\, \mu_2\,)\, \boldsymbol{A} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} - (\mu_1\,,\, \mu_2\,)\, \boldsymbol{b} + \boldsymbol{w}^{\mathrm{T}}_{1:n}\, \boldsymbol{\Sigma}^{-1}\, \boldsymbol{w}_{1:n} \tag{5.28}$$

where

$$\boldsymbol{A} = 2 \begin{bmatrix} \boldsymbol{e}^{\mathrm{T}}_n\, \boldsymbol{S}^{-1}_1\, \boldsymbol{e}_n & \boldsymbol{e}^{\mathrm{T}}_n \left(-\boldsymbol{\Sigma}^{-1}_1\, \boldsymbol{\Sigma}_{21}\, \boldsymbol{S}^{-1}_2\right) \boldsymbol{e}_n \\ \boldsymbol{e}^{\mathrm{T}}_n \left(-\boldsymbol{\Sigma}^{-1}_1\, \boldsymbol{\Sigma}_{21}\, \boldsymbol{S}^{-1}_2\right) \boldsymbol{e}_n & \boldsymbol{e}^{\mathrm{T}}_n\, \boldsymbol{S}^{-1}_2\, \boldsymbol{e}_n \end{bmatrix},\ \boldsymbol{b} = 2 \begin{bmatrix} \boldsymbol{w}^{\mathrm{T}}_{1:n}\, \boldsymbol{\Sigma}^{-1}\, \dfrac{\partial \boldsymbol{\mu}}{\partial \mu_1} \\ \boldsymbol{w}^{\mathrm{T}}_{1:n}\, \boldsymbol{\Sigma}^{-1}\, \dfrac{\partial \boldsymbol{\mu}}{\partial \mu_2} \end{bmatrix} \tag{5.29}$$

To maximize (5.18), (5.28) should be minimized. Its minimum is attained for (Golub and Van Loan 1996 p.490)

$$(\mu_1\,,\, \mu_2\,)^{\mathrm{T}} = \boldsymbol{A}^{-1} \boldsymbol{b} \tag{5.30}$$

The matrices $\boldsymbol{A}$ and $\boldsymbol{b}$ are functions of the other parameters of the bivariate HKp, therefore after substituting their maximum likelihood estimates in (5.30) we obtain $\hat{\mu}_1$ and $\hat{\mu}_2$. However, as mentioned in Section 5.2, this result will not be used in this study.

5.4 Posterior predictive distributions

We assume that $\boldsymbol{x}_{1\,1:(n+k)}$ is the output of the deterministic model and $\boldsymbol{x}_{2\,1:n}$ is the data observed. We wish to find the distribution of $\boldsymbol{x}_{2\,(n+1):(n+m)}$ conditional on $\boldsymbol{x}_{1\,1:(n+m)}$ and $\boldsymbol{x}_{2\,1:n}$. Assuming that $\{\underline{\boldsymbol{x}}_t = (\underline{x}_{1t}, \underline{x}_{2t})\}$, $t = 1, 2, \ldots$ is a bivariate HKp, the probability distribution of $\underline{\boldsymbol{w}}_{1:\,(n+m)}$ is given by (5.18). The $2(n+m)$-by-$2(n+m)$ covariance matrix of the process is given by (5.7) and is partitioned according to (5.31)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{121} & \boldsymbol{\Sigma}_{122} \\ \boldsymbol{\Sigma}_{211} & \boldsymbol{\Sigma}_{2n} & \boldsymbol{\Sigma}_{2nm} \\ \boldsymbol{\Sigma}_{212} & \boldsymbol{\Sigma}_{2mn} & \boldsymbol{\Sigma}_{2m} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}_1 & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_2 \end{bmatrix} \tag{5.31}$$

where $\boldsymbol{\Sigma}_{2m}$ is m-by-m matrix and

$$\boldsymbol{P}_1 = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{121} \\ \boldsymbol{\Sigma}_{211} & \boldsymbol{\Sigma}_{2n} \end{bmatrix},\ \boldsymbol{P}_{21} = \begin{bmatrix} \boldsymbol{\Sigma}_{212} & \boldsymbol{\Sigma}_{2mn} \end{bmatrix},\ \boldsymbol{P}_{12} = \begin{bmatrix} \boldsymbol{\Sigma}_{122} \\ \boldsymbol{\Sigma}_{2nm} \end{bmatrix},\ \boldsymbol{P}_2 = \boldsymbol{\Sigma}_{2m} \tag{5.32}$$

Then the posterior predictive distribution of $\boldsymbol{x}_{2\,(n+1):(n+m)}$ conditional on $\boldsymbol{x}_{1\,1:(n+m)}$, $\boldsymbol{x}_{2\,1:n}$ and $\boldsymbol{\theta}$ is

$$f(\boldsymbol{x}_{2\,(n+1):(n+m)}|\boldsymbol{x}_{1\,1:(n+m)},\boldsymbol{x}_{2\,1:n},\boldsymbol{\theta}) = (2\pi\sigma^2)^{-m/2}\,|\boldsymbol{R}_{m|n}|^{-1/2}\exp[(-1/2\sigma^2)\cdot$$

$$(\boldsymbol{x}_{2\,(n+1):(n+m)} - \boldsymbol{\mu}_{m|n})^{\mathrm{T}}\,\boldsymbol{R}_{m|n}^{-1}\,(\boldsymbol{x}_{2\,(n+1):(n+m)} - \boldsymbol{\mu}_{m|n})] \tag{5.33}$$

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

$$\boldsymbol{\mu}_{m|n} = \mu_2\boldsymbol{e}_m + \boldsymbol{P}_{21}\,\boldsymbol{P}_1^{-1}\,(\,(\boldsymbol{x}_{1\,1:(n+m)}^{\mathrm{T}}\,,\,\boldsymbol{x}_{2\,1:n}^{\mathrm{T}})^{\mathrm{T}} - (\mu_1\boldsymbol{e}_{n+m}^{\mathrm{T}}\;,\,\mu_2\boldsymbol{e}_n^{\mathrm{T}})^{\mathrm{T}}\,) \tag{5.34}$$

$$\boldsymbol{R}_{m|n} = \boldsymbol{P}_2 - \boldsymbol{P}_{21}\,\boldsymbol{P}_1^{-1}\,\boldsymbol{P}_{12} \tag{5.35}$$

Here we mention that in the following $\boldsymbol{\theta}$ will be considered known and equal to its maximum likelihood estimate. In a Bayesian setting we would assume that $\boldsymbol{\theta}$ is a random variable, however this is out of the scope of this study and will be examined in the future. In the Bayesian setting the uncertainty of the prediction would increase, e.g. see Chapter 4. The variables that will be examined in the following will be considered normal. For truncated normal variables the interested reader is referred to Horrace (2005) and Chapter 4. The examination of non-normal variables is out of the scope of this study as well.

5.4.1 Investigation for various values of $\boldsymbol{\theta}$

An investigation for various values of the parameters is performed here.

- For $\rho = 0$, $\underline{\boldsymbol{x}}_1$ and $\underline{\boldsymbol{x}}_2$ are uncorrelated, hence the knowledge added by $\boldsymbol{x}_{1\,1:(n+m)}$ is useless. In this case (5.33) reduces to

$$f(\boldsymbol{x}_{2\,(n+1):(n+m)}|\boldsymbol{x}_{1\,1:(n+m)},\boldsymbol{x}_{2\,1:n},\boldsymbol{\theta}) = f(\boldsymbol{x}_{2\,(n+1):(n+m)}|\boldsymbol{x}_{2\,1:n},\boldsymbol{\theta}) \tag{5.36}$$

which already has been examined in Chapter 4.

- For $H_1 = H_2 = 0.5$ (5.33) reduces to the case of the normal-linear processor examined by Krzysztofowicz (1999a) with the following equivalence between the parameters of normal-linear processor and our model.

$$M = \mu_2 \tag{5.37}$$

$$S = \sigma_2 \tag{5.38}$$

$$a = (\rho\,\sigma_1)/\sigma_2 \tag{5.39}$$

$$\sigma^2 = \sigma_1^2\,(1 - \rho^2) \tag{5.40}$$

$$b = (\sigma_2\,\mu_1 - \rho\,\sigma_1\,\mu_2)/\sigma_2 \tag{5.41}$$

5.5 Case study

We applied our methodology to global temperature data and precipitation data shown in Table 5.1. These data are modelled by a Hurst-Kolmogorov process (Koutsoyiannis and Montanari 2007). We used the 20C3M for the calibration of the model and the SRES scenarios A1B, B1, A2 of the IPCC Fourth Assessment Report (AR4) to improve the prediction of the stochastic model (Table 5.2). The AR4 output could be divided into two time periods. The first time period corresponds to the 20C3M scenario, which simulates the climate of the past, based on greenhouse gasses increasing as observed through the 20th century. The 20C3M scenario approximately covers a time period spanning from 1880 to 2000, albeit the exact time period depends on the developer of the model. A list of model developers is shown in Table 5.3. The second time period corresponds to the A1B, B1, A2 scenarios and simulates the future climate, based on hypotheses mentioned in Table 5.2. We preferred to use the AR4 because the intersection of its second time period with the time period corresponding to the observed histical data is almost 10 years, thus we can inspect the validity of our predictions. The specific GCMs that where used in the study are shown in Table 5.3. Tables B.1-B.4 show the maximum likelihood estimates of the bivariate HKp $\{\underline{\mathbf{x}}_t = (\underline{x}_{1t},\underline{x}_{2t})\}$, where $\{\underline{x}_{1t}\}$ is the process which models the GCM and $\{\underline{x}_{2t}\}$ is the process which models the observations. The time interval for the calibration spans from the maximum starting year of the corresponding 20C3M scenario and the observed data to the minimum of the corresponding 20C3M scenario and the observed data (e.g. see Figure 5.1). We also examined the case where the parameters are estimated separately. Specifically the $\{\underline{x}_{1t}\}$, $\{\underline{x}_{2t}\}$ are assumed to be univariate HKps and their parameters are estimated as in Chapter 3. The sample cross-correlation function is used in this case to estimate ρ.

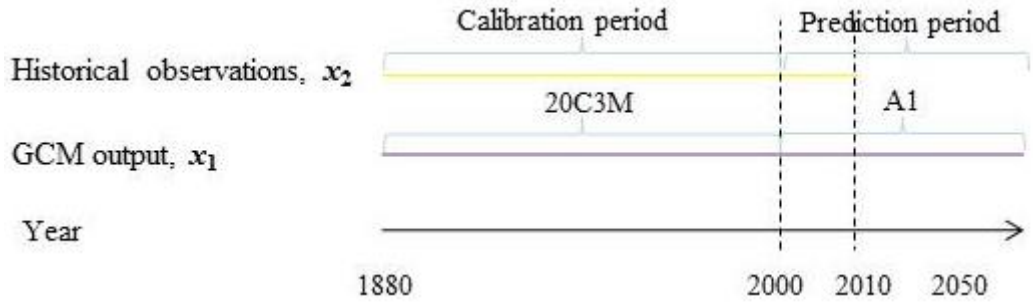


Figure 5.1. Scetch explaining the time periods that are used for model calibration, i.e. estimation of its parameters, and prediction. The specific years depicted in the sketch represent the typical years that were used in case studies (although these may vary in some of them; see Appendix B.

Table 5.1. Study historical time series.

Data	Name	Developed by	Time interval
Temperature	Global Land-Ocean Temperature Index	GISS	1880-2012
Temperature	Annual Global Land and Ocean Temperature Anomalies	NOAA	1880-2012
Temperature	Combined land [CRUTEM4] and marine temperature anomalies	CRU	1850-2012
Precipitation	Precipitation over land areas	CRU	1900-1998

Sources: data.giss.nasa.gov; www.nodc.noaa.gov/General/temperature.html; www.cru.uea.ac.uk/cru/data/temperature/; www.climatedata.info/Precipitation/Precipitation/global.html

Using the simultaneous maximum likelihood estimate of ρ, we obtain the posterior predictive distribution of $\boldsymbol{x}_{2\ (n+1):(n+m)}$ conditional on $\boldsymbol{x}_{1\ 1:(n+m)}$, $\boldsymbol{x}_{2\ 1:n}$ and $\boldsymbol{\theta}$ from (5.33). The other parameters of the bivariate process are estimated again assuming that $\{\underline{x}_{1t}\}$, $\{\underline{x}_{2t}\}$ are univariate HKps, however in this case we use the whole sample, starting from the common starting year of $\{\underline{x}_{1t}\}$ and $\{\underline{x}_{2t}\}$ until the year 2100 for the $\{\underline{x}_{1t}\}$ parameter estimates and the common end year of the corresponding 20C3M scenario and $\{\underline{x}_{2t}\}$ for the $\{\underline{x}_{2t}\}$ parameter estimates. The samples from the posterior predictive probability of $\underline{x}_t|\boldsymbol{x}_n$, $t = n+1, n+2,...$, were used to obtain samples for the variable of interest $\underline{x}_{2\ t(30)}$ given by (4.34).

Table 5.2. IPCC scenarios and their relevance to the study.

	Scenario	Characteristics	Reason for being appropriate or inappropriate
AR4	SRES	Various hypothetical scenarios for the future.	Runs start in the 21st century.
	A1B	A future world of very rapid economic growth, low population growth and rapid introduction of new and more efficient technology. Major underlying themes are economic and cultural convergence and capacity building, with a substantial reduction in regional differences in per capita income. In this world, people pursue personal wealth rather than environmental quality.	
	B1	A convergent world with the same global population as in the A1 storyline but with rapid changes in economic structures toward a service and information economy, with reductions in materials intensity, and the introduction of clean and resource-efficient technologies.	
	A2	A very heterogeneous world. The underlying theme is that of strengthening regional cultural identities, with an emphasis on family values and local traditions, high population growth, and less concern for rapid economic development.	
	COMMIT	Greenhouse gases fixed at year 2000 levels.	Runs start in the 21st century, however it is a conservative scenario.
	1%-2X, 1%-4X	Assume a 1%-per-year increase in CO2, usually starting at year 1850.	Results in CO2 being 570 cm^3/m^3 (ppm) already in 1920, when in fact it was 379 cm^3/m^3 in 2005. Actual 20th century concentrations are required.
	PI-cntrl	Uses pre-industrial greenhouse gas concentrations.	Actual 20th century concentrations are required.
	20C3M	Generated from output of late 19th & 20th century simulations from coupled ocean–atmosphere models, to help assess past climate change.	This scenario is used for calibration.

Sources: Leggett et al. (1992); IPCC (2000); IPCC (2007); IPCC-TGCIA (1999); Hegerl et al. (2003)

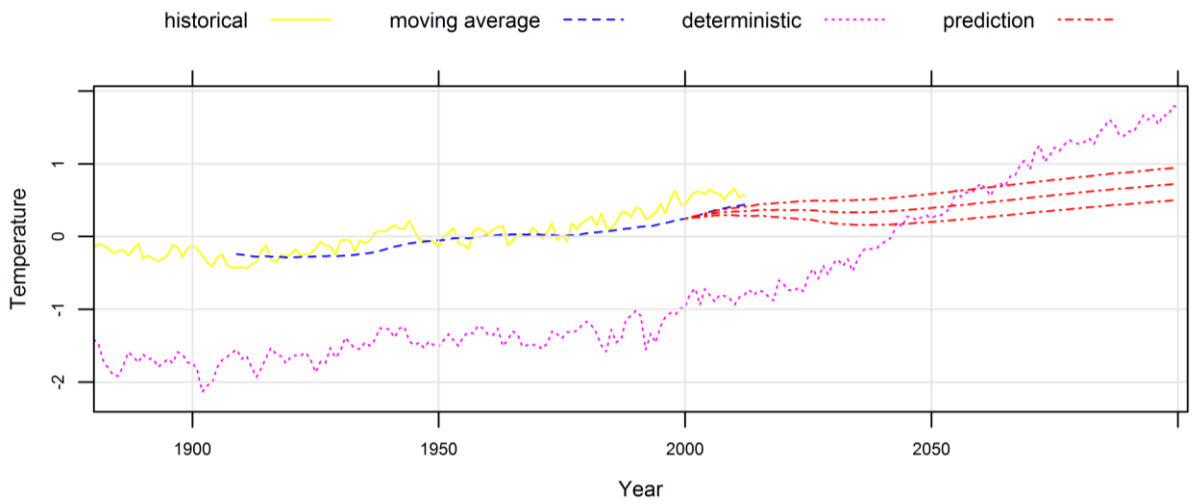


Figure 5.2. 95% confidence region for the predictive 30-moving average temperature (°C) for the A1B scenario of the ECHO-G model, using the NOAA annual global land and ocean temperature anomalies.

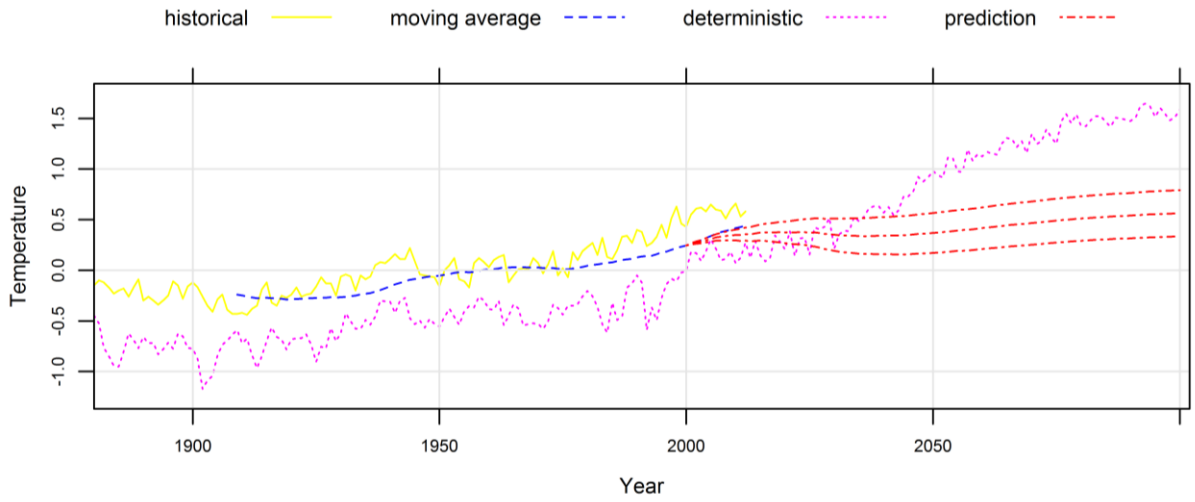


Figure 5.3. 95% confidence region for the predictive 30-moving average temperature (°C) for the B1 scenario of the ECHO-G model, using the NOAA annual global land and ocean temperature anomalies.

We preferred to present some specific examined cases with characteristics presented in Table 5.4. First we examined the scenarios A1B, B1, A2 of the ECHO G model on the NOAA annual global land and ocean temperature anomalies (Figures 5.2-5.4). The estimated ρ for this case was equal to 0.24 and rather moderate. A 95% confidence region for the predictive global climate temperature for the worst scenario A2 was of the order of 0 to 0.5°C more than the 2012 climate temperature. Figures 5.5-5.7 show the results for the scenario A1B of the CGCM3.1 (T63) for all temperature datasets. The estimated ρ took values at the range of 0.12 to 0.24. A 95% confidence region for the predictive global climate temperature for all historical datasets is of the order of −0.2 to 0.6°C more than the 2012 climate temperature. We also examined a case with relatively big estimated ρ's of the order of 0.26 to 0.38 for the

A1B of the UKMO HadGEM1 and for all temperature datasets. Figures 5.8-5.10 show the results for this scenario. A 95% confidence region for the predictive global climate temperature for all historical datasets is of the order of 0.7 to 1.2°C more than the 2012 climate temperature. For the CRU precipitation over land areas dataset we decided to show the results for the scenarios A1B, B1, A2 of the ECHO G model for the CRU precipitation over land areas (Figures 5.11-5.13). It seems that the model's output failed to fit to the historical datasets.

Table 5.3. Main characteristics of the GCMs used in the study.

IPCC report	Name	Developed by	Country
AR4	BCC CM1	Beijing Climate Center	China
	BCCR BCM2.0	Bjerknes Centre for Climate Research	Norway
	CCSM3.0	National Center for Atmospheric Research	USA
	CGCM3.1 (T47)	Canadian Centre for Climate Modelling & Analysis	Canada
	CGCM3.1 (T63)	Canadian Centre for Climate Modelling & Analysis	Canada
	CNRM CM3	Météo-France / Centre National de Recherches Météorologiques	France
	CSIRO Mk3.5	CSIRO Atmospheric Research	Australia
	ECHAM5 MPI-OM	Max Planck Institute for Meteorology	Germany
	ECHO G	Meteorological Institute of the University of Bonn, Meteorological Research Institute of KMA, and Model and Data group.	Germany/Korea
	FGOALS g1.0	LASG / Institute of Atmospheric Physics	China
	GFDL CM2.1	US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory	USA
	GISS ER	NASA / Goddard Institute for Space Studies	USA
	INGV ECHAM4	Instituto Nazionale di Geofisica e Vulcanologia	Italy
	INM CM3.0	Institute for Numerical Mathematics	Russia
	IPSL CM4	Institut Pierre Simon Laplace	France
	MIROC3.2 (medres)	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC)	Japan
	MRI CGCM 2.3.2	Meteorological Research Institute	Japan
	PCM	National Center for Atmospheric Research	USA
	UKMO HadCM3	Hadley Centre for Climate Prediction and Research / Met Office	UK
	UKMO HadGEM1	Hadley Centre for Climate Prediction and Research / Met Office	UK

Sources: http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php; climexp.knmi.nl

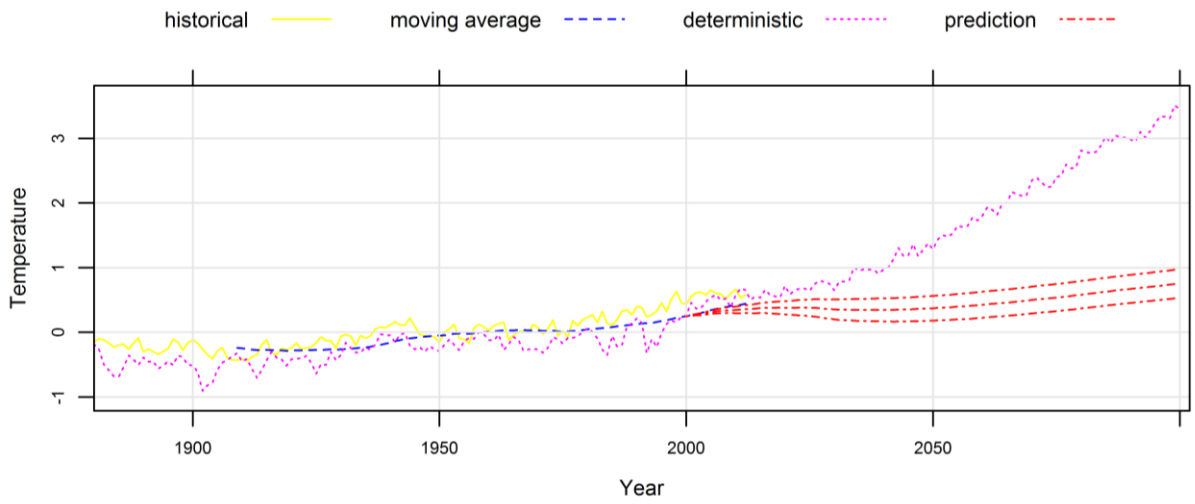


Figure 5.4. 95% confidence region for the predictive 30-moving average temperature (°C) for the A2 scenario of the ECHO-G model, using the NOAA annual global land and ocean temperature anomalies.

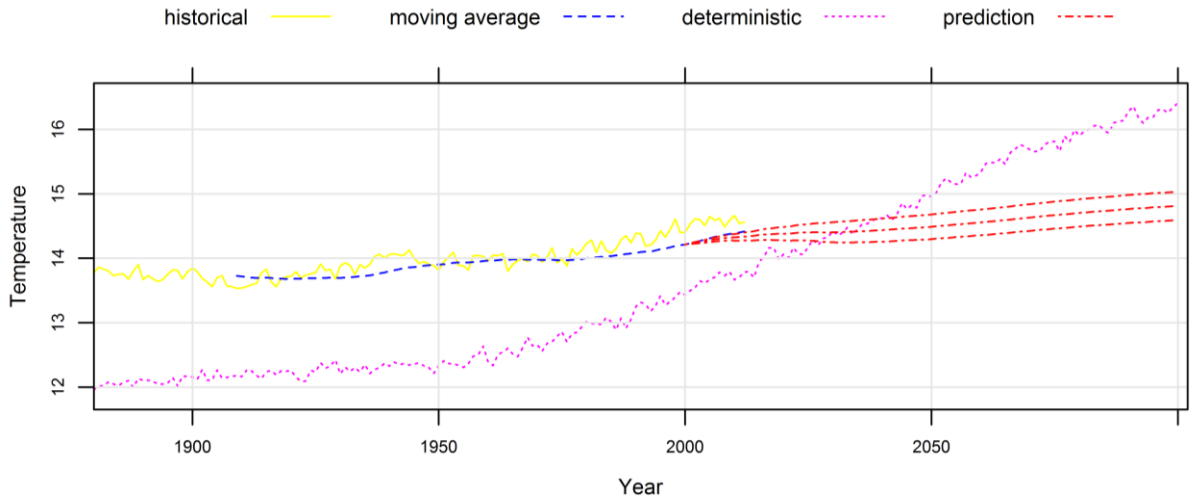


Figure 5.5. 95% confidence region for the predictive 30-moving average temperature (°C) for the A1B scenario of the CGCM3.1 (T63) model, using the GISS global land-ocean temperature index.

Table 5.4. Main characteristics of the cases presented in Figures 5.2-5.13.

Variable	Historical dataset developer	Model	Scenario	Estimated ρ	Corresponding table	Corresponding figure
Temperatue	NOAA	ECHO-G	A1B	0.24	Table B.2	Figure 5.2
Temperatue	NOAA	ECHO-G	B1	0.24	Table B.2	Figure 5.3
Temperatue	NOAA	ECHO-G	A2	0.24	Table B.2	Figure 5.4
Temperatue	GISS	CGCM3.1 (T63)	A1B	0.21	Table B.1	Figure 5.5
Temperatue	NOAA	CGCM3.1 (T63)	A1B	0.24	Table B.2	Figure 5.6
Temperatue	CRU	CGCM3.1 (T63)	A1B	0.12	Table B.3	Figure 5.7
Temperatue	GISS	UKMO HadGEM1	A1B	0.37	Table B.1	Figure 5.8
Temperatue	NOAA	UKMO HadGEM1	A1B	0.38	Table B.2	Figure 5.9
Temperatue	CRU	UKMO HadGEM1	A1B	0.26	Table B.3	Figure 5.10
Precipitation	CRU	ECHO-G	A1B	-0.03	Table B.4	Figure 5.11
Precipitation	CRU	ECHO-G	B1	-0.03	Table B.4	Figure 5.12
Precipitation	CRU	ECHO-G	A2	-0.03	Table B.4	Figure 5.13

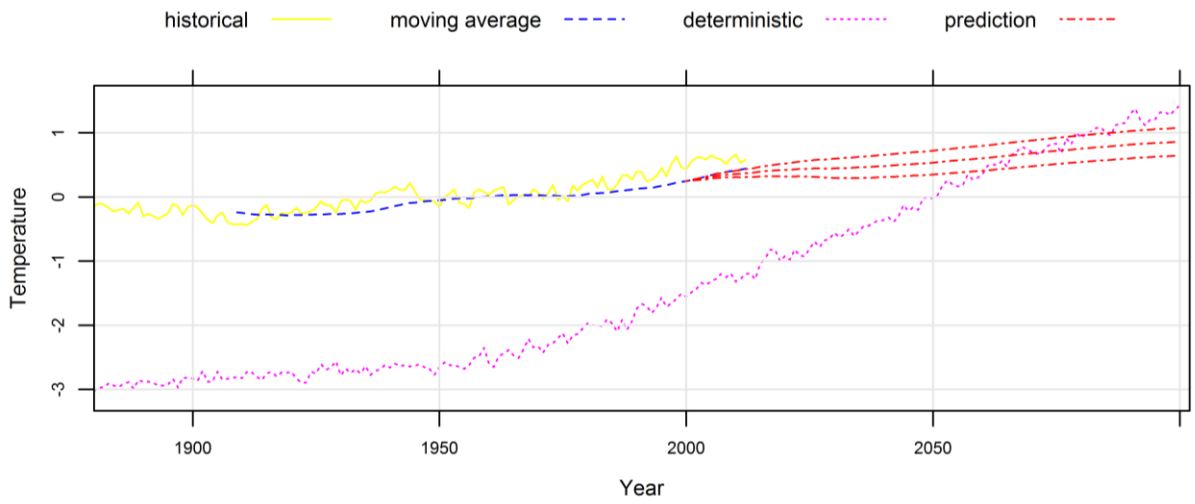


Figure 5.6. 95% confidence region for the predictive 30-moving average temperature (°C) for the A1B scenario of the CGCM3.1 (T63) model, using the NOAA annual global land and ocean temperature anomalies.

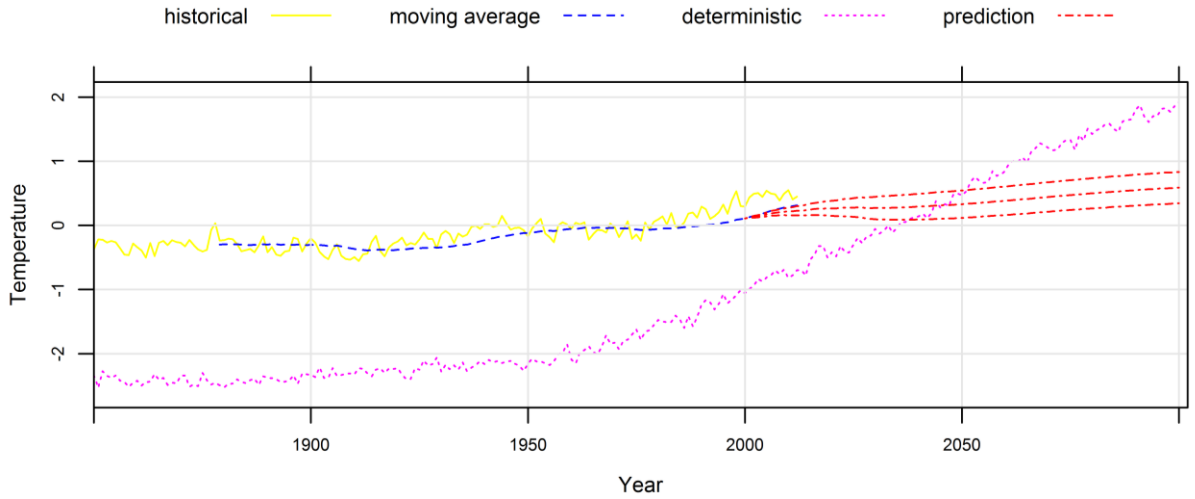


Figure 5.7. 95% confidence region for the predictive 30-moving average temperature (°C) for the A1B scenario of the CGCM3.1 (T63) model, using the CRU combined land [CRUTEM4] and marine temperature anomalies.

5.6 Summary and conclusions

The aim of this Chapter was to predict the future evolution of a LTP process used to model a geophysical phenomenon conditional on historical observations of the phenomenon and the hindcasts and predictions of a deterministic model of the phenomenon. To this end we modelled both time series (histrorical observations and deterministic model outputs) using the bivariate HKp. We derived a new MLE to estimate the parameters of the bivariate HKp. The parameters were given values equal to their estimations, and the distribution of the future variables conditional on the historical observations, the hindcasts and predictions of the deterministic model and the estimated parameters was derived.

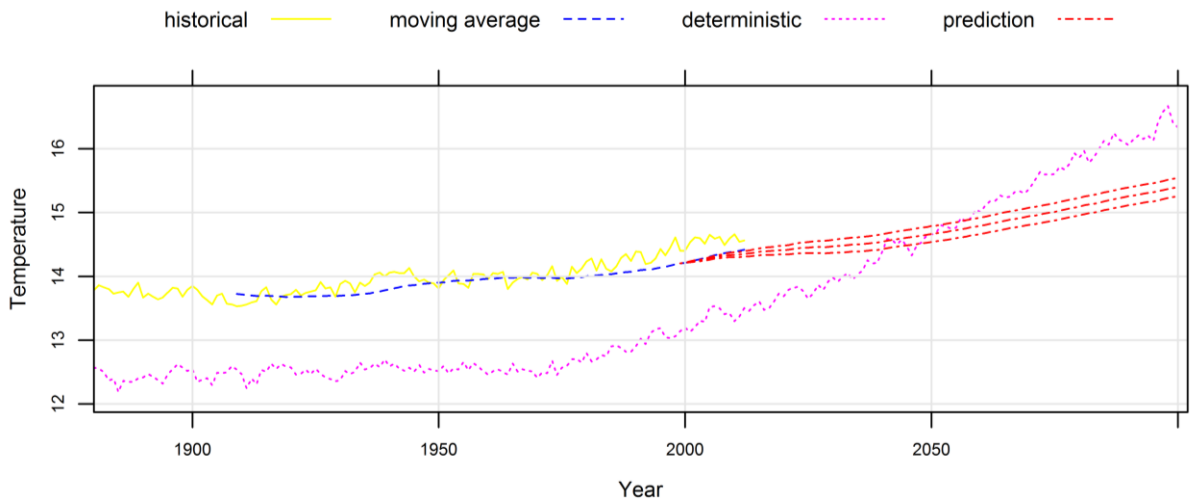


Figure 5.8. 95% confidence region for the predictive 30-moving average temperature (°C) for the A1B scenario of the UKMO HadGEM1 model, using the GISS global land-ocean temperature index.

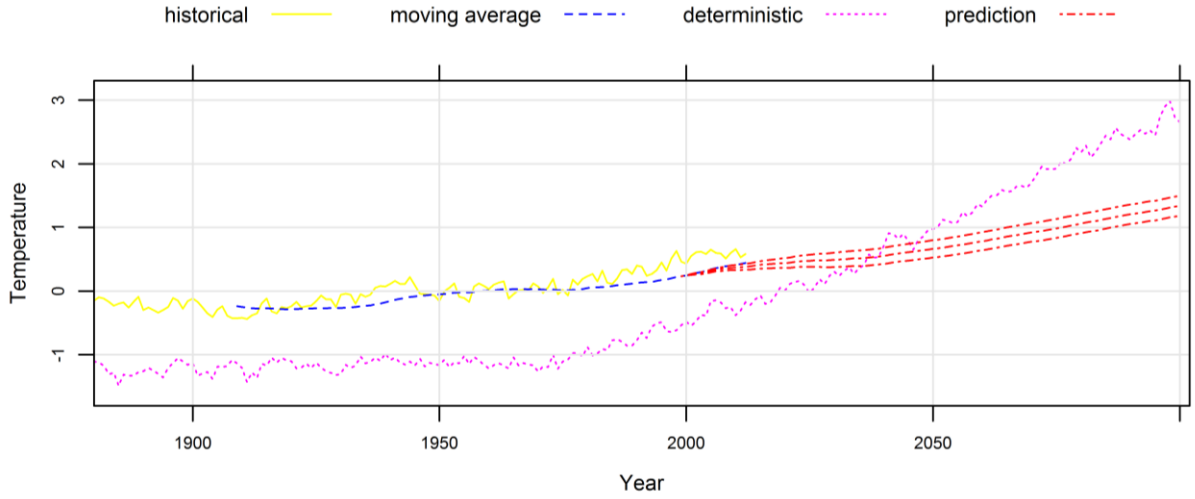


Figure 5.9. 95% confidence region for the predictive 30-moving average temperature (°C) for the A1B scenario of the UKMO HadGEM1 model, using the NOAA annual global land and ocean temperature anomalies.

The methodology was applied to historical global temperature and over land precipitation data. GCMs were used as deterministic models. Using the estimated values of the parameters we provided stochastic prediction of the future climate combining the projections of the GCMs and their corresponding hindcasts with the observed time series. It was found that the estimated values of the cross-correlation between the historical datasets (at global scale) and the hindcasts of the GCMs range from 0 to 0.4, showing that the information added by the GCMs to that contained in the historical datasets is not substantial. Hence the upper bound of the 95% confidence region of the climatic value of temperature at year 2100 was estimated to about 1.2°C more than the current value of this climatic variable. For the precipitation dataset the estimated value of the cross-correlations between the historical datasets and the hindcasts

of the GCMs was almost equal to 0. This meant that the output of the GCM had no effect on the stochastic predictions.

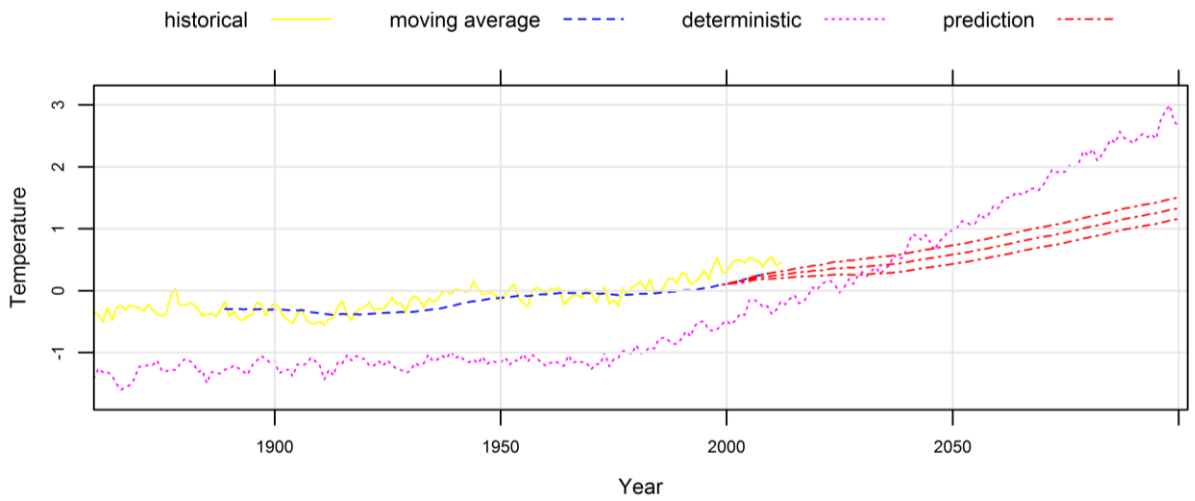


Figure 5.10. 95% confidence region for the predictive 30-moving average temperature (°C) for the A1B scenario of the UKMO HadGEM1 model, using the CRU combined land [CRUTEM4] and marine temperature anomalies.

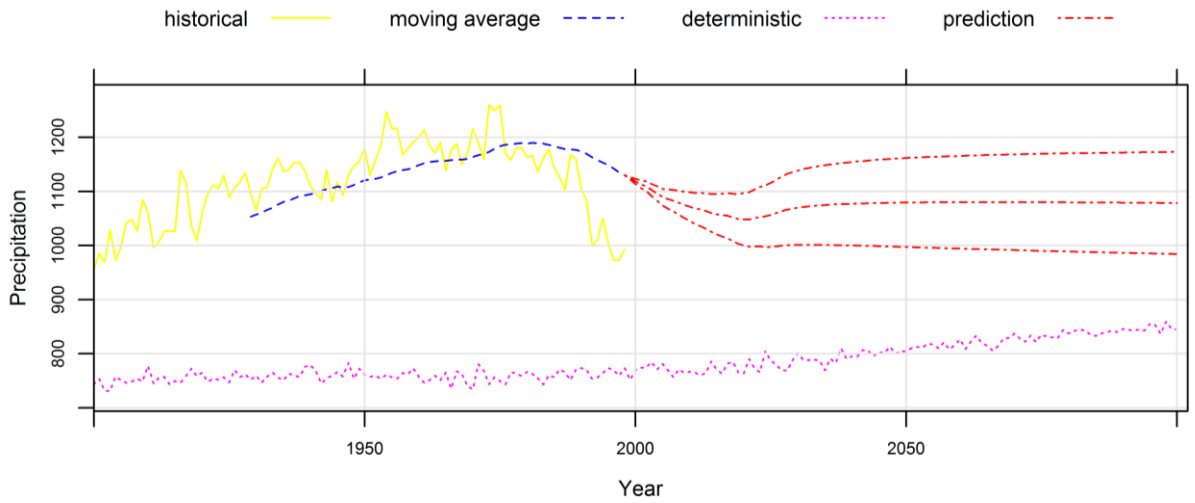


Figure 5.11. 95% confidence region for the predictive 30-moving average precipitation (mm) for the A1B scenario of the ECHO-G model, using the CRU precipitation over land areas.

We emphasize that the estimation of the stochastic model parameters should better be performed using only data that were not used in the GCM fitting/tuning, i.e. for the period after 2000. This would correspond to the so-called split-sample technique, which avoids possible model overfitting on the available data. However this would increase considerably the uncertainty of the estimators of the parameters of the models and practically would result in total neglect of the GCM predictions. Hence we decided to approach the problem more conservatively.

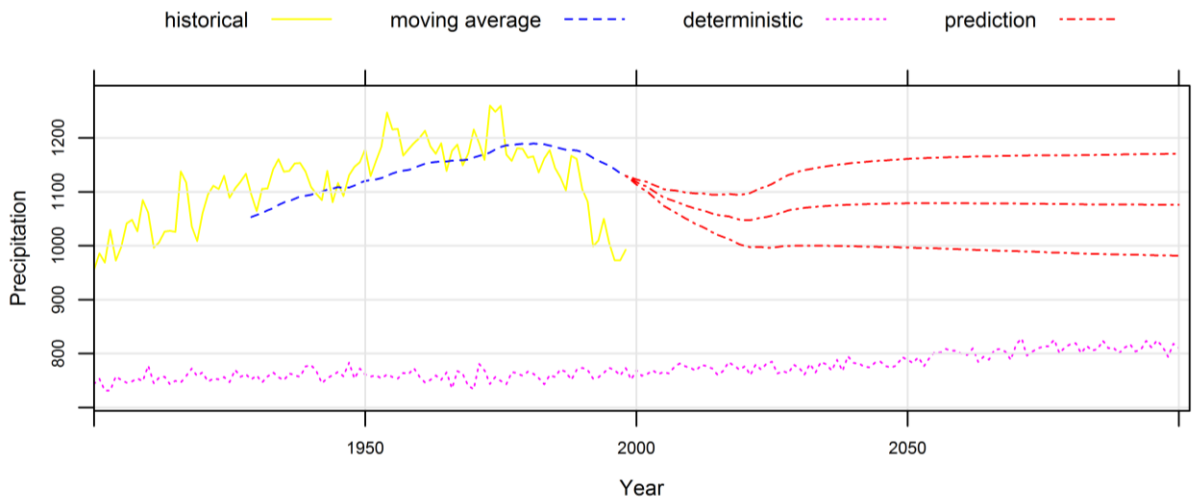


Figure 5.12. 95% confidence region for the predictive 30-moving average precipitation (mm) for the B1 scenario of the ECHO-G model, using the CRU precipitation over land areas.

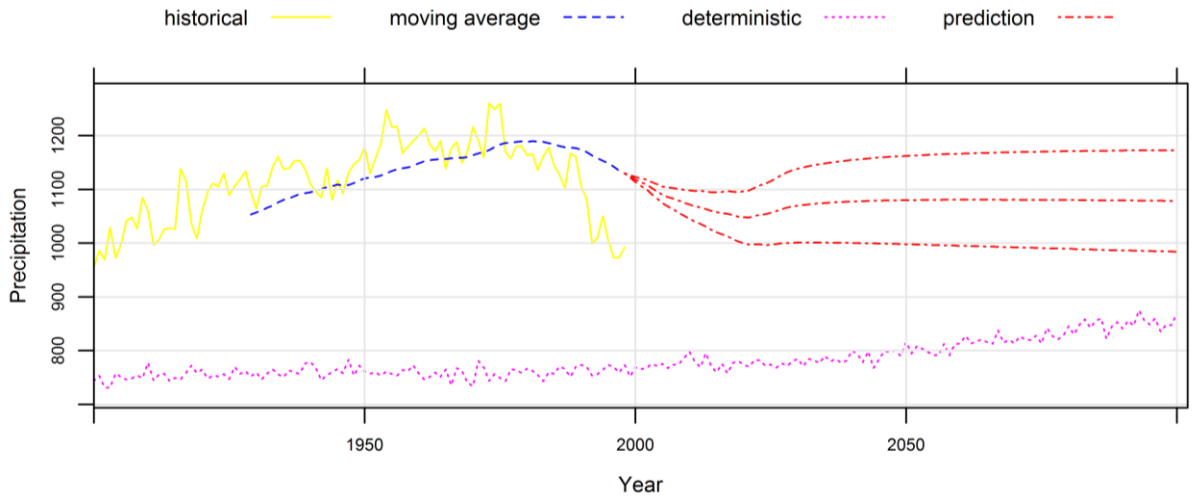


Figure 5.13. 95% confidence region for the predictive 30-moving average precipitation (mm) for the A2 scenario of the ECHO-G model, using the CRU precipitation over land areas.

Our approach is an extension of previous studies (Krzysztofowicz 1999a,b; Wang et al. 2009), which exploited the outputs of deterministic models combined with historical dataset, on persistent stochastic processes. In this study a methodology for LTP processes is proposed whereas in the previous studies only white noise and the AR(1) processes were examined. An expansion of the methodology to a Bayesian setting, in which also the uncertainty of parameters is accounted for, will be a next step.

6. Summary, conclusions and recommendations

The initial aim of this thesis was the development of a stochastic framework for the prediction of hydroclimatic processes using Bayesian techniques. To solve the problem we decided to select a parametric approach. Thus a stochastic model was chosen. The choice was based on well established a priori criteria, namely the second law of thermodynamics, which under certain constraints results to a family of stochastic models exhibiting HK behaviour. A Bayesian approach was adopted to find the posterior predictive distributions of the hydroclimatic variables of interest.

The thesis proceeded linearly as planned from the start. A previous study which developed a stochastic framework was investigated. The results of that study were encouraging. However it was based on a new heuristic algorithm. In this thesis we proved analytically that this algorithm is correct. Due to the limitations of the first approach we decided to solve the problem using Bayesian statistics. A first step to this direction is the assessment of the estimators of the parameters of the stochastic model. The results were again encouraging. Hence in the second step we solved the problem in a Bayesian way using a noninformative prior distribution for the parameters of the stochastic model. Last we decided to use information provided by deterministic models to improve the results of the stochastic model.

Of course this thesis can not solve exhaustively the problem as mentioned in Section 6.3. The state-of-the-art models for climate prediction are deterministic and research is focused on their development, despite their deficiencies. Little research has been conducted in the domain of stochastics. Therefore the stochastic approach could be considered innovative on its own. We hope that the analytical tools developed here add a building block to this effort.

6.1 Methodological contributions

6.1.1 A new algorithm to calculate confidence intervals

In Chapter 2 a Monte-Carlo algorithm for an approximation of a confidence interval of any parameter for any continuous distribution was proposed. It was shown that the algorithm is exact for a single parameter of distribution of either location or scale family. It is also asymptotically equivalent to a Wald-type interval for parameters of regular continuous distributions. After appropriate modification of the algorithm it was made appropriate for calculating confidence intervals for a parameter of multi-parameter distributions and it was shown that it is asymptotically equivalent to a Wald-type interval for regular distributions.

The algorithm was tested in several distributions and was found that in general works well and results in correct coverage probabilities. The algorithm is proposed for an approximation of a confidence interval of any parameter for any continuous distribution because it is easily applicable in every case and gives better approximations than other known algorithms as shown in specific cases. The algorithm was implemented in an earlier study examining future hydroclimatic variables for a better approximation of confidence intervals.

6.1.2 On the estimation of the parameters of the HKp

A simulation study to assess the performance of several estimators of the HK-process was performed in Chapter 3. It was found that three estimators (ML, LSSD and LSV) were more accurate when estimating the Hurst parameter of the HKp, compared to other estimators of the literature, probably because they are based on the structure of the HKp.

The LSV estimator is novel and follows closely the rationale of the construction of the LSSD, deferring in that its construction was based on analytical results. Properties of the LSV were found analytically, namely, the boundedness property of H and the behaviour of the estimator for high values of H. It is mentioned that other estimators, than the ones proposed here, yield estimates of H outside of its proper domain. The MLE was presented after appropriate streamlining, to be used in Chapter 3.

An additional advantage of these three estimators is that, in addition to H, they estimate σ which is essential for the statistical model. It was shown that σ and H are not orthogonal, thus their maximum likelihood estimators are correlated. On the other hand the pairs μ, σ and μ, H are orthogonal, thus the maximum likelihood estimate of σ or H varies only slowly with μ. As a consequence a non simultaneous estimator of σ and H may be suboptimal in terms of robustness comparing to the ML, LSSD or LSV estimators which estimate H and σ simultaneously.

6.1.3 The Bayesian statistical model

In Chapter 4 a Bayesian statistical model was proposed for estimating future confidence regions on the basis of a stationary normal stochastic process. Furthermore the problem for a truncated stationary normal stochastic process was solved as well. A noninformative prior of the parameters was chosen. The posterior distributions of the parameters of the model, the posterior predictive distributions of the variables of the process and the posterior predictive

distribution of the 30-year moving average which is the climatic variable of interest were derived after technical manipulations.

The methodology was applied to runoff, rainfall and temperature datasets from Greece, Vienna and Berlin to five cases. Three kinds of stationary normal processes were examined, namely WN, AR(1) and HKp. It was shown that the use of short-range dependence stochastic processes, i.e. WN and AR(1) is not suitable to model geophysical processes, because they underestimate uncertainty. However the HKp achieved this purpose. In the examined cases it performed well and was able to explain the fluctuations of the process.

Results associated with the estimation of the parameters on a Bayesian framework compared to typical statistical estimators, such as the ML estimator were also derived showing that in general the two estimators are almost equal. Posterior distributions of the parameters were derived and it was shown that they were wider for the HKp.

6.1.4 Incorporating information from deterministic models

Concluding the thesis we tried to encompass information from deterministic models. To this end datasets and deterministic model outputs were time modelled by the bivariate HKp. A new MLE of the parameters of the bivariate HKp was derived. The parameters were given values equal to their estimations, and the distribution of the future variables conditional on the historical observations, the hindcasts and predictions of the deterministic model and the parameters was derived.

We applied the method to historical global temperature and over land precipitation data. GCMs were used as deterministic models. It was shown that the information added by the GCMs to that contained in the historical datasets is not substantial. Hence the upper bound of the 95% confidence region of the climatic value of temperature at year 2100 was estimated to about 1.2°C more than the current value of this climatic variable. For the precipitation dataset the estimated value of the cross-correlations between the historical datasets and the hindcasts of the GCMs was almost equal to 0. This meant that the output of the GCM had no effect on the stochastic predictions.

6.2 Recommendations for further research

6.2.1 Technical issues

Regarding the topics studied in this thesis, there is a lot of space for improvements related to technical issues. Further research is needed to evaluate the influence of the choice of the numerical parameters (increments and the simulation sample size) to the results of the algorithm in Chapter 2.

The prior distribution of the parameters of the HKp in Chapter 4 could be determined automatically, directly derived from the sampling distribution in a noninformative approach, e.g. a Jeffreys prior, a reference prior or a matching prior. A comparison of different simulators for the bivariate HKp in the fashion of Chapter 3, would be a step for a better establishment of the new MLE estimator.

6.2.2 Further research

The stochastic framework developed in this thesis could be improved considerably. We could switch to an informative prior based on prior information, e.g. information from similar observed geophysical processes or use hierarchical models. The framework examines only normal variables. The incorporation of non-normal variables using appropriate transformations will be a next step. An extension of the framework to the multivariate case to examine multiple time series in adjacent regions will be another improvement.

The derivation of the MLE for the multivariate HKp is worth studying. The Bayesian expansion of the framework that incorporates information from deterministic models, will reveal uncertainties in a similar manner to the stochastic framework. Furthermore truncated normal variables could easily be examined within the framework. The examination of its mathematical relationship with the Bayesian Forecasting System will also offer opportunities for new developments. Furthermore the methods developed in this thesis could be applied to more datasets for obtaining more practical results.

6.3 Limitations

This thesis focused on the HKp to model geophysical processes. Despite its parsimony owing to the use of only three parameters it is limited when modelling complicated phenomena. More complex but also parsimonious models should be developed to model such phenomena. However we believe that the methods developed here could serve as building blocks in this

effort. These models for example could be derived setting different constraints when maximizing entropies or could be selected per se. Regional models towards the same direction could also be developed, e.g. the incorporation of deterministic information in the stochastic framework for the multivariate case could be examined.

References

Amblard PO, Coeurjolly JF (2011) Identification of the Multivariate Fractional Brownian Motion. IEEE Transactions on Signal Processing 59(11):5152-5168. doi:10.1109/TSP.2011.2162835

Amblard PO, Coeurjolly JF, Lavancier F, Philippe A (2012) Basic properties of the multivariate fractional Brownian motion. arXiv:1007.0828v2

Anagnostopoulos GG, Koutsoyiannis D, Christofides A, Efstratiadis A, Mamassis N (2010) A comparison of local and aggregated climate model outputs with observed data. Hydrological Sciences Journal 55(7):1094-1110. doi:10.1080/02626667.2010.513518

Bain LJ, Engelhardt M (1975) A Two-Moment Chi-Square Approximation for the Statistic $\mathrm{Log}(\bar{x}/\tilde{x})$. Journal of the American Statistical Association 70(352):948-950. doi:10.1080/01621459.1975.10480328

Bakker AMR, van den Hurk BJJM (2012) Estimation of persistence and trends in geostrophic wind speed for the assessment of wind energy yields in Northwest Europe. Climate Dynamics 39(3-4):767-782. doi:10.1007/s00382-011-1248-1

Beran J (1994) Statistics for Long-Memory Processes. Chapman & Hall/CRC, New York

Berliner LM, Wikle CK, Cressie N (2000) Long-Lead Prediction of Pacific SSTs via Bayesian Dynamic Modeling. Journal of Climate 13:3953-3968. doi:10.1175/1520-0442(2001)013<3953:LLPOPS>2.0.CO;2

Bhaumik DK, Kapur K, Gibbons RD (2009) Testing Parameters of a Gamma Distribution for Small Samples. Technometrics 51(3):326-334. doi:10.1198/tech.2009.07038

Bouette JC, Chassagneux JF, Sibai D, Terron R, Charpentier R (2006) Wind in Ireland: long memory or seasonal effect. Stochastic Environmental Research & Risk Assessment 20(3):141-151. doi:10.1007/s00477-005-0029-y

Casella G, Berger RL (2001) Statistical Inference, second edition. Duxbury Recourse Center, Pacific Grove

Chen C (2008) hurst parameter estimate. MatLab Central, File Exchange. Available online at:http://www.mathworks.com/matlabcentral/fileexchange

Choi SC, Wette R (1969) Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias. Technometrics 11(4):683-690. doi:10.1080/00401706.1969.10490731

Coeurjolly JF (2008) Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. The Annals of Statistics 36(3):1404-1434. doi:10.1214/009053607000000587

Coeurjolly JF, Amblard PO, Achard S (2010) On multivariate fractional Brownian motion and multivariate fractional Gaussian noise. Paper presented at the 18th European Signal Processing Conference 1567-1571

Coeurjolly JF, Amblard PO, Achard S (2013) Wavelet analysis of the multivariate fractional Brownian motion. ESAIM: Probability and Statistics 17:592-564. doi:10.1051/ps/2012011

Cohen AC (1965) Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples. Technometrics 7(4):579-588. doi:10.1080/00401706.1965.10490300

Cox DR, Reid N (1987) Parameter Orthogonality and Approximate Conditional Inference. Journal of the Royal Statistical Society Series B (Methodological) 49(1):1-39

DiCiccio TJ (1984) On Parameter Transformations And Interval Estimation. Biometrica 71(3):477-485. doi:10.1093/biomet/71.3.477

DiCiccio TJ, Romano JP (1995) On Bootstrap Procedures For Second-Order Accurate Confidence Limits In Parametric Models. Statistica Sinica 5:141-160

DiCiccio TJ, Efron B (1996) Bootstrap Confidence Intervals. Statistical Science 11(3):189-228. doi:10.1214/ss/1032280214
Doukhan P, Oppenheim G, Taqqu MS (eds) (2003) Theory and Applications of Long-Range Dependence. Birkhauser, Boston
Duan Q, Ajami NK, Gao X, Sorooshian S (2007) Multi-model ensemble hydrologic prediction using Bayesian model averaging. Advances in Water Resources 30(5):1371-1386. doi:10.1016/j.advwatres.2006.11.014
Eaton ML (2007) Multivariate Statistics: A Vector Space Approach Lecture Notes-Monograph Series 53. Institute of Mathematical Statistics. Beachwood, Ohio
Efron B (1987) Better Bootstrap Confidence Intervals. Journal of the American Statistical Association 82(397):171-185. doi:10.1080/01621459.1987.10478410
Efron B, Hinkley DV (1978) Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information. Biometrica 65(3):457-482. doi:10.1093/biomet/65.3.457
Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman & Hall/CRC, New York
Ehsanzadeh E, Adamowski K (2010) Trends in timing of low stream flows in Canada: impact of autocorrelation and long-term persistence. Hydrological Processes 24:970–980. doi:10.1002/hyp.7533
Embrechts P, Maejima M (2002) Selfsimilar Processes. Princeton University Press, Princeton
Engelhardt M, Bain LJ (1977) Uniformly Most Powerful Unbiased Tests on the Scale Parameter of a Gamma Distribution with a Nuisance Shape Parameter. Technometrics 19(1):77-81
Engelhardt M, Bain LJ (1978) Construction of Optimal Unbiased Inference Procedures for the Parameters of the Gamma Distribution. Technometrics 20(4):485-489. doi:10.1080/00401706.1978.10489703
Esposti F, Ferrario M, Signorini MG (2008) A blind method for the estimation of the Hurst exponent in time series: Theory and application. Chaos: An Interdisciplinary Journal of Nonlinear Science 18(033126). doi:10.1063/1.2976187
Falconer K, Fernadez C (2007) Inference on fractal processes using multiresolution approximation. Biometrica 94(2):313-334. doi:10.1093/biomet/asm025
Fyfe JC, Gillet NP, Zwiers FW (2013) Overestimated global warming over the past 20 years. Nature Climate Change 3:767-769. doi:10.1038/nclimate1972
Gamerman D, Lopes HF (2006) Markov Chain Monte Carlo, Stochastic Simulation for Bayesian Inference, second edition, Chapman & Hall/CRC, Boca Raton
Garthwaite PH, Buckland ST (1992) Generating Monte Carlo Confidence Intervals by the Robbins-Monro Process. Journal of the Royal Statistical Society, Series C (Applied Statistics) 41(1):159-171
Gelman A (1996) Inference and monitoring convergence. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds) Markov Chain Monte Carlo in Practice. Chapman & Hall, New York, pp 131-143
Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian Data Analysis, second edition. Chapman & Hall/CRC, Boca Raton
Gelman A, Rubin DR (1992) A single series from the Gibbs sampler provides a false sense of security. In Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian Statistics 4. Oxford University Press, Oxford, pp 625-632
Golub GH, Van Loan CF (1996) Matrix Computations, third edition. John Hopkins University Press, Baltimore
Grau-Carles P (2005) Tests of Long Memory: A Bootstrap Approach. Computational Economics 25(1-2):103-113. doi:10.1007/s10614-005-6277-6

Guerrero A, Smith LA (2005) A maximum likelihood estimator for long-range persistence. Physica A: Statistical Mechanics and its Applications 355(2-4):619-632. doi:10.1016/j.physa.2005.03.002

Hall P (1988) Theoretical Comparison of Bootstrap Confidence Intervals. The Annals of Statistics 16(3):927-953. doi:10.1214/aos/1176350933

Handorf D, Dethloff K (2012) How well do state-of-the-art atmosphere-ocean general circulation models reproduce atmospheric teleconnection patterns?. Tellus A 64:19777. doi:10.3402/tellusa.v64i0.19777

Hegerl GC, Meehl G, Covey C, Latif M, McAvaney B, Stouffer R (2003) 20C3M: CMIP collecting data from 20th century coupled model simulations. CLIVAR Exchanges 26 8(1)

Heidelberger P, Welch PD (1983) Simulation run length control in the presence of an initial transient. Operations Research 31(6):1109-1144. doi:10.1287/opre.31.6.1109

Hemelrijk J (1966) Underlining random variables. Statistica Neerlandica 20(1):1-7. doi:10.1111/j.1467-9574.1966.tb00488.x

Hillier G, Armstrong M (1999) The Density of the Maximum Likelihood Estimator. Econometrica 67(6):1459-1470. doi:10.1111/1468-0262.00086

Horn RA, Zhang F (2005) Basic Properties of the Schur Complement. In Zhang F (ed) The Schur Complement and Its Applications. Springer US, NY. doi:10.1007/b105056

Horrace WC (2005) Some results on the multivariate truncated normal distribution. Journal of Multivariate Analysis 94(1):209-221. doi:10.1016/j.jmva.2004.10.007

Hurst HE (1951) Long term storage capacities of reservoirs. Transactions of the American Society of Civil Engineers 116:776-808 (published in 1950 as Proceedings Separate no.11)

IPCC (2000) Special Report on Emissions Scenarios. Nakicenovic N, Swart RJ (eds). Cambridge University Press, Cambridge, United Kingdom

IPCC (2007) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tingor M, Miller HL (eds). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

IPCC-TGCIA (1999) Guidelines on the use of scenario data for climate impact and adaptation assessment. Carter TR, Hulme M, Lal M (prepared by). Intergovernmental Panel on Climate Change, Task Group on Scenarios For Climate Impact Assessment

Itia research group (2009-2012) Hydrognomon: Hydrological time series processing software. Available online at: hydrognomon.org

Keenan DJ (2014) Statistical Analyses of Surface Temperatures in the IPCC Fifth Assessment Report. Available online at: http://www.informath.org/AR5stat.pdf

Kisielinska J (2012) The exact bootstrap method shown on the example of the mean and variance estimation. Computational Statistics 28(3)1061-1077. doi:10.1007/s00180-012-0350-0

Kite GW (1988) Frequency and Risk Analyses in Hydrology. Water Recources Publications, Fort Collins

Kolmogorov AE (1940) Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum. Dokl Akad Nauk URSS 26:115–118

Koutsoyiannis D (1997) Statistical Hydrology, fourth edition. National Technical University of Athens, Athens

Koutsoyiannis D (2000) A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. Water Resources Research 36(6):1519-1533. doi:10.1029/2000WR900044

Koutsoyiannis D (2002) The Hurst phenomenon and fractional Gaussian noise made easy. Hydrological Sciences Journal 47(4):573–595. doi:10.1080/02626660209492961
Koutsoyiannis D (2003) Climate change, the Hurst phenomenon, and hydrological statistics. Hydrological Sciences Journal 48(1):3–24. doi:10.1623/hysj.48.1.3.43481
Koutsoyiannis D (2006a) A toy model of climatic variability with scaling behaviour. Journal of Hydrology 322(1-4):25–48. doi:10.1016/j.jhydrol.2005.02.030
Koutsoyiannis D (2006b) Nonstationarity versus scaling in hydrology. Journal of Hydrology 324(1-4):239–254. doi:10.1016/j.jhydrol.2005.09.022
Koutsoyiannis D (2010) A random walk on water. Hydrology and Earth System Sciences 14:585–601. doi:10.5194/hess-14-585-2010
Koutsoyiannis D (2011) Hurst-Kolmogorov dynamics as a result of extremal entropy production. Physica A: Statistical Mechanics and its Applications 390(8):1424–1432. doi:10.1016/j.physa.2010.12.035
Koutsoyiannis D (2015) Generic and parsimonious stochastic modelling for hydrology and beyond. Hydrological Sciences Journal. doi:10.1080/02626667.2015.1016950
Koutsoyiannis D, Efstratiadis A, Georgakakos KP (2007) Uncertainty Assessment of Future Hydroclimatic Predictions: A Comparison of Probabilistic and Scenario-Based Approaches. Journal of Hydrometeorology 8(3):261-281. doi:10.1175/JHM576.1
Koutsoyiannis D, Efstratiadis A, Mamassis N, Christofides A (2008a) On the credibility of climate predictions. Hydrological Sciences Journal 53(4):671-684. doi:10.1623/hysj.53.4.671
Koutsoyiannis D, Kozanis S (2005) A simple Monte Carlo methodology to calculate generalized approximate confidence intervals. Hydrologic Research Center. San Diego 25. Available online at:https://itia.ntua.gr/en/docinfo/692/
Koutsoyiannis D, Montanari A (2007) Statistical analysis of hydroclimatic time series: Uncertainty and insights. Water Resources Research 43(W05429). doi:10.1029/2006WR005592
Koutsoyiannis D, Yao H, Georgakakos A (2008b) Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. Hydrological Sciences Journal 53(1):142-164. doi:10.1623/hysj.53.1.142
Krzysztofowicz R (1987a) Bayesian models of forecasted time series. Journal of the American Water Resources Association 21(5):805-814. doi:10.1111/j.1752-1688.1985.tb00174.x
Krzysztofowicz R (1987b) Markovian Forecast Processes. Journal of the American Statistical Association 82(397):31-37. doi:10.1080/01621459.1987.10478387
Krzysztofowicz R (1999a) Bayesian Forecasting via Deterministic Model. Risk Analysis 19(4):739-749. doi:10.1111/j.1539-6924.1999.tb00443.x
Krzysztofowicz R (1999b) Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research 35(9):2739-2750. doi:10.1029/1999WR900099
Krzysztofowicz R (2001) Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. Journal of Hydrology 249(1-4):69-85. doi:10.1016/S0022-1694(01)00413-9
Krzysztofowicz R (2002) Bayesian system for probabilistic river stage forecasting. Journal of Hydrology 268(1-4):16-40. doi:10.1016/S0022-1694(02)00106-3
Krzysztofowicz R, Evans WB (2008) The role of climatic autocorrelation in probabilistic forecasting. Monthly Weather Review 136(12):4572-4592. doi:10.1175/2008MWR2375.1
Krzysztofowicz R, Maranzano CJ (2004) Bayesian system for probabilistic stage transition forecasting. Journal of Hydrology 299(1-2):15-44. doi:10.1016/j.jhydrol.2004.02.013

Lavancier F, Philippe A, Surgailis D (2009) Covariance function of vector self-similar processes. Statistics & Probability Letters 79(23):2415-2421. doi:10.1016/j.spl.2009.08.015

Lawless JF (2003) Statistical Models and Methods for Lifetime Data, second edition. John Wiley & Sons, New Jersey

Leggett J, Pepper WJ, Swart RJ (1992) Emissions scenarios for the IPCC: an update. In Houghton TJ, Berger JO, Callander BA, Varney SK (eds) Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment. Cambridge University Press, Cambridge, UK, pp 69-95

Macilwain C (2014) A touch of the random. Science 344(6189):1221-1223. doi:10.1126/science.344.6189.1221

Maity R, Nagesh Kumar D (2006) Bayesian dynamic modeling for monthly Indian summer monsoon using El Nino-Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINOO). Journal of Geophysical Research: Atmospheres (1984–2012) 111(D7). doi:10.1029/2005JD006539

Mandelbrot BB, van Ness JW (1968) Fractional Brownian motion, fractional noises and applications. SIAM Review 10(4):422–437. doi:10.1137/1010093

Markonis Y, Koutsoyiannis D (2013) Climatic variability over time scales spanning nine orders of magnitude: Connecting Milankovitch cycles with Hurst-Kolmogorov dynamics. Surveys in Geophysics 34(2):181–207. doi:10.1007/s10712-012-9208-9

Martin AD, Quinn KM, Park JH (2011) MCMCpack: Markov chain Monte Carlo in R. Journal of Statistical Software 42(9): 1-21

McLeod AI, Hippel KW (1978) Preservation of the Rescaled Adjusted Range 1. A Reassesment of the Hurst Phenomenon. Water Resources Research 14(3):491-508. doi:10.1029/WR014i003p00491

McLeod AI, Yu H, Krougly ZL (2007) Algorithms for linear time series analysis: With R package. Journal of Statistical Software 23(5):1-26

McNider RT, Steeneveld GJ, Holtslag AAM, Pielke Sr RA, Mackaro S, Pour-Biazar A, Walters J, Nair U, Christy J (2012) Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing. Journal of Geophysical Research: Atmospheres 117(D14). doi:10.1029/2012JD017578

Mielniczuk J, Wojdyllo P (2007) Estimation of Hurst exponent revisited. Computational Statistics & Data Analysis 51(9):4510-4525. doi:10.1016/j.csda.2006.07.033

Montanari A, Grossi G (2008) Estimating the uncertainty of hydrological forecasts: A statistical approach. Water Resources Research 44(W00B08). doi:10.1029/2008WR006897

Musicus BR (1988) Levinson and Fast Cholesky Algorithms for Toeplitz and Almost Toeplitz Matrices. Research Laboratory of Electronics Massachusetts Institute of Technology Cambridge 538. Available online at:http://citeseerx.ist.psu.edu

Nagesh Kumar D, Maity R (2008) Bayesian dynamic modeling for nonstationary hydroclimatic time series forecasting along with uncertainty quantification. Hydrological Processes 22(17):3488-3499. doi:10.1002/hyp.6951

Palma W (2007) Long-Memory Time Series Theory and Methods. John Wiley & Sons. New Jersey

Papoulis A, Pillai SU (2002) Probability, Random Variables and Stochastic Processes, fourth edition. McGraw-Hill. New York

Phillips NA (1956) The general circulation of the atmosphere: A numerical experiment. Quarterly Journal of the Royal Meteorological Society 82(352):123-164. doi:10.1002/qj.49708235202

Plummer M, Best N, Cowles K, Vines K (2006) CODA: Convergence Diagnosis and Output Analysis for MCMC. R News 6:7-11

Pokhrel P, Robertson DE, Wang QJ (2013) A Bayesian joint probability post-processor for reducing errors and quantifying uncertainty in monthly streamflow predictions. Hydrology and Earth Systems Sciences 17:795-804. doi:10.5194/hess-17-795-2013

Raftery AL, Lewis S (1992) How many iterations in the Gibbs sampler? In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian Statistics 4. Oxford University Press, Oxford, pp 763-774

Rea W, Oxley L, Reale M, Brown J (2013) Not all estimators are born equal: The empirical properties of some estimators of long memory. Mathematics and Computers in Simulation 93:29-42. doi:10.1016/j.matcom.2012.08.005

Ripley BD (1987) Stochastic Simulation. John Wiley & Sons. New York

Robert CP (2007) The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, second edition. Springer. New York

Robert CP, Casella G (2004) Monte Carlo Statistical Methods, second edition, Springer. New York

Robinson PM (1995) Gaussian semiparametric estimation of long range dependence. The Annals of Statistics 23(5):1630-1661. doi:10.1214/aos/1176324317

Robinson PM (ed) (2003) Time Series with Long Memory. Oxford University Press, Oxford

Román-Montoya Y, Rueda M, Arcos M (2008) Confidence intervals for quantile estimation using Jackknife techniques. Computational Statistics 23(4):573-585. doi:10.1007/s00180-007-0099-z

Rozos E, Efstratiadis A, Nalbantis I, Koutsoyiannis D (2004) Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows. Hydrological Sciences Journal 49(5):819-842. doi:10.1623/hysj.49.5.819.55130

Scafetta N (2013) Solar and Planetary Oscillation Control on Climate Change: Hind-Cast, Forecast and a Comparison with the CMIP5 GCMS. Energy & Environment 24(3-4):455-496. doi:10.1260/0958-305X.24.3-4.455

Shao J (2003) Mathematical Statistics, second edition, Springer. New York

Shimotsu K (2004) Matlab codes for local Whittle estimation and exact local Whittle estimation of the memory parameter (d) in fractionally integrated (I(d)) time series. Available online at:qed.econ.queensu.ca/faculty/shimotsu/

Smith BJ (2007) boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. Journal of Statistical Software 21(11):1-37

Smith PJ, Beven KJ, Weerts AH, Leedal D (2012) Adaptive correction of deterministic models to produce probabilistic forecasts. Hydrology and Earth Systems Sciences 16:2783-2799. doi:10.5194/hess-16-2783-2012

Son YS, Oh M (2006) Bayesian Estimation of the Two-Parameter Gamma Distribution. Communications in Statistics-Simulation and Computation 35(2):285-293. doi:10.1080/03610910600591925

Spencer RW, Braswell WD (2011) On the Misdiagnosis of Surface Temperature Feedbacks from Variations in Earth's Radiant Energy Balance. Remote Sensing 3(8):1603-1613. doi:10.3390/rs3081603

Stevens B, Bony S (2013) What Are Climate Models Missing?. Science 340(6136):1053-1054. doi:10.1126/science.1237554

Stoev S (2008) Simulation of Fractional Gaussian Noise *EXACT*. MatLab Central, File Exchange. Available online at:www.mathworks.com/matlabcentral/fileexchange

Taqqu MS, Teverovsky V, Willinger W (1995) Estimators for long-range dependence: an empirical study. Fractals 3(4):785-798. doi:10.1142/S0218348X95000692

Tyralis H, Koutsoyiannis D (2011) Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process. Stochastic Environmental Research & Risk Assessment 25(1):21-33. doi:10.1007/s00477-010-0408-x

Tyralis H, Koutsoyiannis D (2014) A Bayesian statistical model for deriving the predictive distribution of hydroclimatic variables. Climate Dynamics 42(11-12):2867-2883. doi:10.1007/s00382-013-1804-y

Tyralis H, Koutsoyiannis D (2015) On the prediction of persistent processes using the output of deterministic models. To be submitted

Tyralis H, Koutsoyiannis D, Kozanis S (2013) An algorithm to construct Monte Carlo confidence intervals for an arbitrary function of probability distribution parameters. Computational Statistics 28(4):1501-1527. doi:10.1007/s00180-012-0364-7

Wang QJ, Robertson DE, Chiew FHS (2009) A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. Water Resources Research 45(W05407). doi:10.1029/2008WR007355

Wei WWS (2006) Time Series Analysis, Univariate and Multivariate Methods, second edition. Pearson Addison Wesley

Weron R (2002) Estimating long-range dependence: finite sample properties and confidence intervals. Physica A: Statistical Mechanics and its Applications 312(1-2):285-299. doi:10.1016/S0378-4371(02)00961-5

Wilks SS (1938) Shortest Average Confidence Intervals from Large Samples. The Annals of Mathematical Statistics 9(3):166-175

Yang Z, Xie M, Wong ACM (2007) A unified confidence interval for reliability-related quantities of two-parameter Weibull distribution. Journal of Statistical Computation and Simulation 77(5):365-378. doi:10.1080/00949650701227452

Zhang Q, Xu CY, Yang T (2009) Scaling properties of the runoff variations in the arid and semi-arid regions of China: a case study of the Yellow River basin. Stochastic Environmental Research & Risk Assessment 23(8):1103-1111. doi:10.1007/s00477-008-0285-8

Zhao L, Duan Q, Schaake J, Ye A, Xia J (2011) A hydrologic post-processor for ensemble streamflow predictions. Advances in Geosciences 29:51-59. doi:10.5194/adgeo-29-51-2011

Appendix A Notations and distributions

The notations used in this thesis unless stated otherwise are summarized in Table A.1.

Table A.1. Notations.

Symbol	Notation
x, y, z	Observations
$\underline{x}, \underline{y}, \underline{z}$	Random variables
$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$	Vectors or Matrices
f, g, h	Densities
$f(x\|\theta)$	Density of $\underline{x}$ conditional on the parameter θ
π	Densities for parameters (Bayesian setting)
F, G, H	Distribution functions
$F(x\|\theta)$	Distribution function of $\underline{x}$ conditional on the parameter θ

For easy reference, the details of the distribution functions used in this thesis are summarized in Table A.2.

Table A.2. Distributions used in the Bayesian framework.

Distribution	Notation	Parameters	Density function
Chi-square	$\underline{x} \sim \chi^2(\nu)$	degrees of freedom $\nu > 0$	$f_{\chi^2}(x\|\nu) = (1/2)^{\nu/2} [\Gamma(\nu/2)]^{-1} x^{\nu/2-1} \exp(-x/2)$, $x > 0$
Inverse chi-square	$\underline{x} \sim \text{Inv-}\chi^2(\nu)$	degrees of freedom $\nu > 0$	$F_{\text{Inv-}\chi^2}(x\|\nu) = (1/2)^{\nu/2} [\Gamma(\nu/2)]^{-1} x^{-(\nu/2+1)} \exp(-1/2x)$, $x > 0$
Exponential	$\underline{x} \sim \text{EXP}(\sigma)$	scale $\sigma > 0$	$f_{\text{EXP}}(x\|\sigma) = (1/\sigma)\exp(-x/\sigma)$
Gamma	$\underline{x} \sim \text{gamma}(\alpha,\beta)$	shape $\alpha > 0$ scale $1/\beta > 0$	$f_{\text{G}}(x\|\alpha,\beta) = \beta^{\alpha} [\Gamma(\alpha)]^{-1} x^{\alpha-1} \exp(-\beta x)$, $x > 0$
Inverse-gamma	$\underline{x} \sim \text{Inv-gamma}(\alpha,\beta)$	$\alpha > 0$ $\beta > 0$	$f_{\text{IG}}(x\|\alpha,\beta) = \beta^{\alpha} [\Gamma(\alpha)]^{-1} x^{-(\alpha+1)} \exp(-\beta/x)$, $x > 0$
Normal	$\underline{x} \sim \text{N}(\mu,\sigma^2)$	location μ scale $\sigma > 0$	$f_{\text{N}}(x\|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp[(-1/2\sigma^2)(x-\mu)^2]$
Truncated normal	$\underline{x} \sim \text{TN}(\mu,\sigma^2,a,b)$	location μ scale $\sigma > 0$ a minimum value b maximum value	$f_{\text{TN}}(x\|\mu,\sigma^2,a,b) = [\Phi((b-\mu)/\sigma) - \Phi((a-\mu)/\sigma)]^{-1}(1/\sigma)\,\Phi((x-\mu)/\sigma)$ $x \in [a,b]$, $\Phi(x) := f_{\text{N}}(x\|0,1^2)$
Multivariate normal	$\underline{\boldsymbol{x}} \sim \text{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ (implicit dimension n)	location $\boldsymbol{\mu}$ symmetric, pos. definite n x n variance matrix $\boldsymbol{\Sigma}$	$f_{\text{MN}}(\boldsymbol{x}\|\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{-n/2} \lvert\boldsymbol{\Sigma}\rvert^{-1/2} \exp[(-1/2)(\boldsymbol{x}-\boldsymbol{\mu})^{\text{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})]$
Student-t	$\underline{x} \sim \text{t}_n(\mu,\sigma^2)$	degrees of freedom n location μ scale $\sigma > 0$	Not needed in the manuscript
Weibull	$\underline{x} \sim \text{Weibull}(a,b)$	scale $a > 0$ shape $b > 0$	$f_{\text{W}}(x\|a,b) = (b/a)\,(x/a)^{b-1} \exp(-(x/a)^b)$

Appendix B Results for deterministic models

Table B.1. Maximum likelihood estimates for the parameters of the bivariate HKp for the GISS global land-ocean temperature index.

GCM		Time	Simultaneous MLE					Separate MLE						
			ρ	σ_1	σ_2	H_1	H_2	ρ	μ_1	μ_2	σ_1	σ_2	H_1	H_2
BCC CM1	itas_bcc_cm1_20c3m_0-360E_-90-90N_n_su_00	1871-2003	0.37	0.15	0.40	0.89	0.98	0.84	16.82	14.00	0.20	0.50	0.95	0.99
	itas_bcc_cm1_20c3m_0-360E_-90-90N_n_su_01	1871-2003	0.44	0.14	0.36	0.88	0.98	0.89	16.81	14.00	0.20	0.50	0.96	0.99
BCCR BCM2.0	itas_bccr_bcm2_0_20c3m_0-360E_-90-90N_n_su	1850-1999	-0.05	0.26	0.47	0.97	0.99	0.53	12.56	13.97	0.22	0.41	0.97	0.99
CGCM3.1 (T63)	itas_cccma_cgcm3_1_t63_20c3m_0-360E_-90-90N_n_su	1850-2000	0.21	0.87	0.31	0.995	0.97	0.86	12.53	13.97	0.92	0.42	0.995	0.99
CNRM CM3	itas_cnrm_cm3_20c3m_0-360E_-90-90N_n_su	1860-1999	0.08	0.49	0.43	0.96	0.99	0.80	13.09	13.97	0.49	0.41	0.97	0.99
CSIRO Mk3.5	itas_csiro_mk3_5_20c3m_0-360E_-90-90N_n_su_01	1871-2000	0.08	0.77	0.44	0.99	0.99	0.76	15.13	13.97	0.73	0.42	0.99	0.99
ECHAM5 MPI-OM	itas_mpi_echam5_20c3m_0-360E_-90-90N_n_su_03	1860-2000	0.06	0.28	0.45	0.90	0.99	0.56	14.23	13.97	0.27	0.42	0.90	0.99
ECHO G	itas_miub_echo_g_20c3m_0-360E_-90-90N_n_su_00	1860-2000	0.27	0.46	0.40	0.98	0.98	0.79	13.57	13.97	0.49	0.42	0.99	0.99
	itas_miub_echo_g_20c3m_0-360E_-90-90N_n_su_02	1860-2000	0.26	0.52	0.39	0.99	0.98	0.78	13.51	13.97	0.55	0.42	0.99	0.99
FGOALS g1.0	itas_iap_fgoals1_0_g_20c3m_0-360E_-90-90N_n_su_00	1850-1999	0.05	0.27	0.44	0.71	0.99	0.30	12.42	13.97	0.25	0.41	0.72	0.99
	itas_iap_fgoals1_0_g_20c3m_0-360E_-90-90N_n_su_01	1850-1999	0.02	0.30	0.44	0.75	0.99	0.51	12.35	13.97	0.28	0.41	0.76	0.99
	itas_iap_fgoals1_0_g_20c3m_0-360E_-90-90N_n_su_02	1850-1999	0.02	0.27	0.44	0.68	0.99	0.34	12.41	13.97	0.25	0.41	0.69	0.99
GFDL CM2.1	itas_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_su_00	1861-2000	0.13	0.49	0.42	0.95	0.99	0.73	13.31	13.97	0.52	0.42	0.96	0.99
	itas_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_su_01	1861-2000	0.17	0.60	0.42	0.98	0.99	0.70	13.33	13.97	0.61	0.42	0.98	0.99
	itas_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_su_02	1861-2000	0.22	0.60	0.42	0.97	0.99	0.68	13.30	13.97	0.63	0.42	0.98	0.99
GISS ER	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_00	1880-2003	0.35	0.39	0.45	0.99	0.99	0.85	14.01	14.00	0.44	0.50	0.99	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_01	1880-2003	0.28	0.51	0.45	0.99	0.99	0.85	14.00	14.00	0.53	0.50	0.99	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_07	1880-2003	0.45	0.41	0.43	0.99	0.99	0.87	14.02	14.00	0.46	0.50	0.99	0.99
INM CM3.0	itas_inmcm3_0_20c3m_0-360E_-90-90N_n_su	1871-2000	0.14	0.68	0.42	0.99	0.99	0.78	12.75	13.97	0.67	0.42	0.99	0.99
IPSL CM4	itas_ipsl_cm4_20c3m_0-360E_-90-90N_n_su	1860-2000	0.12	0.37	0.43	0.96	0.99	0.79	13.08	13.97	0.37	0.42	0.97	0.99
MIROC3.2 (medres)	itas_miroc3_2_medres_20c3m_0-360E_-90-90N_n_su_00	1850-2000	0.29	0.38	0.40	0.98	0.98	0.79	13.37	13.97	0.41	0.42	0.99	0.99
	itas_miroc3_2_medres_20c3m_0-360E_-90-90N_n_su_01	1850-2000	0.31	0.48	0.39	0.99	0.98	0.76	13.41	13.97	0.49	0.42	0.99	0.99
	itas_miroc3_2_medres_20c3m_0-360E_-90-90N_n_su_02	1850-2000	0.28	0.48	0.41	0.99	0.98	0.70	13.45	13.97	0.49	0.42	0.99	0.99
MRI CGCM 2.3.2	itas_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_su_01	1851-2000	0.04	0.50	0.45	0.98	0.99	0.79	12.82	13.97	0.46	0.42	0.99	0.99
UKMO HadGEM1	itas_ukmo_hadgem1_20c3m_0-360E_-90-90N_n_su_00	1860-1999	0.37	0.36	0.38	0.98	0.98	0.78	12.63	13.97	0.39	0.41	0.99	0.99

Source: climexp.knmi.nl

Table B.2. Maximum likelihood estimates for the parameters of the bivariate HKp for the NOAA annual global land and ocean temperature anomalies.

GCM		Time	Simultaneous MLE					Separate MLE				
			ρ	σ_1	σ_2	H_1	H_2	ρ	σ_1	σ_2	H_1	H_2
BCC CM1	itas_bcc_cm1_20c3m_0-360E_-90-90N_n_su_00	1871-2003	0.36	0.15	0.41	0.88	0.98	0.85	0.20	0.52	0.95	0.99
	itas_bcc_cm1_20c3m_0-360E_-90-90N_n_su_01	1871-2003	0.41	0.14	0.38	0.88	0.98	0.90	0.20	0.52	0.96	0.99
BCCR BCM2.0	itas_bccr_bcm2_0_20c3m_0-360E_-90-90N_n_su	1850-1999	-0.06	0.26	0.50	0.97	0.99	0.50	0.22	0.45	0.97	0.99
CCSM3.0	itas_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_su_02	1870-1999	0.29	0.72	0.36	0.99	0.98	0.86	0.76	0.45	0.995	0.99
	itas_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_su_05	1870-1999	0.31	0.42	0.40	0.98	0.98	0.84	0.49	0.45	0.99	0.99
CGCM3.1 (T63)	itas_cccma_cgcm3_1_t63_20c3m_0-360E_-90-90N_n_su	1850-2000	0.24	0.85	0.34	0.995	0.98	0.85	0.92	0.45	0.995	0.99
CNRM CM3	itas_cnrm_cm3_20c3m_0-360E_-90-90N_n_su	1860-1999	0.09	0.49	0.46	0.96	0.99	0.79	0.49	0.45	0.97	0.99
CSIRO Mk3.5	itas_csiro_mk3_5_20c3m_0-360E_-90-90N_n_su_00	1871-2000	0.00	0.62	0.50	0.98	0.99	0.69	0.56	0.45	0.98	0.99
	itas_csiro_mk3_5_20c3m_0-360E_-90-90N_n_su_01	1871-2000	0.09	0.77	0.47	0.99	0.99	0.74	0.73	0.45	0.99	0.99
ECHAM5 MPI-OM	itas_mpi_echam5_20c3m_0-360E_-90-90N_n_su_03	1860-2000	0.07	0.28	0.48	0.89	0.99	0.55	0.27	0.45	0.91	0.99
ECHO G	itas_miub_echo_g_20c3m_0-360E_-90-90N_n_su_00	1860-2000	0.24	0.46	0.44	0.98	0.99	0.79	0.49	0.45	0.99	0.99
	itas_miub_echo_g_20c3m_0-360E_-90-90N_n_su_02	1860-2000	0.28	0.52	0.42	0.99	0.99	0.78	0.55	0.45	0.99	0.99
	itas_miub_echo_g_20c3m_0-360E_-90-90N_n_su_03	1860-2000	0.17	0.36	0.46	0.98	0.99	0.75	0.38	0.45	0.98	0.99
FGOALS g1.0	itas_iap_fgoals1_0_g_20c3m_0-360E_-90-90N_n_su_00	1850-1999	0.04	0.27	0.47	0.71	0.99	0.28	0.25	0.45	0.72	0.99
	itas_iap_fgoals1_0_g_20c3m_0-360E_-90-90N_n_su_01	1850-1999	0.02	0.30	0.48	0.750	0.99	0.50	0.28	0.45	0.76	0.99
	itas_iap_fgoals1_0_g_20c3m_0-360E_-90-90N_n_su_02	1850-1999	0.02	0.27	0.48	0.69	0.99	0.33	0.25	0.45	0.69	0.99
GFDL CM2.1	itas_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_su_00	1861-2000	0.11	0.50	0.46	0.95	0.99	0.73	0.52	0.45	0.96	0.99
	itas_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_su_01	1861-2000	0.15	0.61	0.46	0.98	0.99	0.70	0.61	0.45	0.98	0.99
GISS ER	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_00	1880-2003	0.36	0.39	0.47	0.99	0.99	0.85	0.44	0.52	0.99	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_01	1880-2003	0.29	0.51	0.48	0.99	0.99	0.85	0.53	0.52	0.99	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_05	1880-2003	0.35	0.50	0.47	0.99	0.99	0.82	0.52	0.52	0.99	0.99
INM CM3.0	itas_inmcm3_0_20c3m_0-360E_-90-90N_n_su	1871-2000	0.14	0.68	0.46	0.99	0.99	0.78	0.67	0.45	0.99	0.99
IPSL CM4	itas_ipsl_cm4_20c3m_0-360E_-90-90N_n_su	1860-2000	0.11	0.37	0.46	0.96	0.99	0.78	0.37	0.45	0.97	0.99
MIROC3.2 (medres)	itas_miroc3_2_medres_20c3m_0-360E_-90-90N_n_su_02	1850-2000	0.24	0.49	0.45	0.99	0.99	0.68	0.49	0.45	0.99	0.99
MRI CGCM 2.3.2	itas_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_su_01	1851-2000	0.05	0.49	0.48	0.98	0.99	0.78	0.46	0.45	0.99	0.99
UKMO HadCM3	itas_ukmo_hadcm3_20c3m_0-360E_-90-90N_n_su_00	1860-1999	0.07	0.33	0.47	0.96	0.99	0.62	0.32	0.45	0.96	0.99
UKMO HadGEM1	itas_ukmo_hadgem1_20c3m_0-360E_-90-90N_n_su_00	1860-1999	0.38	0.36	0.42	0.98	0.99	0.77	0.39	0.45	0.99	0.99

Source: climexp.knmi.nl

Table B.3. Maximum likelihood estimates for the parameters of the bivariate HKp for the CRU combined land [CRUTEM4] and marine temperature anomalies.

GCM		Time	Simultaneous MLE					Separate MLE				
			ρ	σ_1	σ_2	H_1	H_2	ρ	σ_1	σ_2	H_1	H_2
BCC CM1	itas_bcc_cm1_20c3m_0-360E_-90-90N_n_su_00	1871-2003	0.51	0.16	0.33	0.89	0.965	0.87	0.20	0.45	0.95	0.99
	itas_bcc_cm1_20c3m_0-360E_-90-90N_n_su_01	1871-2003	0.51	0.15	0.31	0.89	0.96	0.89	0.19	0.45	0.95	0.99
BCCR BCM2.0	itas_bccr_bcm2_0_20c3m_0-360E_-90-90N_n_su	1850-1999	-0.04	0.29	0.44	0.97	0.98	0.43	0.23	0.35	0.97	0.98
CCSM3.0	itas_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_su_01	1870-1999	0.23	0.59	0.32	0.99	0.97	0.83	0.60	0.37	0.99	0.98
	itas_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_su_02	1870-1999	0.20	0.73	0.28	0.99	0.95	0.85	0.76	0.37	0.995	0.98
	itas_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_su_03	1870-1999	0.21	0.52	0.35	0.99	0.97	0.77	0.51	0.37	0.99	0.98
	itas_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_su_04	1870-1999	0.20	0.66	0.34	0.99	0.97	0.75	0.65	0.37	0.99	0.98
	itas_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_su_05	1870-1999	0.27	0.45	0.32	0.98	0.97	0.81	0.48	0.37	0.99	0.98
CGCM3.1 (T63)	itas_cccma_cgcm3_1_t63_20c3m_0-360E_-90-90N_n_su	1850-2000	0.12	0.99	0.33	0.995	0.96	0.79	0.93	0.35	0.995	0.98
CNRM CM3	itas_cnrm_cm3_20c3m_0-360E_-90-90N_n_su	1860-1999	0.09	0.57	0.38	0.97	0.97	0.76	0.53	0.35	0.97	0.98
CSIRO Mk3.5	itas_csiro_mk3_5_20c3m_0-360E_-90-90N_n_su_00	1871-2000	0.01	0.69	0.42	0.99	0.98	0.64	0.61	0.37	0.99	0.98
	itas_csiro_mk3_5_20c3m_0-360E_-90-90N_n_su_01	1871-2000	0.07	0.81	0.40	0.99	0.98	0.69	0.73	0.37	0.99	0.98
	itas_csiro_mk3_5_20c3m_0-360E_-90-90N_n_su_02	1871-2000	0.14	0.61	0.39	0.99	0.98	0.70	0.57	0.37	0.99	0.98
ECHAM5 MPI-OM	itas_mpi_echam5_20c3m_0-360E_-90-90N_n_su_03	1860-2000	0.07	0.30	0.41	0.89	0.98	0.51	0.26	0.36	0.90	0.98
FGOALS g1.0	itas_iap_fgoals1_0_g_20c3m_0-360E_-90-90N_n_su_00	1850-1999	-0.01	0.33	0.43	0.78	0.98	-0.01	0.27	0.35	0.78	0.98
	itas_iap_fgoals1_0_g_20c3m_0-360E_-90-90N_n_su_01	1850-1999	-0.02	0.38	0.44	0.82	0.98	0.14	0.30	0.35	0.81	0.98
	itas_iap_fgoals1_0_g_20c3m_0-360E_-90-90N_n_su_02	1850-1999	-0.01	0.31	0.44	0.72	0.98	0.10	0.25	0.35	0.72	0.98
GFDL CM2.1	itas_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_su_00	1861-2000	0.08	0.55	0.40	0.96	0.98	0.73	0.50	0.36	0.96	0.98
	itas_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_su_01	1861-2000	0.09	0.68	0.39	0.98	0.98	0.71	0.62	0.36	0.98	0.98
	itas_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_su_02	1861-2000	0.17	0.65	0.38	0.98	0.97	0.71	0.63	0.36	0.98	0.98
GISS ER	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_00	1880-2003	0.34	0.39	0.37	0.99	0.98	0.86	0.44	0.45	0.99	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_01	1880-2003	0.24	0.52	0.37	0.99	0.98	0.86	0.53	0.45	0.99	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_03	1880-2003	0.28	0.53	0.34	0.99	0.97	0.88	0.57	0.45	0.995	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_04	1880-2003	0.32	0.46	0.35	0.99	0.97	0.86	0.50	0.45	0.99	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_06	1880-2003	0.28	0.49	0.40	0.99	0.98	0.84	0.51	0.45	0.99	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_07	1880-2003	0.40	0.42	0.37	0.99	0.98	0.85	0.46	0.45	0.99	0.99
	itas_giss_model_e_r_20c3m_0-360E_-90-90N_n_su_08	1880-2003	0.27	0.43	0.38	0.99	0.98	0.84	0.46	0.45	0.99	0.99
INGV ECHAM4	itas_ingv_echam4_20c3m_0-360E_-90-90N_n_su	1870-2000	0.19	0.62	0.36	0.99	0.97	0.76	0.58	0.37	0.99	0.98
INM CM3.0	itas_inmcm3_0_20c3m_0-360E_-90-90N_n_su	1871-2000	0.11	0.70	0.39	0.99	0.98	0.75	0.65	0.37	0.99	0.98
IPSL CM4	itas_ipsl_cm4_20c3m_0-360E_-90-90N_n_su	1860-2000	0.11	0.39	0.39	0.96	0.98	0.75	0.35	0.36	0.97	0.98
MRI CGCM 2.3.2	itas_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_su_00	1851-2000	0.15	0.60	0.37	0.99	0.97	0.78	0.54	0.36	0.99	0.98
	itas_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_su_01	1851-2000	0.074	0.51	0.41	0.98	0.98	0.77	0.44	0.36	0.98	0.98

GCM		Time	Simultaneous MLE					Separate MLE				
			ρ	σ_1	σ_2	H_1	H_2	ρ	σ_1	σ_2	H_1	H_2
	itas_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_su_02	1851-2000	0.20	0.69	0.36	0.99	0.97	0.81	0.64	0.36	0.991	0.98
	itas_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_su_03	1851-2000	0.30	0.59	0.31	0.99	0.96	0.84	0.58	0.36	0.99	0.98
	itas_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_su_04	1851-2000	0.29	0.57	0.32	0.99	0.96	0.83	0.57	0.36	0.99	0.98
PCM	itas_ncar_pcm1_20c3m_0-360E_-90-90N_n_su_00	1890-1999	0.23	0.41	0.34	0.98	0.97	0.78	0.46	0.39	0.98	0.98
	itas_ncar_pcm1_20c3m_0-360E_-90-90N_n_su_01	1890-1999	0.29	0.30	0.32	0.96	0.97	0.79	0.37	0.39	0.98	0.98
	itas_ncar_pcm1_20c3m_0-360E_-90-90N_n_su_02	1890-1999	0.23	0.40	0.34	0.98	0.97	0.78	0.45	0.39	0.99	0.98
	itas_ncar_pcm1_20c3m_0-360E_-90-90N_n_su_03	1890-1999	0.30	0.41	0.32	0.98	0.97	0.81	0.49	0.39	0.98	0.98
UKMO HadCM3	itas_ukmo_hadcm3_20c3m_0-360E_-90-90N_n_su_00	1860-1999	0.06	0.36	0.41	0.96	0.978	0.57	0.31	0.35	0.96	0.98
UKMO HadGEM1	itas_ukmo_hadgem1_20c3m_0-360E_-90-90N_n_su_00	1860-1999	0.26	0.49	0.35	0.99	0.97	0.73	0.46	0.35	0.99	0.98

Source: climexp.knmi.nl

Table B.4. Maximum likelihood estimates for the parameters of the bivariate HKp for the CRU precipitation over land areas.

GCM		Time	Simultaneous MLE					Separate MLE						
			ρ	σ_1	σ_2	H_1	H_2	ρ	μ_1	μ_2	σ_1	σ_2	H_1	H_2
CCSM3.0	ipr_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_5lan_su_00	1870-1999	0.02	11.35	161.75	0.69	0.99	0.23	756.00	1082.68	11.44	160.35	0.69	0.99
	ipr_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_5lan_su_01	1870-1999	-0.02	10.20	160.86	0.73	0.99	-0.05	756.56	1082.68	10.29	160.35	0.74	0.99
	ipr_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_5lan_su_02	1870-1999	-0.02	12.34	164.05	0.71	0.99	0.15	753.87	1082.68	12.34	160.35	0.71	0.99
	ipr_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_5lan_su_03	1870-1999	0.00	12.25	162.61	0.76	0.99	0.05	756.33	1082.68	12.31	160.35	0.76	0.99
	ipr_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_5lan_su_04	1870-1999	-0.02	10.03	162.80	0.72	0.99	-0.01	753.96	1082.68	10.08	160.35	0.72	0.99
	ipr_ncar_ccsm3_0_20c3m_0-360E_-90-90N_n_5lan_su_05	1870-1999	0.00	10.19	162.29	0.62	0.99	0.15	756.44	1082.68	10.24	160.35	0.62	0.99
CGCM3.1 (T47)	ipr_cccma_cgcm3_1_20c3m_0-360E_-90-90N_n_5lan_su_00	1850-2000	-0.03	9.80	160.99	0.70	0.99	-0.09	685.01	1082.68	9.87	160.35	0.70	0.99
	ipr_cccma_cgcm3_1_20c3m_0-360E_-90-90N_n_5lan_su_01	1850-2000	0.02	9.88	163.61	0.62	0.99	0.08	686.24	1082.68	9.94	160.35	0.63	0.99
	ipr_cccma_cgcm3_1_20c3m_0-360E_-90-90N_n_5lan_su_02	1850-2000	-0.04	11.33	161.11	0.78	0.99	-0.03	687.32	1082.68	11.41	160.35	0.78	0.99
	ipr_cccma_cgcm3_1_20c3m_0-360E_-90-90N_n_5lan_su_03	1850-2000	0.03	12.15	162.83	0.77	0.99	0.09	686.65	1082.68	12.22	160.35	0.77	0.99
	ipr_cccma_cgcm3_1_20c3m_0-360E_-90-90N_n_5lan_su_04	1850-2000	-0.01	11.09	161.99	0.76	0.99	-0.05	687.51	1082.68	11.14	160.35	0.76	0.99
CGCM3.1 (T63)	ipr_cccma_cgcm3_1_t63_20c3m_0-360E_-90-90N_n_5lan_su	1850-2000	-0.02	11.42	162.18	0.62	0.99	-0.07	698.15	1082.68	11.48	160.35	0.62	0.99
CSIRO Mk3.5	ipr_csiro_mk3_5_20c3m_0-360E_-90-90N_n_5lan_su_00	1871-2000	0.01	22.70	162.83	0.62	0.99	-0.03	677.67	1082.68	22.80	160.35	0.62	0.99
	ipr_csiro_mk3_5_20c3m_0-360E_-90-90N_n_5lan_su_01	1871-2000	0.00	22.00	162.51	0.65	0.99	-0.05	673.90	1082.68	22.11	160.35	0.65	0.99
	ipr_csiro_mk3_5_20c3m_0-360E_-90-90N_n_5lan_su_02	1871-2000	-0.03	19.49	162.46	0.65	0.99	-0.08	678.32	1082.68	19.60	160.35	0.65	0.99
ECHAM5 MPI-OM	ipr_mpi_echam5_20c3m_0-360E_-90-90N_n_5lan_su_03	1860-2000	-0.01	11.05	162.67	0.55	0.99	-0.03	678.27	1082.68	11.11	160.35	0.55	0.99
ECHO G	ipr_miub_echo_g_20c3m_0-360E_-90-90N_n_5lan_su_00	1860-2000	-0.03	10.53	164.31	0.61	0.99	0.05	757.54	1082.68	10.58	160.35	0.61	0.99
	ipr_miub_echo_g_20c3m_0-360E_-90-90N_n_5lan_su_01	1860-2000	0.01	10.51	162.46	0.68	0.99	0.10	758.15	1082.68	10.57	160.35	0.68	0.99
	ipr_miub_echo_g_20c3m_0-360E_-90-90N_n_5lan_su_02	1860-2000	0.00	10.96	162.05	0.65	0.99	0.16	758.50	1082.68	11.02	160.35	0.65	0.99
GFDL CM2.1	ipr_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_5lan_su_00	1861-2000	-0.03	28.72	161.12	0.49	0.99	-0.10	749.37	1082.68	28.86	160.35	0.49	0.99
	ipr_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_5lan_su_01	1861-2000	-0.01	25.85	162.76	0.48	0.99	0.02	747.31	1082.68	25.98	160.35	0.48	0.99
	ipr_gfdl_cm2_1_20c3m_0-360E_-90-90N_n_5lan_su_02	1861-2000	-0.02	25.63	162.34	0.61	0.99	-0.07	750.61	1082.68	25.77	160.35	0.61	0.99
GISS ER	ipr_giss_model_e_r_20c3m_0-360E_-90-90N_n_5lan_su_01	1880-2003	0.01	9.88	161.94	0.77	0.99	0.04	878.52	1082.68	9.93	160.35	0.77	0.99
	ipr_giss_model_e_r_20c3m_0-360E_-90-90N_n_5lan_su_03	1880-2003	-0.03	8.96	160.95	0.56	0.99	-0.16	880.45	1082.68	9.01	160.35	0.56	0.99
	ipr_giss_model_e_r_20c3m_0-360E_-90-90N_n_5lan_su_04	1880-2003	-0.04	11.10	164.06	0.66	0.99	-0.14	880.03	1082.68	11.14	160.35	0.65	0.99
	ipr_giss_model_e_r_20c3m_0-360E_-90-90N_n_5lan_su_05	1880-2003	-0.02	10.22	162.58	0.67	0.99	-0.11	879.16	1082.68	10.27	160.35	0.67	0.99
	ipr_giss_model_e_r_20c3m_0-360E_-90-90N_n_5lan_su_06	1880-2003	0.01	8.65	163.75	0.64	0.99	-0.11	880.93	1082.68	8.69	160.35	0.64	0.99
	ipr_giss_model_e_r_20c3m_0-360E_-90-90N_n_5lan_su_07	1880-2003	-0.04	9.87	161.64	0.68	0.99	-0.17	879.71	1082.68	9.96	160.35	0.69	0.99
	ipr_giss_model_e_r_20c3m_0-360E_-90-90N_n_5lan_su_08	1880-2003	0.02	10.55	164.42	0.65	0.99	-0.17	880.12	1082.68	10.59	160.35	0.64	0.99
INGV ECHAM4	ipr_ingv_echam4_20c3m_0-360E_-90-90N_n_5lan_su	1870-2000	0.01	10.79	162.22	0.75	0.99	0.10	755.09	1082.68	10.86	160.35	0.75	0.99
INM CM3.0	ipr_inmcm3_0_20c3m_0-360E_-90-90N_n_5lan_su	1871-2000	0.03	12.28	156.60	0.70	0.99	0.43	693.42	1082.68	12.48	160.35	0.72	0.99
IPSL CM4	ipr_ipsl_cm4_20c3m_0-360E_-90-90N_n_5lan_su	1860-2000	-0.01	9.93	163.12	0.60	0.99	0.12	656.51	1082.68	9.98	160.35	0.60	0.99
MIROC3.2 (medres)	ipr_miroc3_2_medres_20c3m_0-360E_-90-90N_n_5lan_su_00	1850-2000	-0.03	19.40	160.50	0.81	0.99	-0.34	807.72	1082.68	19.72	160.35	0.82	0.99
	ipr_miroc3_2_medres_20c3m_0-360E_-90-90N_n_5lan_su_01	1850-2000	0.06	20.02	163.83	0.86	0.99	-0.09	799.97	1082.68	19.97	160.35	0.85	0.99
	ipr_miroc3_2_medres_20c3m_0-360E_-90-90N_n_5lan_su_02	1850-2000	0.11	18.44	164.95	0.84	0.99	0.00	801.56	1082.68	18.53	160.35	0.84	0.99
MRI CGCM 2.3.2	ipr_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_5lan_su_00	1851-2000	0.02	9.92	162.66	0.50	0.99	0.10	710.52	1082.68	9.97	160.35	0.50	0.99

GCM		Time	Simultaneous MLE					Separate MLE						
			ρ	σ_1	σ_2	H_1	H_2	ρ	μ_1	μ_2	σ_1	σ_2	H_1	H_2
	ipr_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_5lan_su_02	1851-2000	-0.03	10.76	162.83	0.65	0.99	-0.04	711.06	1082.68	10.80	160.35	0.64	0.99
	ipr_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_5lan_su_03	1851-2000	-0.02	9.60	162.31	0.59	0.99	-0.02	712.80	1082.68	9.65	160.35	0.59	0.99
	ipr_mri_cgcm2_3_2a_20c3m_0-360E_-90-90N_n_5lan_su_04	1851-2000	-0.02	12.04	160.75	0.57	0.99	-0.14	709.93	1082.68	12.11	160.35	0.57	0.99
PCM	ipr_ncar_pcm1_20c3m_0-360E_-90-90N_n_5lan_su_00	1890-1999	0.05	11.40	163.51	0.57	0.99	0.11	758.02	1082.68	11.45	160.35	0.56	0.99
	ipr_ncar_pcm1_20c3m_0-360E_-90-90N_n_5lan_su_01	1890-1999	-0.01	12.00	163.06	0.57	0.99	0.07	759.18	1082.68	12.06	160.35	0.57	0.99
	ipr_ncar_pcm1_20c3m_0-360E_-90-90N_n_5lan_su_02	1890-1999	0.00	12.50	162.45	0.60	0.99	-0.04	757.91	1082.68	12.56	160.35	0.60	0.99
	ipr_ncar_pcm1_20c3m_0-360E_-90-90N_n_5lan_su_03	1890-1999	0.00	12.05	163.16	0.44	0.99	-0.06	757.87	1082.68	12.11	160.35	0.44	0.99
UKMO HadCM3	ipr_ukmo_hadcm3_20c3m_0-360E_-90-90N_n_5lan_su_00	1860-1999	-0.03	13.07	162.90	0.43	0.99	-0.06	768.84	1082.68	13.14	160.35	0.43	0.99
	ipr_ukmo_hadcm3_20c3m_0-360E_-90-90N_n_5lan_su_01	1860-1999	-0.03	13.46	159.40	0.47	0.99	-0.14	768.47	1082.68	13.52	160.35	0.48	0.99
UKMO HadGEM1	ipr_ukmo_hadgem1_20c3m_0-360E_-90-90N_n_5lan_su_00	1860-1999	-0.04	13.81	159.55	0.72	0.99	-0.26	805.32	1082.68	13.95	160.35	0.73	0.99

Source: climexp.knmi.nl

Acknowledgements

I would like to thank everyone contributing to the completion of this thesis directly or indirectly. Your help was invaluable.

First of all I thank Professor Demetris Koutsoyiannis for offering me the opportunity to work with him. You spent so many hours pointing to me even the most insignificant details. Foremost you helped me to develop my critical thinking. The largest part of this thesis originated from our collaboration and your original ideas. Credit for the quote "*which is the right thing?*" when I had doubts about solving a scientific-personal problem.

I thank Reader Christian Onof for reviewing this thesis, pointing out significant details and for his encouraging words during this research. Unfortunately Associate Professor Matthew Karlaftis, the second member of the advisory committe passed away a few days before completion of this thesis. I appreciate his willingness to join this commitee and I am sorry that I could not present to him a complete thesis timely. I appreciate and thank Assistant Professor Nikolaos Lagaros for accepting to join the advisory committee, and rapidly commenting on the thesis. Dr. Federico Lombardo and five anonumous reviewers of our published papers which support this thesis contributed considerably to the improvement of our work.

The members of the examination committee, Professor Apostolos Burnetas, Professor Alberto Montanari, Associate Professor Harry Pavlopoulos and Assistant Professor Andreas Langousis helped to improve the draft thesis with their comments. I especially thank Professor Apostolos Burnetas for being my first teacher in stochastic processes, which use is important in this thesis.

Olga Kitsou and Katerina Tzouka helped me so much, to deal with the bureaucratic stuff. You saved me a lot of time hanging around from one office to another. The Itia Research group and especially Panayiotis Dimitriadis, Yiannis Markonis and Simon Papalexiou, my fellow Ph.D students were partial hosts in their office for various conversations on scientific themes. I hope we will work together in a future project. I thank Xanthoula Michail for designing a beautiful front cover.

Without the help of my Hellenic Air Force colleagues during periods of pressure, it would not be possible to fulfil all my obligations in the university and in my service.

My parents, my brother, his wife and their baby were patient with me and all my odd demands. They considerably facilitated my everyday life. I hope my parents feel proud for the

accomplishments of decades of study.

Finally I thank Kwan. We met during my visit to NU. Her moral support while accomplishing the most part of this work was invaluable.

About the author

Hristos Tyralis was born in the small city of Serres in North Greece, Macedonia at the 29th April 1979. From a young age he became interested in perfect information, combinatorial games such chess, however playing intuitively and gambling at all phases of the game, but even most interested in other team sports, spending most of his time in football fields or basketball courts, but without neglecting to study once in a while his favourite course, mathematics. Due to the educational system in his last two years he was forced to study mathematics in every detail and loved this exact science with its strict structure.

He decided to join the Air Force Academy, which was organized in a strict, similar to mathematics, manner, to study civil engineering and during the first semester he became interested in statistics. After graduating and while at the same time serving the Hellenic Air Force as an engineer officer, Hristos decided to attend an MSc program in Statistics to renew his interest for the science of mathematics. He then decided to repeat three years of civil engineering studies in the National Technical University of Athens, to be certified for the study and supervision of civil structures.

During his studies he became interested in the course of Stochastic Hydrology which was the only course which demanded a deep theoretical background in statistics. After taking a break from the academia and while continuing working for the Hellenic Air Force in the Air Force Infrastructure Wing, being busy in studying and supervising the maintenance and construction of building structures and airfield pavements, he decided to return, starting his Ph.D research at the Water Resources Department. During his research he improved his handling of statistical tools, but most importantly the communication with the rest of the research group broadened his horizons and sharpened his critical thinking.

In his little spare time Hristos continues to play basketball to preserve his form, as demanded by his service and chess, however reducing the amount of sharp, intuitive gambling moves. In the immediate future he will continue working for the Hellenic Air Force, but his distant future is uncertain, as proposed by the HK Bayesian setting.

Publications

Publications in scientific journals

Tyralis H, Koutsoyiannis D (2011) Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process. Stochastic Environmental Research & Risk Assessment 25(1):21-33. doi:10.1007/s00477-010-0408-x

Tyralis H, Koutsoyiannis D (2014) A Bayesian statistical model for deriving the predictive distribution of hydroclimatic variables. Climate Dynamics 42(11-12):2867-2883. doi:10.1007/s00382-013-1804-y

Tyralis H, Koutsoyiannis D (2015) On the prediction of persistent processes using the output of deterministic models. To be submitted

Tyralis H, Koutsoyiannis D, Kozanis S (2013) An algorithm to construct Monte Carlo confidence intervals for an arbitrary function of probability distribution parameters. Computational Statistics 28(4):1501-1527. doi:10.1007/s00180-012-0364-7

Conference publications and presentations with evaluation of abstract

Filippidou AM, Andrianopoulos A, Argyrakis C, Chomata LE, Dagalaki V, Grigoris X, Kokkoris TS, Nasioka M, Papazoglou KA, Papalexiou SM, Tyralis H, Koutsoyiannis D (2014) Comparison of climate time series produced by General Circulation Models and by observed data on a global scale. European Geosciences Union. European Geosciences Union General Assembly 2014, Geophysical Research Abstracts 16(EGU2014-8529). Vienna. Available online at:http://itia.ntua.gr/en/docinfo/1440/

Kossieris P, Koutsoyiannis D, Onof C, Tyralis H, Efstratiadis A (2012) HyetosR: An R package for temporal stochastic simulation of rainfall at fine time scales. European Geosciences Union. European Geosciences Union General Assembly 2012, Geophysical Research Abstracts 14(EGU2012-11718). Vienna. Available online at:http://itia.ntua.gr/en/docinfo/1200/

Koutsoyiannis D, Kozanis S, Tyralis H (2011) A general Monte Carlo method for the construction of confidence intervals for a function of probability distribution parameters. European Geosciences Union. European Geosciences Union General Assembly 2011, Geophysical Research Abstracts 13(EGU2011-1489). Vienna. Available online at:http://itia.ntua.gr/en/docinfo/1125/

Tyralis H, Koutsoyiannis D (2010) Performance evaluation and interdependence of parameter estimators of the Hurst-Kolmogorov stochastic process. European Geosciences Union. European Geosciences Union General Assembly 2010, Geophysical Research Abstracts 12(EGU2010-10476). Vienna. Available online at:http://itia.ntua.gr/en/docinfo/982/

Tyralis H, Koutsoyiannis D (2012) A Bayesian approach to hydroclimatic prognosis using the Hurst-Kolmogorov stochastic process. European Geosciences Union. European Geosciences Union General Assembly 2012, Geophysical Research Abstracts, Vol. 14. Vienna. Available online at:http://itia.ntua.gr/en/docinfo/1202/

Tyralis H, Koutsoyiannis D (2013) Simultaneous use of observations and deterministic model outputs to forecast persistent stochastic processes. European Geosciences Union, International Association of Hydrological Sciences, International Union of Geodesy and Geophysics. Facets of Uncertainty: 5th EGU Leonardo Conference - Hydrofractals 2013 - STAHY '13. Kos Island, Greece. Available online at:http://itia.ntua.gr/en/docinfo/1401/

Software

Tyralis H (2014) HKprocess, R software for the Hurst-Kolmogorov process. Available online at:http://itia.ntua.gr/en/softinfo/31/