Zeytu Gashaw Asfaw

Inference and Prediction in Non-stationary Stochastic Models: Survival Analysis and Kriging Interpolation

Thesis for the degree of Philosophiae Doctor

Trondheim, August 2014

Norwegian University of Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Mathematical Sciences

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-326-0342-8 (printed ver.)
ISBN 978-82-326-0343-5 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2014:210

Printed by NTNU-trykk

Preface

This thesis is submitted in partial fulfilment of the requirements for the degree of Philosophiae Doctor (PhD) at the Norwegian University of Science and Technology (NTNU). The research is financed by the Norwegian Educational Loan (Quota program) along with the NOMA program, Mathematical and Statistical Modeling (MASTMO) where the latter is administered by Prof. Henning Omre, NTNU and Dr. Ayele Taye, Hawassa University (HU), Ethiopia.

Acknowledgement

I am grateful to all of you who have contributed towards shaping this thesis and where I am today.

First and foremost, I would like to express my special appreciation and thanks to my main supervisor Prof. Henning Omre, you have been a tremendous mentor for me. I would like to thank you for your gentle encouragement, excellent and well scheduled guidance, constructive suggestions added for the enrichment of the thesis, and for providing necessary logistics for my PhD study. Your advice on both research as well as on my academic career have been prized.

I am deeply grateful to my co-supervisor Prof. Bo Lindqvist for his encouragement, guidance, sharing experience and support, it has been greatly appreciated. He has also provided insightful discussions about the research and his comments and positive energy has been invaluable. He is a fanciful supervisor and his guidance has been much appreciated and will be missed.

I gratefully thank you Prof. Henning Omre and Dr.Ayele Taye for financial support under NOMA-MASTMO project along with the Norwegian Educational Loan (Quota program). I would like to thank all administrative staff members of Department of Mathematical Sciences and International Relation office at NTNU who have been kind enough to advise and help in their respective roles.

A very special thanks to my family. Words cannot express how grateful

I am to my mother, W/ro Zeleka Meshesha, for all of the sacrifices that you have made on my behalf. Your prayer for me was what sustained me thus far. Moreover, my hearted thanks are reserved for my brothers and sisters for their moral support and encouragement throughout my life, and especially these years that I was far away from them.

I would also like to thank all of my friends who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. I am especially grateful to Zerihun Kinfe, Dr. Tesfa Yigerem, Dr. Maru Alamirew, Asfafaw Haileselasie and their lovely families that made my time at Trondheim enjoyable at large part. My warmest thanks go to Zerihun Kinfe, Mamuye Busrie, Kiros Gebrearegawi, Dejene Yigzaw, Negalign Legesse, Assefa Derbew and Gizat Dereb for your unreseved help during the family reunification process and I have very much enjoyed the friendly environment created by you. I also extend my sincere thanks to fellow PhD students David Lindberg and Petter Arnesen for your fruitful academic discussion during course work and unrestricted help in Matlab.

I would like to thank the Trondheim Sunday School of the Ethiopian Orthodox Tewhaido church and its members for supporting me to nourish spiritually throughout my study.

Finally, I would like express my gratitude and great thanks to my beloved wife Hana Yitagesu, who I spent sleepless nights with and was always my backbone in the moments when there was no one to answer my queries. Moreover, I heartedly thank you for your golden gift-give birth to our wonderful Daughter named Natania. She has made me more responsible and energetic to pull-off my PhD study. Thank you very much both of you!

Last but not least, I would like to give honor to the Almighty GOD, through whom everything is possible.

Zeytu Gashaw Asfaw
Trondheim, Norway, August 2014

Thesis Outline

The thesis consists of four parts, Part I: Introduction, Part II: Report I, Part III: Report II and Part IV: Appendix. The latter part, the Appendix, contains three papers (Paper I.A, Paper I.B and Paper II.A). The two first papers are based on Report I while the last paper is based on Report II. The reports could be read independently of each other but Paper I.B may be considered as a follow-up to Paper I.A.

Introduction

Report I: A Simulation Study of Statistical Inference in Non-homogeneous Poisson Processes with Emphasis on Frailty and Dynamic Behavior
Zeytu Gashaw Asfaw
Preprint Series in Statistics no.6, 2013, Department of Mathematical Sciences, NTNU, Trondheim, Norway. Partly presented at the 25th Nordic conference in Mathematical statistics (NORDSTAT), June 2014, Turku University, Finland.

Report II: Localized/Shrinkage Kriging Predictors
Zeytu Gashaw Asfaw and Henning Omre
Preprint Series in Statistics no.1, 2014, Department of Mathematical Sciences, NTNU, Trondheim, Norway.

Appendix:

Paper I.A: Unobserved heterogeneity in the power law nonhomogeneous Poisson process
Zeytu Gashaw Asfaw and Bo Henry Lindqvist
Submitted for publication (In revision)

Paper I.B: Extending minimal repair models for repairable systems: A comparision of dynamic and heterogeneous extensions of a nonhomogeneous Poisson process
Zeytu Gashaw Asfaw and Bo Henry Lindqvist
Submitted for publication

Paper II.A: Localized/Shrinkage Kriging Predictors
Zeytu Gashaw Asfaw and Henning Omre
Submitted for publication

Part I:

Introduction

1 Introduction

Survival analysis and kriging interpolation is widely studied in statistical litterature and frequently applied in biology and medicine, technological systems and earth science mappings.

Survival analysis have been utilized since the first life table, which was the oldest methods for analyzing survival/failure time data, was disclosed by John Graunt in 1662, see Kreager [6]. The name survival analysis may indicate the analysis of actual survival, death rates or mortality. But, pertaining to the analysis of the time of occurrence of events of various types, it has a wider meaning that covers reliability theory/reliability analysis in engineering, duration analysis/duration modeling in economics and event history analysis in sociology.

Kriging as a special term in geostatistics was introduced by Matheron in 1963, to acknowledge the M.Sc Thesis of Daniel G. Krige. The philosophical ideas behind G.Matheron's work is presented in Matheron [8]. The adoption of geostatistics and kriging to applications was initially made in the celebrated book Journel and Huijbregts [5]. Currently kriging is used as a term for best linear unbiased predictors in mainstream statistical litterature. The methodology is being used in a large variety of applications, for example: natural resource explorations, climate modelling, medical image analysis and econometrics.

1.1 Non-stationary stochastic models

Stochastic models as temporal processes or as spatial fields are discussed in this thesis. We coin them survival analysis and kriging interpolation models, respectively. Non-stationary versions of these models are of special interest, and both model formulation and model parameter inference are discussed.

Survival analysis and kriging interpolation models are presented in the next Sections. These two model types defines the backbone of the two major reports in the Thesis. The exposure of the models emphasizes the similarity in model construction in spite one is in the temporal and one is in the spatial domain. We also outline model parameter inference techniques under varying

model assumptions.

1.1.1 Survival Analysis Model

A stochastic process $\{N(t); t \in T \subset \Re_+\}$ with temporal reference variable $t \in T$, is a counting process if $N(t) \in \mathbf{N}_+$ and $N(s) \leq N(t)$, for $s < t$, see Rausand and Høyland [9].

A counting process is a non-stationary Poisson process (NSPP) with rate function $\{w(t); t \in T\}$, if:

1. N(0)=0.
2. $Pr\{N(t + \Delta t) - N(t) = 1\} = w(t)\Delta t + o(\Delta t)$, for all t where $\frac{O(\Delta t)}{\Delta t} \to 0$ as $\Delta t \to 0$ and $N(t)$ is the number of events occurring within $(0, t]$
3. $Pr\{N(t + \Delta t) - N(t) \geq 2\} = o(\Delta t)$, which means that the system will not experience more than one event at the same time.

An NSPP model is fully characterized by the intensity function $\{w(t); t \in T\}$, commonly denoted by Rate of ocurrence of failures (ROCOF). The associated cumulative ROCOF (CROCOF) is defined as:

$$W(t) = \int_0^t w(s)ds; t \in T.$$

The number of events in a time interval $(0, t]$, $N(t)$, is Poisson distributed with parameter $W(t)$, for any $t \in T$:

$$N(t) \sim Pois[W(t)]$$

The probability of n events in the interval $(0, t]$ is then:

$$f(n) = \frac{[W(t)]^n}{n!} exp\{-W(t)\}; n = 0, 1, ...$$

The expectation and variance are defined by the model parameters for NSPPs:

$$E[N(t)] = W(t)$$

$$Var[N(t)] = W(t)$$

The counting process $\{N(t); t \in T \subset \Re_+\}$ is said to be a stationary Poisson process (SPP) having constant and time independent rate function:

$$w(t) = \lambda; t \in T$$

$$W(t) = \lambda t; t \in T$$

Parametric Model

Several parametric models illustrate the ROCOF of an NSPP but we use the most celeberated parametrization of the NSPP, namely the power law model. One reason for its popularity is that the ROCOF as a function of t is of the same form as the hazard rate of a Weibull distribution. Hence the time to first event of the power law NSPP is Weibull distributed. Because of this, the power law model is sometimes denoted the Weibull process.

The CROCOF of the power law is given by, see Rausand and Høyland [9]:

$$W(t; \lambda, \beta) = \lambda t^{\beta}, \qquad \text{for } \lambda > 0,\ \beta > 0;\ t \in T$$

The parameter β in the power law model gives information about the system as follows: if $0 < \beta < 1$, then the system is improving, if $\beta > 1$, then the system is deteriorating and if $\beta = 1$, then the model is reduces to an SPP.

Hierarchical Model

The hierarchical model is phrased in a Bayesian setting with an underlying random model parameter a, as:

$$[N(t) \mid a] \sim Pois[a\, W(t; \lambda, \beta)]$$

with prior model which is Gamma distributed:

$$a \sim Gam[\delta_a, \frac{1}{\delta_a}]$$

where, $\delta_a \in \Re_+$ is a coupled location/scale parameter. This entails that $E[a] = 1$ and $Var[a] = \delta_a$. The corresponding pdf is:

$$f(a) = \left[\Gamma(\tfrac{1}{\delta_a})\delta_a^{\frac{1}{\delta_a}}\right]^{-1} \left[a^{\frac{1}{\delta_a}-1} exp^{\frac{-a}{\delta_a}}\right],\ a > 0$$

The associated marginal pdf for $N(t)$ is then,

$$N(t) \sim NegBin\left[\frac{1}{\delta_a}, \frac{1}{\delta_a W(t;\lambda,\beta)+1}\right]$$

The negative binomial distribution is also known as the Gamma-Poisson (mixture) distribution.

Dynamic Model

The dynamic model is defined by the conditional intensity function, see Babykina and Couallier [1] and Le Gat [7]:

$$w_d(t;\lambda,\beta,\gamma) = (1+\gamma N(t-))w(t)$$

where d stands for dynamic, $w(t) = \lambda\beta t^{\beta-1}$ is the *baseline* intensity function and $N(t-)$ denotes the number of events experienced in the time interval from $(0,t]$.

The interpretation of $\{w_d(t); t \in T\}$ as a conditional intensity function is as follows,

$$\text{P(event in } (t,t+h) \mid N(t-)=n) \approx (1+\gamma n)w(t)h$$

for small h.

As shown by Le Gat [7], the marginal distribution of $N(t)$ is negative binomial:

$$N(t) \sim NegBin\left[\frac{1}{\gamma}, exp\left\{-\gamma W(t;\lambda,\beta)\right\}\right].$$

1.1.2 Kriging Interpolation Model

The random field (RF) $\{R(x); x \in \mathbf{D} \subset \Re^m\}$ with spatial reference variable $x \in \mathbf{D}$ and $R(x) \in \Re$, is a preferable model for continuous, or almost continuous, spatial variables due to its simplicity in inferences and analytical tractability.

A RF $\{R(x); x \in \mathbf{D} \subset \Re^m\}$ is denoted a Gaussian RF (GRF) with spatial

expectation $\{\mu(x); x \in \mathbf{D}\}$, spatial variance $\{\sigma^2(x); x \in \mathbf{D}\}$ and spatial correlation function $\left\{\rho(x^{'}, x^{''}); x^{'}, x^{''} \in \mathbf{D}^2\right\}$, see Cressie [3] if:

$$\mathbf{R} \sim Gauss[\boldsymbol{\mu}, \Gamma\Omega\Gamma^T]$$

for $\mathbf{R} = [R(x_1), ..., R(x_n)]^T$, $\forall$ conf $(x_1, \cdots, x_n) \in \mathbf{D}^n$ and $\forall n \geq 1$.

Where $\boldsymbol{\mu} = [\mu(x_1), ..., \mu(x_n)]^T$ is a $[n \times 1]$-vector, Γ is a diagonal $[n \times n]$-matrix with diagonal elements $[\sigma(x_1), ..., \sigma(x_n)]$ and Ω is a $[n \times n]$-matrix with elements $\rho(x_i, x_j); i, j \in 1, ..., n$.

Hence the GRF is fully characterized by $\{\mu(x); x \in \mathbf{D}\}$, $\{\sigma^2(x); x \in \mathbf{D}\}$ and $\left\{\rho(x^{'}, x^{''}); x^{'}, x^{''} \in \mathbf{D}^2\right\}$. We consider the two former to be unknown, while the latter is considered to be a known positive definite function, which will not be further discussed here. The corresponding pdf can be written as:

$$f(\mathbf{r}) = (2\pi)^{-\frac{n}{2}} \left|\Gamma\Omega\Gamma^T\right|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}(\mathbf{r} - \boldsymbol{\mu})^T[\Gamma\Omega\Gamma^T]^{-1}(\mathbf{r} - \boldsymbol{\mu})\right\},$$

The expectation and variance are defined by the model parameters for GRFs:

$$E\left[R(x)\right] = \mu(x)$$
$$Var\left[R(x)\right] = \sigma^2(x)$$

The GRF $\{R(x); x \in \mathbf{D} \subset \Re^m\}$ is said to be stationary if the model parameters are:

$$\mu(x) = \mu; \quad x \in \mathbf{D}$$
$$\sigma^2(x) = \sigma^2; \quad x \in \mathbf{D}$$
$$\rho(x^{'}, x^{''}) = \rho(x^{'} - x^{''}); x^{'}, x^{''} \in \mathbf{D}^2$$

Parametric Model

The so called universal kriging model is defined by, see Chiles and Delfiner [2]:

$$\mu(x; \alpha_1, ..., \alpha_L) = \sum_{l=1}^{L} \alpha_l g_l(x); \quad x \in \mathbf{D}$$

$$\sigma^2(x) = \sigma^2 \qquad \text{for } \sigma^2 \geq 0; \quad x \in \mathbf{D}$$

where, $\alpha = [\alpha_1, ..., \alpha_L]^T$ and σ^2 are unknown and must be assessed from the observed values while $\{g_l(x); x \in \mathbf{D}\}$; $l = 1, ..., L$ need to be known functions. This parametric model can be interpreted as a regression model in the known regressor surfaces $\{g_l(x); x \in \mathbf{D}\}$; $l = 1, ..., L$. The spatial correlation function $\left\{\rho(x^{'} - x^{''}); x^{'}, x^{''} \in \mathbf{D}^2\right\}$ is considered to be known.

Hierarchical Model

In the hierarchical Bayesian representation we let the model parameters in the stationary GRF μ and σ^2 be represented by random variables m and s^2. By conditioning on $[m, s^2]$, the GRF is fully specified:

$$[\mathbf{R} \mid m, s^2] \sim Gauss[m\boldsymbol{i}_n, s^2\Omega]$$

where, m and s^2 are univariate random variables and $\boldsymbol{i}_n$ is a unit $[n \times 1]$-vector. The stationary correlation function is considered to be known.

Assume the following prior model for $[m\ ,\ s^2]$:

$$[m \mid s^2] \sim Gauss[\mu_m, \tau_m s^2]$$

$$s^2 \sim InvGam[\xi_s, \gamma_s]$$

with $\mu_m \in \Re$, $\tau_m \in \Re_+$ and $InvGam[\xi_s, \gamma_s]$ representing the inverse gamma pdf:

$$f(s^2) = [\Gamma(\xi_s)]^{-1}\gamma_s^{\xi_s}[s^2]^{-(\xi_s+1)}exp\left\{-\gamma_s[s^2]^{-1}\right\},\ s^2 > 0.$$

where $\Gamma(x)$ is the gamma function, $\xi_s \in \Re_+$ is a shape parameter and $\gamma_s \in \Re_+$ is a scale parameter.

Note that the prior models for $[m, s^2]$ are conjugate models with respect to multi-Gaussian models, hence the posterior models conditioned on a set of observations belong to the same pdf classes.

The associated marginal pdf for $\mathbf{R}$ is then, see Røislien and Omre [10]:

$$\mathbf{R} \sim T - dist\left[\mu_m i_n, \Gamma\Omega\Gamma^T, \nu\right]$$

where, $\nu \in \Re_+$ is the degrees of freedom which is defined by $\nu = 2\xi_s$.

Dynamic Model

Since there is no ordering in the spatial reference variable $x \in \mathbf{D}$, it is not easy to see how a spatial dynamic model could be defined.

1.2 Inference of model parameters

In this section we discuss the survival analysis and kriging interpolation models separately, although these have many similarities in the inference approaches.

Survival Analysis Model

Inference is based on the observations in time $(0, \tau]$ for m independent systems with n_j events for each system $j = 1, ..., m$ and $0 < t_{ij} < \tau; i = 1, ..., n_j, j = 1, ..., m$ being the times to events.

The stationary model parameter λ is assessed by maximum likelihod:

$$\hat{\lambda} = \frac{\sum_{i=1}^{m} n_j}{m\tau}$$

In the parametric model, the actual set of model parameters are $[\lambda, \beta]$ and their maximum likelihood estimates are:

$$\hat{\lambda} = \frac{\sum_{i=1}^{m} n_j}{m[\tau^{\hat{\beta}}]}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{m} n_j}{\log(\tau) \sum_{i=1}^{m} n_j - \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij}}$$

The latter equation gives an explicit solution for $\hat{\beta}$ which can afterwards be substituted in the former so that we have both estimates.

In the hierachical model, the actual set of model parameters are $[\lambda, \beta, \delta]$.

The λ and β parameters have the same expression as the parametric solution but numerical solution is needed to determine δ. The maximum likelihood solution require the following equation to be solved:

$$\left[\frac{-1}{\delta_a^2}\right]\left\{\sum_{j=1}^{m}\psi(n_j+\frac{1}{\delta_a})-m\psi(\frac{1}{\delta_a})-m\log(\delta_a)+m\right\}$$
$$-\left[\frac{-1}{\delta_a^2}\right]\left\{m\log\left[\hat{\lambda}T^{\hat{\beta}}+\frac{1}{\delta_a}\right]-\frac{n+\frac{m}{\delta_a}}{\hat{\lambda}T^{\hat{\beta}}+\frac{1}{\delta_a}}\right\}=0$$

This defines the estimate $\hat{\delta}_a$, and consequently $\widehat{Var}[a]=\hat{\delta}_a$.

In the dynamic model, the actual set of model parameters are $[\lambda,\beta,\gamma]$. Numerical solution is also recommended here.

Kriging Interpolation Model

The inference is based on the observations in locations $(x_1^o,...,x_{n_o}^o)\in\mathbf{D}^{n_o}$, hence $\mathbf{r}_o=(r(x_1^o),...,r(x_{n_o}^o))$ in one realization $\{r(x);x\in\mathbf{D}\}$.

In a stationary GRF model, the actual set of model parameters are $[\mu,\sigma^2]$ and their maximum likelihood estimates are:

$$\hat{\mu}=[\boldsymbol{i}_{n_o}^T\Omega_{oo}^{-1}\mathbf{r}_o][\boldsymbol{i}_{n_o}^T\Omega_{oo}^{-1}\boldsymbol{i}_{n_o}]^{-1}$$

$$\hat{\sigma^2}=\tfrac{1}{n_o}(\mathbf{r}_o-\hat{\mu}\boldsymbol{i}_{n_o})^T\Omega_{oo}^{-1}(\mathbf{r}_o-\hat{\mu}\boldsymbol{i}_{n_o})$$

where Ω_{oo} is the correlation $[n_o\times n_o]$-matrix between observations.

We also use a localized estimators of $[\mu,\sigma^2]$ centered at arbitrary location $x_+\in\mathbf{D}$ based on $\mathbf{r}_{+o}^k=G_+^k\mathbf{r}_o$, where G_+^k is a selection $[k\times n_o]$-matrix which selects the k-closest observations to x_+.

The localized estimators are:

$$\hat{\mu}_+^k=[\boldsymbol{i}_k^T[G_+^k\Omega_{oo}[G_+^k]^T]^{-1}G_+^k\mathbf{r}_o][\boldsymbol{i}_k^T[G_+^k\Omega_{oo}[G_+^k]^T]^{-1}\boldsymbol{i}_k]^{-1}$$

$$\hat{\sigma}_+^{k2}=\tfrac{1}{k}(G_+^k\mathbf{r}_o-\hat{\mu}_+^k\boldsymbol{i}_k)^T[G_+^k\Omega_{oo}[G_+^k]^T]^{-1}(G_+^k\mathbf{r}_o-\hat{\mu}_+^k\boldsymbol{i}_k)$$

In the parametric model, the parameters are $(\alpha_1, ..., \alpha_L, \sigma^2)$ the estimators can easily be assessed by a maximum likelihood criterion.

In hierarchical, stationary Gaussian RF model, we choose to make the assessment in an empirical Bayes setting based on the observations $\mathbf{r}_o$. The localized estimates in the observation locations $\hat{\mu}_i^k, \hat{\sigma}_i^{k2}; i = 1, ..., n_o$ are considered to be a super-population of $\hat{\mu}^k$ and $\hat{\sigma}^{k2}$. Based on this empirical sample we infer the model parameters $\hat{\mu}_m^k$ and $\hat{\tau}_m^k$ in the prior model for $[m \mid s^2]$.

The corresponding localized estimators for the posterior expectation m centered at location $x_+ \in \mathbf{D}$ based on observations $\mathbf{r}_{+o}^k$=$G_+^k \mathbf{r}_o$ is based on the conjugate properties of the prior model:

$$\begin{aligned}\hat{m}_+^k &= E[m \mid s^2, G_+^k \mathbf{r}_o] \\ &= \hat{\mu}_m^k + \hat{\tau}_m^k \boldsymbol{i}_k^T \left[\hat{\tau}_m^k \boldsymbol{i}_k \boldsymbol{i}_k^T + [G_+^k \Omega_{oo} [G_+^k]^T\right]^{-1} \left[G_+^k \mathbf{r}_o - \hat{\mu}_m^k \boldsymbol{i}_k\right]\end{aligned}$$

which is independent of s^2. This estimator appears as shrinkage estimator of $\hat{\mu}_+^k$ towards $\hat{\mu}_m^k$.

The estimates for the inverse gamma prior model parameters for variance are more complicated. We consider the squared crossvalidation errors of observation locations $s_i^2 = (r_{oi} - \hat{\mu}_m^k)^2; i = 1, ..., n_o$ to be a super-population of $\hat{s}^2$. Based on this empirical sample we infer the model parameters $\hat{\xi}_s$ and $\hat{\gamma}_s$ in the prior model for s^2.

The corresponding localized estimator for the posterior variance s^2 centered at location $x_+ \in \mathbf{D}$ based on observations $\mathbf{r}_{+o}^k = G_+^k \mathbf{r}_o$ is based on the conjugate properties of the prior model.

$$\begin{aligned}\hat{s}_+^{k2} &= E[s^2 \mid G_+^k \mathbf{r}_o] \\ &= \frac{\hat{\gamma}_s + \frac{1}{2}\left[[G_+^k \mathbf{r}_o - \hat{\mu}_m^k \boldsymbol{i}_k]^T \left[[G_+^k \Omega_{oo} [G_+^k]^T] + \hat{\tau}_m^k \boldsymbol{i}_k \boldsymbol{i}_k^T\right]^{-1} [G_+^k \mathbf{r}_o - \hat{\mu}_m^k \boldsymbol{i}_k]\right]}{\hat{\xi}_s + \frac{k}{2} - 1}\end{aligned}$$

This estimator also appears as a shrinkage estimator.

The kriging interpolation model is used for spatial predictions. In order

to define these predictors the model parameters must be assessed, either explicitely or implicitely. The empirical Bayes approach is used to define the prediction $\hat{r}_+^k$ of $r(x_+)$ in arbitrary location $x_+ \in \mathbf{D}$ based on the k-closest observations defined by G_+^k.

1.3 Summary of Thesis

In the following we summarize the Reports and Papers in the current Thesis. The thesis consists of two reports (Report I and Report II) and three papers (Paper I.A, Paper I.B and Paper II.A). The two first papers are based on Report I while the last paper is based on Report II. Report I, including Paper I.A and I.B are written for a statistical audience with particular attention to a survival and reliability theory audience. Report II, including Paper II.A, is written for a geostatistical audience.

Report I: A Simulation Study of Statistical Inference in Nonhomogeneous Poisson Processes with Emphasis on Frailty and Dynamic Behavior. A stochastic process $\{N(t); t \in T \subset \Re_+\}$ with temporal reference variable $t \in T$ is presented. The basic model is the non-stationary Poisson process (NSPP) with rate function $\{w(t); t \in T\}$. Its properties and the relevant probabilistic distribution with the model parameters are discussed. The reduced form of NSPP, stationary Poisson processes (SPP) is defined and its model parameters are stated. We also define the most common parametrization of the NSPP, namely the power law model. When several similar systems are observed, the assumption that the corresponding processes are independent and identically distributed is often questionable. In practice there may be an unobserved common heterogeneity between the systems. We consider two different approaches for analysis of such dependencies, namely the hierarchical and dynamic models. The consequences of ignoring heterogeneity and the advantages of considering a hierarchical model over a parametric model are studied. The relation between the hierachical and dynamic model approaches is investigated, both theoretically and by simulation. Detailed derivations of likelihood functions are provided, and maximum likelihood is used as the inference tool throughout the report. The conclusion is that the two approaches appear as very similar, hence hierarchical models may be viewed as alternatives to dynamic models.

Paper I.A: Unobserved heterogeneity in the power law nonhomogeneous Poisson process, presents a subset of the work in Report I, and in addition a real data example. Possible consequences of heterogeneity in the failure intensity of repairable systems is presented. The basic model is the non-stationary Poisson process with power law intensity function. The heterogeneity is modeled by introduction of unobserved gamma distributed frailties. The relevant likelihood function is derived, and maximum likelihood estimation is illustrated. In a simulation study we compare results from using a power law model without heterogeneity, with the results obtained when the heterogeneity is accounted for. A motivating data example is given.

Paper I.B: Extending minimal repair models for repairable systems: A comparision of dynamic and heterogeneous extensions of a nonhomogeneous Poisson process, contains a subset of the work in Report I, and some additional theoretical discussion as well as the analysis of a real data set. For many applications of repairable systems, the minimal repair assumption, which leads to non-stationary Poisson processes (NSPP), is not adequate. We review and study two extensions of the NSPP, the dynamic NSPP and the heterogeneous NSPP. Both extensions are motivated by specific aspects of potential applications. It has long been known, however, that the two paradigms are essentially indistinguisable in an analysis of failure data. We investigate the connection between the two approaches by extending NSPP models, both theoretically and in data studies including simulated failure processes. In particular we review the computation of the likelihood functions for the two situations and demonstrate a somewhat surprising similarity between them. This similarity is empirically confirmed in the numerical examples.

The major contributions of Report I, including Paper I.A and Paper I.B, are a study of the consequences of overlooking heterogeneity/frailities in similar repariable systems. Review of a dynamic extension of a minimal repair model is presented. Likelihood functions are established for parametric, hierarchical and dynamic models. The interrelation between the latter two are theoretically investigated. Maximum likelihood estimators for parameters of parametric and hierarchical models are derived. A simulation study is conducted and it demonstrates the effects of heterogeneity and its ignorance in models.

Report II: Localized/Shrinkage Kriging Predictors. The random field (RF) $\{R(x); x \in \mathbf{D} \subset \Re^m\}$ with spatial reference variable $x \in \mathbf{D}$ is defined to be a Gaussian RF (GRF) with spatial expectation $\{\mu(x); x \in \mathbf{D}\}$, spatial variance $\{\sigma^2(x); x \in \mathbf{D}\}$ and spatial correlation function $\{\rho(x^{'}, x^{''}); x^{'}, x^{''} \in \mathbf{D}^2\}$. We consider mostly stationary GRF and hierarchical stationary GRF in the study. The objective of the study is to improve the robustness and flexibility of spatial kriging predictors with respect to deviations from spatial stationarity assumptions. A predictor based on a non-stationary Gaussian random field is defined. The model parameters are inferred in an empirical Bayesian setting, using observations in a local neighborhood and a prior model assessed from the global set of observations. The localized predictor appears with a shrinkage effect and is coined a localized/shrinkage kriging predictor. The predictor is compared to traditional localized kriging predictors in a case study on observations of annual cumulated precipitation. A crossvalidation criterion is used in the comparision. The shrinkage predictor appears as uniformly preferable to the traditional kriging predictors. A simulation study on prediction in non-stationary Gaussian random fields is conducted. The results from this study confirms that the shrinkage predictor is favorable to the traditional ones. Moreover, the crossvalidation criterion is found to be suitable for selection of predictor. Lastly, the shrinkage predictor appears as particularly robust towards spatially varying expectation functions.

Paper II.A: Localized/Shrinkage Kriging Predictors, contain a subset of the work in Report II, and some additional theoretical discussion. The objective of the study is to improve the robustness and flexibility of spatial kriging predictors with respect to deviations from spatial stationarity assumptions. A predictor based on a non-stationary Gaussian random field is defined. The model parameters are inferred in an empirical Bayesian setting, using observations in a local neighborhood and a prior model assessed from the global set of observations. Lastly, the computational demands of localized predictors are very modest, hence the localized/shrinkage predictors are suitable for large scale spatial prediction problems.

The major contribution of Report II, including Paper II.A, is the introduction of a shrinkage concept in spatial prediction. It extends the empirical Bayesian idea to spatial interpolation. Moreover, a technique for calibration of localized predictors to global statistics is defined, and it provides crossvalidated calibrated (CVC) predictors. Lastly, real data and simulation studies

demonstrate that the CVC localized, shrinkage predictors are favorable to the traditional predictors.

References

1. Babykina, G. and Couallier, V.; 2010: "Modelling Recurrent Events for Repairable Systems Under Worse Than Old Assumption". In: Advances in Degradation Modeling (Nikulin MS et al., eds.) Statistics for Industry and Technology, DOI 10.1007/978-0-8176-4924-1 22, Birkhauser Boston, 2010.

2. Chiles, J.P and Delfiner,J.P; 1999: *Geostatistics: Modeling spatial uncertainty(Wiley series in probability and statistics)*, Wiley, New York, 695p.

3. Cressie, N.; 1993: *Statistics for Spatial Data*, John Wiley and Sons, New York, 2nd. edition.

4. Journel, A.G and Huijbregts, C.J; 1978: *Mining Geostatistics*, Academic Press Inc, London.

5. Kreager, P.; 1988: "New light on Graunt". Population Studies, Vol.42, pp.129-140.

6. Le Gat, Y.; 2013: "Extending the Yule process to model recurrent pipe failures in water supply networks". Urban Water Journal. Taylor and Francis. DOI: 10.1080/1573062X.2013.783088

7. Matheron, G.;1989: *Estimating and Choosing.* Springer-Verlag, Berlin Heidelberg.

8. Rausand, M. and Høyland, A.; 2004: *System Reliability Theory Models, statistical Methods, and Applications*, Second Edition. John Wiley and Sons, Inc., Hoboken, New Jersey.

9. Røislien, J. and Omre, H.; 2006: "T-distributed random fields: A parametric model for heavy-tailed well-log data," Mathematical Geosciences, Vol.38, No.7, pp.821-849.

Part II:

Report I

Report I

A Simulation Study of Statistical Inference in Non-homogeneous Poisson Processes with Emphasis on Frailty and Dynamic Behavior

Zeytu Gashaw Asfaw

Preprint Series in Statistics no.6, 2013, Department of Mathematical Sciences, NTNU, Trondheim, Norway.

Partly presented at the 25th Nordic conference in Mathematical statistics (NORDSTAT), June 2014, Turku University, Finland.

Abstract

A study of recurrent events for repairable systems is presented. The basic model is the nonhomogeneous Poisson process with power law intensity function. When several similar systems are under observation, the assumption that the corresponding processes are independent and identically distributed is often questionable. In practice there may be an unobserved heterogeneity among the systems. We consider two seemingly different approaches for analysis of such differences, namely by using frailities and by using dynamic models. The relation between the two approaches is investigated, both theoretically and in a simulation study. Detailed derivations of likelihood functions are provided, and maximum likelihood is used as the inference tool throughout the paper. A possible conclusion is that the two approaches are very similar, so that frailty models may be viewed as an alternative to dynamic models.

2 A Simulation Study of Statistical Inference in Non-homogeneous Poisson Processes with Emphasis on Frailty and Dynamic Behavior

2.1 Introduction

Survival analysis involves the modeling of time to event data. Classically, death or failure are considered as "events" in the survival literature, considering only single events, after which the individual or machine is dead or broken. More recently, many concepts in survival analysis have been modelled by counting process theory, which adds flexibility in that it allows modeling, for example recurrent events.

In the reliability literature, systems are generally classified into non-repairable and repairable. Non-repairable systems are those that do not get repaired when they fail. Thus, non-repairable system can fail only once, and a lifetime model such as the Weibull distribution provides the distribution of the time at which such a system fails. Most household products can be good examples of non-repairable systems.

On the other hand, repairable systems are those systems (machines, industrial plants, software, etc.) which, in the event of a failure, can be restored to satisfactory operation by any action, including parts replacements or changes to adjustable settings. A repairable system is often modeled by means of a counting process. But, to what extent can the system perform after being returned back to its regular operation? We may have that the system's performance is in the same state that the system had at the start of the operation, which means a renewal process or as good as new condition. Or, its performance may be in the same state as before the failure, which leads to a non-homogeneous Poisson process (NHPP), i.e. as bad as old condition.

NHPPs which is the main concern of this paper are useful due to their flexible assumption that events are occurring randomly in time at varying rates, instead of events being just as likely to occur in all intervals of equal size, which is the property of homogeneous Poisson processes (HPP).

There are three primary approaches to evaluating multivariate survival processes: Marginal models, Frailty models and Dynamic models (see Aalen et. al, 2008). We are focused on the last two due to the fact that marginal models, unlike fraility and dynamic models, focus on parts of the avaliable data instead of giving more realistic descriptions of the full data sets. For each of these model types we consider parametric modelling and inference. Although there is a fairly rich literature on the corresponding models and methods, all their features and particularly their interrelations are yet not fully understood nor fully investigated. The objective of the study is to perform such a study, by considering comparable models both theoretically and in a simulation study.

2.2 Notation

t	Failure time
S	Starting time
T	Ending time
τ_j	Ending time of observation for system j
n_j	Total number of failure per system
n	Total number of failure
m	Total number of system
$w(t)$	Failure rate (ROCOF)
$W(t)$	Cumulative failure rate (CROCOF)
$N(t)$	Number of failure in (0, t)
$E[N(t)]$	Expected number of failures in (0, t)
$Var[N(t)]$	Variance of number of failures in (0, t)
λ	Parameter of Power law model
β	Parameter of Power law model
δ	Parameter of Fraility model
γ	Parameter of LEYP model
$z_1(t)$	Hazard rate of T_1
$z_2(t)$	Hazard rate of T_2
$z_3(t)$	Hazard rate of T_3
$G_1(t)$	Hazard rate of T_1
$G_2(t)$	Hazard rate of T_2
$G_3(t)$	Hazard rate of T_3

2.3 Definition and Properties of NHPPs

A counting process is a non-homogeneous (or non-stationary) Poisson process with rate function $w(t)$ for $t \geq 0$, if :

1. N(0)=0.
2. $Pr\{N(t+\Delta t) - N(t) = 1\} = w(t)\Delta t + o(\Delta t)$, for all t where $\frac{O(\Delta t)}{\Delta t} \to 0$ as $\Delta t \to 0$ and $N(t)$ is the number of events occurring within $(0, t]$
3. $Pr\{N(t+\Delta t) - N(t) \geq 2\} = o(\Delta t)$, which means that the system will not experience more than one failure at the same time.

The NHPP is fully characterized by ROCOF(rate of occurrence of failure) and usually denoted by $w(t)$. This function is also called the peril rate of the NHPP.

Its cumulative rate of the process is

$$W(t) = \int_0^t w(s)ds$$

(later called the CROCOF)
Then, the probability of seeing n events in the interval $(0, t]$ is

$$Pr[N(t) = n] = \frac{[W(t)]^n}{n!} e^{-W(t)}$$

for $n = 0, 1, 2, ...$

The mean number of failures in $(0, t]$ is therefore

$$E[N(t)] = W(t)$$

and its variance is

$$Var[N(t)] = W(t)$$

Likewise, the probability of seeing n events in the interval $(t, t+s]$ is

$$Pr[(N(t+s) - N(t)) = n] = e^{-[W(t+s)-W(t)]} \frac{[W(t+s)-W(t)]^n}{n!}$$

What is said above is that $N(t+s)-N(t)$ is Poisson distributed with expected value $\int_t^{t+s} w(s)ds$ where $w(s)$ is the time dependent intensity function.

Another probabilistic property of NHPP which can help us to simulate the events of NHPP from that of HPP is stated as follows. If $t_1, t_2, ...$ are event times in a unit HPP, then $W^{-1}(t_1), W^{-1}(t_2), ...$ are event times in an NHPP with cumulative intensity function $W(t)$. Let us use the CROCOF of power law model to show how NHPP events are simulated from HPP. The CROCOF of power law is,

$$W(t) = \lambda t^{\beta}$$

Then, equate W(t) to the exponentially distributed random number having parameter one, $u \sim exp(1)$.

$$u = \lambda t^{\beta}$$
$$\Rightarrow t = [\tfrac{u}{\lambda}]^{\frac{1}{\beta}}$$

Graphically,

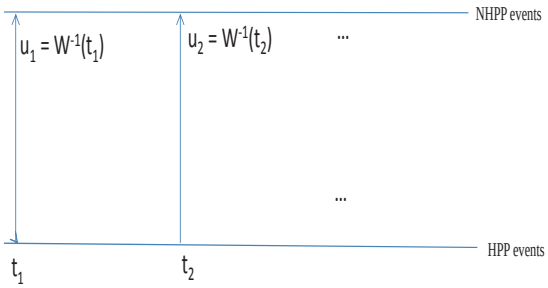

Figure 1: Simulation of NHPP events from HPP

The basic difference of NHPP from HPP is that the rate of occurrence of failures varies with time rather than being a constant. This implies that for an NHPP model the inter occurrence times are neither independent nor identically distributed. In line with this, frequently NHPP is used to model repairable systems that are subject to a minimal repair strategy, with negligible repair times. Minimal repair means that a failed system is restored just back to functioning state and the system continues as if nothing had

happened.This implies that the likelihood of system failure is the same immediately before and after a failure.

2.3.1 Parametric Models of NHPPs

Several parametric models have been established to portray the ROCOF of an NHPP, but here we are concerned to the most celebrated process model which is the power law model. This model is favored for several reasons.The first reason for the popularity of this model is that it has a very practical foundation in terms of minimal repair. The second reason here is that if the time to first failure follows the Weibull distribution, then each succeeding failure is governed by the Power law model in the case of minimal repair. From the aforementioned discussion we can say the Power law model is an extension of the Weibull distribution.

In the power law model the ROCOF of the NHPP is defined as

$$w(t) = \lambda\beta t^{\beta-1} \quad \text{for} \quad \lambda \geq 0,\ \beta \geq 0 \text{ and } \ t \geq 0$$

Its cumulative rate of ocurrence of failure (CROCOF) is

$$\begin{aligned} W(t) &= \int_0^t w(s)ds \\ &= \lambda t^{\beta} \end{aligned}$$

This intensity function was introduced in Crow (1972) as a stochastic model for the Duane reliability growth postulate. Moreover, it is referred to as a Weibull Poisson Process or the Power law Poisson Process.

The parameter β in the Power law model can give information about the system as follows:

If $0 < \beta < 1$, then the system is improving(happy).
If $\beta > 1$, then the system is deteriorating(sad).
If $\beta = 1$, then the model is reduces to an HPP.
The case $\beta = 2$ is seen to give a linearly increasing ROCOF.

2.3.2 Maximum Likelihood Estimation of Power Law Model

Suppose that the number of systems under study is m and the j^{th} system is observed continuously from time S_j to time T_j, $j = 1, 2, 3, ..., m$. During

the period $[S_j, T_j]$, let n_j be the number of failures experienced by the j^{th} system and let $t_{i,j}$ be the age of this system at the i^{th} occurrence of failure, $i = 1, 2, ..., n_j$.

It is also possible that the system boundaries S_j and T_j may be observed failure times for the j^{th} system. If $t_{n_j j} = T_j$, then the data on the j^{th} system are said to be failure terminated and T_j is a random variable with n_j fixed. If $t_{n_j j} < T_j$, then the data on the j^{th} system are said to be time terminated with n_j a random variable. Suppose that data are available from m independent systems with the same intensity function $w(t)$ and system j is observed in the interval $[S_j, T_j]$, $j = 1, 2, ..., m$, and the system j recurrence times are denoted by $t_{1j}, t_{2j}, ..., t_{n_j j}$.

Then, the NHPP likelihood function is simply the product of the individual system likelihoods

$$L = \prod_{j=1}^{m} \left[\prod_{i=1}^{n_j} [w(t_{ij})] \, exp\left\{ -[W(T_j) - W(S_j)] \right\} \right]$$

Due to the monotonicity characteristics of log transformation and for theoretical as well as technical reasons it is well recommended to work with the logarithm of the likelihood function or with the negative logarithm of it. Although the shape of these(likelihood and log-likelihood) functions are different, they have their maximum point at the same value.

Hence,

The log-likelihood function of NHPP is

$$\begin{aligned} l &= \log(L) \\ &= \sum_{j=1}^{m} \left[\sum_{i=1}^{n_j} [\log w(t_{ij})] - [W(T_j) - W(S_j)] \right]. \end{aligned}$$

The log-likelihood function of power law model with intensity function $w(t) = \lambda \beta t^{\beta - 1}$ is,

$$
\begin{aligned}
l &= \log(L) \\
&= \sum_{j=1}^{m} \left[\sum_{i=1}^{n_j} \left[\log(\lambda\beta t_{ij}^{\beta-1}) \right] - \left[\lambda T_j^{\beta} - \lambda S_j^{\beta} \right] \right] \\
&= \sum_{j=1}^{m} \left[\sum_{i=1}^{n_j} \left[\log\lambda + \log\beta + (\beta-1)\log t_{ij} \right] - \left[\lambda T_j^{\beta} - \lambda S_j^{\beta} \right] \right] \\
&= \sum_{j=1}^{m} \left[n_j \log\lambda + n_j \log\beta + (\beta-1) \sum_{i=1}^{n_j} \log t_{ij} - \lambda \left[T_j^{\beta} - S_j^{\beta} \right] \right] \\
&= \sum_{j=1}^{m} n_j \log\lambda + \sum_{j=1}^{m} n_j \log\beta + (\beta-1) \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij} - \lambda \sum_{j=1}^{m} \left[T_j^{\beta} - S_j^{\beta} \right] \\
&= n\log\lambda + n\log\beta + (\beta-1) \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij} - \lambda \sum_{j=1}^{m} \left[T_j^{\beta} - S_j^{\beta} \right] \\
&= \sum_{j=1}^{m} n_j \log\lambda + \sum_{j=1}^{m} n_j \log\beta + (\beta-1) \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij} - \lambda \sum_{j=1}^{m} \left[T_j^{\beta} - S_j^{\beta} \right] \\
&= n\log\lambda + n\log\beta + (\beta-1) \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij} - \lambda m T^{\beta}
\end{aligned}
$$

where $n = \sum_{j=1}^{m} n_j$

In the last line above we set $S_j = 0$ i.e. all systems have the same intial point which is zero and all $T_j = T$, where T is a constant number. The standard, analytical method of finding the MLEs is to take the first partial derivatives of the likelihood/log-likelihood function with respect to each parameter in the model and equate to zero.

Hence,

$$
\begin{aligned}
\frac{\partial l}{\partial \lambda} &= \frac{n}{\lambda} - mT^{\beta} \\
\hat{\lambda} &= \frac{n}{m[T^{\hat{\beta}}]}
\end{aligned}
$$

Similarly,

$$\frac{\partial l}{\partial \beta} = \frac{n}{\beta} + \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij} - \lambda m T^{\beta} [\log(T)]$$

Setting this equal to zero and using the above $\hat{\lambda}$ we get

$$\hat{\beta} = \frac{n}{n \log(T) - \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij}}$$

This gives an explicit solution for $\hat{\beta}$ which can afterwards be substituted in the expression for $\hat{\lambda}$.

We might consider the Fisher information matrix for the computation of variances and covariances of the MLEs. Fisher information matrix is used to measure the amount of information that the observed data carries about the unknown parameters. The log-likelihood function is twice differentiable with respect to each parameter,and

$$\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda^2} = \frac{-n}{\lambda^2}$$

Similarly for β parameter

$$\frac{\partial^2 l(\lambda, \beta)}{\partial \beta^2} = \frac{-n}{\beta^2} - \lambda m T^{\beta} (log(T))$$

Second mixed-partial derivatives of the log-likelihood,

$$\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda \partial \beta} = -m T^{\beta} (log(T))$$

Thus, the Fisher information matrix is

$$I(\lambda, \beta) = \begin{bmatrix} \frac{n}{\lambda^2} & m T^{\beta} (log(T)) \\ m T^{\beta} (log(T)) & \frac{n}{\beta^2} + \lambda m T^{\beta} (log(T)) \end{bmatrix}$$

For large sample size, maximum likelihood estimate have an approximate normal distribution centered on the true parameter and the variance, which is given by Fisher information matrix after substituting the maximum likelihood estimates for λ and β. Thus, asymptotically, maximum likelihood estimator is normally distributed.

Once $\hat{\lambda}$ and $\hat{\beta}$ have been estimated, the maximum likelihood estimate of the intensity function is given by:

$$w(t) = \hat{\lambda}\hat{\beta}t^{\hat{\beta}-1}, t > 0$$

and then we can draw failure intensity versus time.

2.4 Frailties in NHPP

2.4.1 Definition and Parametric Model

The notion of fraility provides a convenient way to introduce random effects, association and unobserved heterogeneity into models for survival variables. It may be considered as unmeasured risk factors, where the relevant covariates are not included in the model's specification and unknown to exist. This may otherwise be a problem in having inconsistent parameter estimates and wrong standard estimate values.

The term fraility itself was introduced by Vaupel et al. (1979) for univariate survival models, but was substantially promoted by applications to multivariate survival data from around 1980. Fraility models extend popular models such as the Cox model. Normally, survival analysis implicitly assumes a homogeneous population to be studied. In many applications, however, the study population can not be assumed to be homogeneous but must be considered as a heterogeneous sample.

Here we consider parametric models for NHPP, with a serious consideration of frailties (hidden heterogeneity) among systems. This is done in accordance with the definition of frailties in connection with the power law model. Recall that the CROCOF of power law model is

$$W(t) = \lambda t^{\beta} \text{ where } \lambda > 0,\ \beta > 0 \text{ and } t \geq 0$$

With a consideration of frailties this model can be written as $W(t) = a\lambda t^{\beta}$ where $\lambda > 0, \beta > 0, t \geq 0$ and a is a gamma distributed random number with mean 1 and variance δ. The idea is then that in the case of m systems, each system has its own value of a, i.e. $a_1, a_2, ..., a_m$, which are assumed to be independent and identically distributed with the distribution just given.

Although we have several potential frailty models to choose for the above "a's" we choose gamma frailties deliberately due to the following reason:There is no physical justification to prefer gamma frailties instead of the other but only in the line of computational and analytical aspect we prefer it. From a computational and analytical perspective, it fits very well to failure data because it is easy to derive the closed form expressions of unconditional survival, cumulative density and hazard function. This is due to the simplicity of the Laplace transform. The density of the two-parameter gamma distribution is given as

$$h_a(a) = \frac{a^{k-1}e^{-\frac{a}{\theta}}}{\theta^k \Gamma(k)}$$

where $a \geq 0$, k is shape parameter and θ is scale parameter.
Moreover, E(a)=kθ and Var(a)=k θ^2 But we want to have E(a)=1 and Var(a)=δ. Thus we have k=$\frac{1}{\delta}$ and θ=δ and density

$$h_a(a) = \frac{a^{\frac{1}{\delta}-1}exp^{\frac{-a}{\delta}}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}}$$

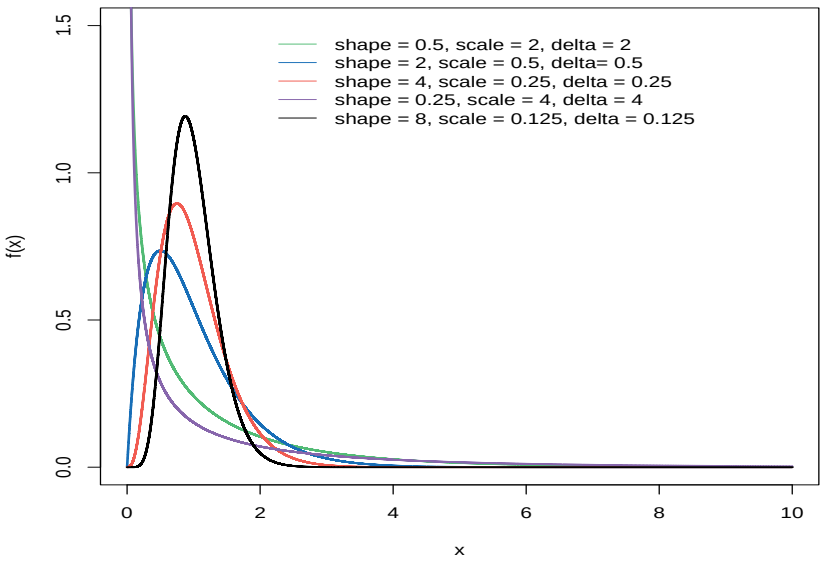

Figure 2: Graph of Gamma Densities with expected value 1

2.4.2 Maximum Likelihood Estimation of Power Law Model with Gamma Distributed Frailty

We considered the likelihood function for m systems without consideration of frailty but in this subsection we are eager to see the change in parameter estimation with a consideration of frailty a. We use similar argument as before, but now with "δ" as an additional parameter and CROCOF should be multiplied with the frailty.

Individual system likelihood function is:

$$L_j(a_j) = \prod_{i=1}^{n_j} a_j w(t_{ij}) exp\left[-a_j \left[W(T_j) - W(S_j)\right]\right]$$

Since a_j is a random variable we should find the expected value of $L_j(a_j)$ with respect to the distribution of a_j. In our case the distribution of a_j is gamma with expected value 1 and its probability density function is

$$h(a_j) = \frac{a_j^{\frac{1}{\delta}-1} e^{\frac{-a_j}{\delta}}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}}$$

The expected value of $L_j(a_j)$ is

$$\begin{aligned}
L_j &= E\left[L_j(a_j)\right] \\
&= \int L_j(a_j)h(a_j)da_j \\
&= \int \prod_{i=1}^{n_j} a_j w(t_{ij}) exp\left[-a_j\left[W(T_j) - W(S_j)\right]\right] h(a_j)da_j \\
&= \int \prod_{i=1}^{n_j} a_j w(t_{ij}) exp\left[-a_j\left[W(T_j) - W(S_j)\right]\right] \frac{a_j^{\frac{1}{\delta}-1} e^{\frac{-a_j}{\delta}}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}} da_j \\
&= \int \prod_{i=1}^{n_j} a_j \left[\lambda\beta t_{ij}^{\beta-1}\right] exp\left[-a_j\left[\lambda T_j^{\beta} - \lambda S_j^{\beta})\right]\right] \frac{a_j^{\frac{1}{\delta}-1} e^{\frac{-a_j}{\delta}}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}} da_j \\
&= \int_0^{\infty} a_j^{n_j} \lambda^{n_j}\beta^{n_j} \left(\prod_{i=1}^{n_j} t_{ij}\right)^{\beta-1} exp\left[-a_j\left[\lambda T_j^{\beta} - \lambda S_j^{\beta})\right]\right] \frac{a_j^{\frac{1}{\delta}-1} e^{\frac{-a_j}{\delta}}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}} da_j \\
&= \left[\frac{\lambda^{n_j}\beta^{n_j}\left(\prod_{i=1}^{n_j} t_{ij}\right)^{\beta-1}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}}\right] \int_0^{\infty} a_j^{n_j+\frac{1}{\delta}-1} exp\left[-a_j\left[\lambda T_j^{\beta} - \lambda S_j^{\beta} + \frac{1}{\delta}\right]\right] da_j
\end{aligned}$$

Let $r_j = n_j + \frac{1}{\delta} - 1$ and $s_j = \lambda T_j^{\beta} - \lambda S_j^{\beta} + \frac{1}{\delta}$

Then the above equation can be written as:

$$L_j = \left[\frac{\lambda^{n_j}\beta^{n_j}\left(\prod_{i=1}^{n_j} t_{ij}\right)^{\beta-1}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}}\right] \int_0^{\infty} a_j^{r_j} e^{-a_j s_j} da_j$$

To have the integrand expression of the above integration let us substitute $v = a_j s_j$. After a certain mathematical operation we have the following expression:

$$L_j = \left[\frac{\lambda^{n_j}\beta^{n_j}\left(\prod_{i=1}^{n_j} t_{ij}\right)^{\beta-1}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}}\right] \left[\frac{1}{s_j^{r_j+1}} \int_0^{\infty} v^{r_j} e^{-v} dv\right]$$

But, $\frac{1}{s_j^{r_j+1}} \int_0^\infty v^{r_j} e^{-v} dv$ equals $\frac{1}{s_j^{r_j+1}} \Gamma(r_j + 1)$ by using gamma function

Hence,

$$L_j = \frac{\lambda^{n_j} \beta^{n_j} \left(\prod_{i=1}^{n_j} t_{ij} \right)^{\beta-1} \Gamma(n_j + \frac{1}{\delta})}{\Gamma(\frac{1}{\delta}) \delta^{\frac{1}{\delta}} \left[\lambda T_j^\beta - \lambda S_j^\beta + \frac{1}{\delta} \right]^{n_j + \frac{1}{\delta}}}$$

Although power law model can have a potential to model systems that start from any time t, we restrict to time zero as starting operation time of all systems i.e $S_j = 0 \ \forall j = 1, 2, ..., m$ in this study. Then, the aforementioned individual likelihood function simplifies to

$$L_j = \frac{\lambda^{n_j} \beta^{n_j} \left(\prod_{i=1}^{n_j} t_{ij} \right)^{\beta-1} \Gamma(n_j + \frac{1}{\delta})}{\Gamma(\frac{1}{\delta}) \delta^{\frac{1}{\delta}} \left[\lambda T_j^\beta + \frac{1}{\delta} \right]^{n_j + \frac{1}{\delta}}}$$

Thus, the total likelihood function is

$$L = \prod_{j=1}^{m} L_j$$

The log likelihood function is

$$
\begin{aligned}
l(\lambda,\beta,\delta) &= \log L \\
&= \log\left[\prod_{j=1}^{m} L_j\right] \\
&= \log\left[\prod_{j=1}^{m} \frac{\lambda^{n_j}\beta^{n_j}\left(\prod_{i=1}^{n_j} t_{ij}\right)^{\beta-1}\Gamma(n_j+\frac{1}{\delta})}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}\left[\lambda T_j^{\beta}+\frac{1}{\delta}\right]^{n_j+\frac{1}{\delta}}}\right] \\
&= \sum_{j=1}^{m}\left\{\log\left[\lambda^{n_j}\beta^{n_j}\left(\sum_{i=1}^{n_j} t_{ij}\right)^{\beta-1}\Gamma(n_j+\frac{1}{\delta})\right] - \log\left[\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}\left[\lambda T_j^{\beta}+\frac{1}{\delta}\right]^{n_j+\frac{1}{\delta}}\right]\right\} \\
&= \sum_{j=1}^{m}\left\{n_j\log\lambda + n_j\log\beta + (\beta-1)\log\left(\sum_{i=1}^{n_j} t_{ij}\right) + \log\Gamma(n_j+\frac{1}{\delta})\right\} \\
&- \sum_{j=1}^{m}\left\{\log\Gamma(\frac{1}{\delta}) + (\frac{1}{\delta})\log\delta + \left[n_j+\frac{1}{\delta}\right]\log\left[\lambda T_j^{\beta}+\frac{1}{\delta}\right]\right\} \\
&= n\log\lambda + n\log\beta + (\beta-1)\sum_{j=1}^{m}\log\left(\sum_{i=1}^{n_j} t_{ij}\right) + \sum_{j=1}^{m}\log\Gamma(n_j+\frac{1}{\delta}) \\
&- \left[m\log\Gamma(\frac{1}{\delta}) + m\frac{1}{\delta}\log\delta + \sum_{j=1}^{m}\left[\left[n_j+\frac{1}{\delta}\right]\log\left[\lambda T_j^{\beta}+\frac{1}{\delta}\right]\right]\right]
\end{aligned}
$$

Hereafter, let all T_j=T and T=τ.

$$
\begin{aligned}
&= n\log\lambda + n\log\beta + (\beta-1)\sum_{j=1}^{m}\sum_{i=1}^{n_j}\log t_{ij} + \sum_{j=1}^{m}\log\Gamma(n_j+\frac{1}{\delta}) \\
&- \left[m\log\Gamma(\frac{1}{\delta}) + m\frac{1}{\delta}\log\delta + \sum_{j=1}^{m}\left[\left[n_j+\frac{1}{\delta}\right]\log\left[\lambda T^{\beta}+\frac{1}{\delta}\right]\right]\right] \\
&= n\log\lambda + n\log\beta + (\beta-1)\sum_{j=1}^{m}\sum_{i=1}^{n_j}\log t_{ij} + \sum_{j=1}^{m}\log\Gamma(n_j+\frac{1}{\delta}) \\
&- \left[m\log\Gamma(\frac{1}{\delta}) + m\frac{1}{\delta}\log\delta + \left[n+\frac{m}{\delta}\right]\log[\lambda T^{\beta}+\frac{1}{\delta}]\right]
\end{aligned}
$$

Partial derivative of $l(\lambda, \beta, \delta)$ with respect to λ is

$$\frac{\partial l(\lambda, \beta, \delta)}{\partial \lambda} = \frac{n}{\lambda} - \left[\frac{\tau^{\beta}}{\lambda\tau^{\beta} + \frac{1}{\delta}}\right]\left[n + \frac{m}{\delta}\right]$$

Then,

$$\frac{\partial l(\lambda, \beta, \delta)}{\partial \lambda} = 0$$
$$\Rightarrow \frac{n}{\lambda} = \left[\frac{\tau^{\beta}}{\lambda\tau^{\beta} + \frac{1}{\delta}}\right]\left[n + \frac{m}{\delta}\right]$$

$$\Rightarrow \hat{\lambda} = \frac{n}{m\tau^{\hat{\beta}}}$$

Similarly, Partial derivative of $l(\lambda, \beta, \delta)$ with respect to β is

$$\frac{\partial l(\lambda, \beta, \delta)}{\partial \beta} = \frac{n}{\beta} + \sum_{j=1}^{m}\sum_{i=1}^{n_j}\log t_{ij} - \left[\frac{\lambda\tau^{\beta}\log(\tau))}{\lambda\tau^{\beta} + \frac{1}{\delta}}\right]\left[n + \frac{m}{\delta}\right]$$

Hence,

$$\frac{\partial l(\lambda, \beta, \delta)}{\partial \beta} = 0$$
$$\Rightarrow \frac{n}{\beta} + \sum_{j=1}^{m}\sum_{i=1}^{n_j}\log t_{ij} - n\log(\tau)$$
$$\Rightarrow \hat{\beta} = \frac{n}{n\log(\tau) - \sum_{j=1}^{m}\sum_{i=1}^{n_j}\log t_{ij}}$$

Thus, $\hat{\lambda}$ and $\hat{\beta}$ are exactly the same as for the power law without frailities.

Likewise, using the digamma function ψ defined by

$$\psi(x) = \frac{d}{dx}\log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

, the partial derivative of $l(\lambda, \beta, \delta)$ with respect to δ is

$$\begin{aligned}
\frac{\partial l(\lambda,\beta,\delta)}{\partial \delta} &= -\frac{1}{\delta^2}\sum_{j=1}^{m}\psi(n_j+\frac{1}{\delta}) + \frac{m}{\delta^2}\psi(\frac{1}{\delta}) - m\left[-\frac{1}{\delta^2}\log(\delta)+\frac{1}{\delta^2}\right] \\
&\quad - \left[-\frac{m}{\delta^2}\log[\lambda\tau^\beta+\frac{1}{\delta}] - \frac{1}{\delta^2}\left[\frac{n+\frac{m}{\delta}}{\lambda\tau^\beta+\frac{1}{\delta}}\right]\right] \\
&= -\frac{1}{\delta^2}\sum_{j=1}^{m}\psi(n_j+\frac{1}{\delta}) + \frac{m}{\delta^2}\psi(\frac{1}{\delta}) + \frac{m}{\delta^2}\log(\delta) \\
&\quad - \frac{m}{\delta^2} + \frac{m}{\delta^2}\log[\lambda\tau^\beta+\frac{1}{\delta}] + \frac{1}{\delta^2}\left[\frac{n+\frac{m}{\delta}}{\lambda\tau^\beta+\frac{1}{\delta}}\right] \\
&= \left[\frac{-1}{\delta^2}\right]\left\{\sum_{j=1}^{m}\psi(n_j+\frac{1}{\delta}) - m\psi(\frac{1}{\delta}) - m\log(\delta) + m\right\} \\
&\quad - \left[\frac{-1}{\delta^2}\right]\left\{m\log\left[\lambda\tau^\beta+\frac{1}{\delta}\right] - \frac{n+\frac{m}{\delta}}{\lambda\tau^\beta+\frac{1}{\delta}}\right\}
\end{aligned}$$

It might be difficult to have the explicit solution of this expression by equating to zero so that using an iterative procedure is recommendable. Therefore, a function of δ will be utilized in The Newton-Raphson Method.

Since we have the explicit formula for λ and β estimate, which is independent of δ, $\frac{\partial l(\lambda,\beta,\delta)}{\partial \delta}$ is a function of δ only and we can denote it by, $f(\delta)$

$$f(\delta) = \left[\frac{-1}{\delta^2}\right]\left\{\sum_{j=1}^{m}\psi(n_j+\frac{1}{\delta}) - m\psi(\frac{1}{\delta}) - m\log(\delta) + m - m\log\left[\lambda\tau^\beta+\frac{1}{\delta}\right] - \frac{n+\frac{m}{\delta}}{\lambda\tau^\beta+\frac{1}{\delta}}\right\}$$

Note that in Matlab, psi means digamma function and psi(x) computes the digamma function of x. Similarly, psi(k,x) computes the polygamma function of x, which is the k^{th} derivative of the digamma function at x, denoted by $\psi^k(k,x)$.

2.5 Dynamic Models: Extending the NHPP

2.5.1 Introduction

In the previous section we concentrated on modeling recurrent events for repairable systems by non-homogeneous Poisson processes with and without

frailty. But, it might be difficult to quantify the effect of the repair by an amount proportional to the current intensity of the processes. Moreover, the number of repair actions up to the current time may have a heavier impact on failure intensity than aging. Due to this fact, in the following we are interested in the dynamic aspect of repairable systems to make a comparison between them.

2.5.2 Maximum Likelihood Estimation in Dynamic Model

Here we are considering maximum likelihood estimation of a dynamic model. An intensity process that depends on previous repair actions is termed as conditional intensity. The LEYP model (Linear Extension of Yule Process) (Babykina and Couallier, 2009; Le Gat, 2013) assumes that the conditional intensity evolves as

$$E[dN_j(t)|N_{t_-}] = w_j(t)dt,$$

where $N_j(t)$ counts the number of events for process j, and the history N_{t_-} contains information on (fixed and time-dependent) covariates as well as censoring and observed events in all counting processes prior to time t. Here we look at the situation where

$$w_j(t) = [1 + \gamma N_j(t)]\lambda\beta t^{\beta-1}$$

We suppose that the data concerns m systems with a consideration of these systems in a calender time interval [S,T] where S and T are the starting and the ending time of observation.

The likelihood function for the j^{th} process may be expressed as

$$L_j(\theta) = \left[\prod_{i=1}^{n_j} w_j\left(t_{ij}\right)\right] exp\left[-W_j(\tau_j)\right] \qquad (*)$$

To write the explicit form of this likelihood function we should define the ROCOF $w_j\left(t_{ij}\right)$ to be $(1 + \gamma N_j(t_{ij}))w_0(t_{ij})$ where $N_j(t_{ij})$ is the number of previous observed failures for process j and $w_0(t_{ij})$ could be power law model i.e $w_0(t_{ij}) = \lambda\beta t_{ij}^{\beta-1}$.

Next we show how to obtain W(t) for the LEYP model.

In fact we can use any positive number as a starting time of a single recurrent event process but for simplicity we consider t=0 as initial point. Let $0 \leq T_1 < T_2 < ...$ denote the event times, where T_k and T_{k+1} are the time of the k^{th} and $(k+1)^{th}$ events, in respective order. In counting processes $[N(t), 0 \leq t]$ records the cumulative number of events generated by the process but while we look in depth on the processes, the counting processes can be written as $N(t) = \sum_{k=1}^{\infty} I(T_k \leq t)$ counting the number of events occuring over the time interval [0, t].

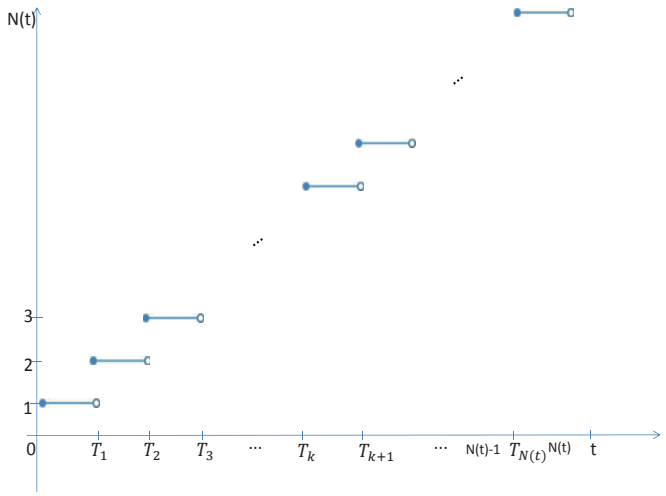

Figure 3: Counting processes representation of data on recurrent events

Now W(t) for LEYP model can derived as follows, being a function of $T_1, T_2, ..., T_{N(t)}$

$$\begin{aligned}
W(t) &= \int_0^t w(u)du \\
&= \int_0^t (1+\gamma N(u))w_0(u)du \\
&= \sum_{k=0}^{N(t)-1} \int_{T_k}^{T_{k+1}} (1+\gamma k))w_0(u)du + \int_{T_{N(t)}}^{t} (1+\gamma N(u)))w_0(u)du \\
&= \sum_{k=0}^{N(t)-1} (1+\gamma k)(W_0(T_{k+1}) - W_0(T_k)) + (1+\gamma N(t))(W_0(t) - W_0(T_{N(t)})) \\
&= \sum_{k=0}^{N(t)-1} (W_0(T_{k+1}) - W_0(T_k)) + (W_0(t) - W_0(T_{N(t)})) \\
&+ \gamma \sum_{k=0}^{N(t)-1} k(W_0(T_{k+1}) - W_0(T_k)) + \gamma N(t)(W_0(t) - W_0(T_{N(t)}))
\end{aligned}$$

By manipulating through all k values i.e $k = 0, 1, 2, ..., N(t) - 1$ we arrive at

$$W(t) = W_0(t) + \gamma \left[N(t)W_0(t) - \sum_{k=1}^{N(t)} W_0(T_k) \right]$$

Hence, from (*) on page 34,

$$L_j(\theta) = \left[\prod_{i=1}^{n_j} (1+\gamma N_j(t_{ij}))w_0(t_{ij}) \right] exp\left[-\left[W_j(\tau_j) \right] \right]$$

Substituting the ROCOF of power law model on the above equation is

$$L_j(\theta) = \left[\prod_{i=1}^{n_j} (1+\gamma N_j(t_{ij})\lambda\beta t_{ij}^{\beta-1}) \right] exp\left[-\left[\lambda\tau_j^{\beta} + \gamma \left[N(\tau_j)\lambda\tau_j^{\beta} - \sum_{i=1}^{n_j} \lambda t_{ij}^{\beta} \right] \right] \right]$$

Thus, the total likelihood function is

$$L = \prod_{j=1}^{m} L_j(\theta)$$

The log likelihood function is

$$
\begin{aligned}
l &= \log L \\
&= \log \left[\prod_{j=1}^{m} L_j(\theta) \right] \\
&= \log \left[\prod_{j=1}^{m} \left\{ \left[\prod_{i=1}^{n_j} (1 + \gamma N_j(t_{ij}) \lambda \beta t_{ij}^{\beta-1}) \right] exp \left[- \left[\lambda \tau_j^{\beta} + \gamma \left[n_j \lambda \tau_j^{\beta} - \sum_{i=1}^{n_j} \lambda t_{ij}^{\beta} \right] \right] \right] \right\} \right] \\
&= \sum_{j=1}^{m} \left\{ \left[\sum_{i=1}^{n_j} \log(1 + \gamma N_j(t_{ij}) \lambda \beta t_{ij}^{\beta-1}) \right] - \left[\lambda \tau_j^{\beta} + \gamma \left[n_j \lambda \tau_j^{\beta} - \sum_{i=1}^{n_j} \lambda t_{ij}^{\beta} \right] \right] \right\} \\
&= \sum_{j=1}^{m} \left[\sum_{i=1}^{n_j} \left[\log(1 + \gamma N_j(t_{ij})) + \log \lambda + \log \beta + (\beta - 1) \log(t_{ij}) \right] \right] \\
&- \sum_{j=1}^{m} \left[\lambda \tau_j^{\beta} + \gamma \left[n_j \lambda \tau_j^{\beta} - \sum_{i=1}^{n_j} \lambda t_{ij}^{\beta} \right] \right] \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log \left(1 + \gamma N_j(t_{ij})\right) + \sum_{j=1}^{m} n_j \log \lambda + \sum_{j=1}^{m} n_j \log \beta + (\beta - 1) \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log \left(t_{ij}\right) \\
&- \lambda \sum_{j=1}^{m} \left[\tau_j^{\beta} + \gamma \left[n_j \tau_j^{\beta} - \sum_{i=1}^{n_j} t_{ij}^{\beta} \right] \right] \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log \left(1 + \gamma(i-1)\right) + n \log \lambda + n \log \beta + (\beta - 1) \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log \left(t_{ij}\right) \\
&- \lambda \sum_{j=1}^{m} \left[\tau^{\beta} + \gamma \left[n_j \tau^{\beta} - \sum_{i=1}^{n_j} t_{ij}^{\beta} \right] \right]
\end{aligned}
$$

Note that $N_j(t_{ij}) = i - 1$ since before the event at t_{ij} we had $i - 1$ events.

Recall property of gamma function:

$$\Gamma[k+1] = k\Gamma[k]$$

So,

$$\begin{aligned}\Gamma[\frac{1}{\gamma}+n_j] &= \Gamma[\frac{1}{\gamma}+n_j-1+1]\\ &= [\frac{1}{\gamma}+n_j-1]\Gamma[\frac{1}{\gamma}+n_j-1]\\ &= [\frac{1}{\gamma}+n_j-1][\frac{1}{\gamma}+n_j-2]\Gamma[\frac{1}{\gamma}+n_j-2]\\ &= [\frac{1}{\gamma}+n_j-1][\frac{1}{\gamma}+n_j-2][\frac{1}{\gamma}+n_j-3]\Gamma[\frac{1}{\gamma}+n_j-3]\\ &\vdots\\ &= \frac{1}{\gamma^{n_j}}\left[[1+\gamma(n_j-1)][1+\gamma(n_j-2)][1+\gamma(n_j-3)]\cdots[1+\gamma]\right]\Gamma[\frac{1}{\gamma}]\end{aligned}$$

Hence,

$$\log[\Gamma[\frac{1}{\gamma}+n_j]] = -n_j\log(\gamma)+\sum_{i=1}^{n_j}\log[1+\gamma(i-1)]+\log[\Gamma(\frac{1}{\gamma})]$$

Thus, for all systems,

$$\sum_{j=1}^{m}\log[\Gamma[\frac{1}{\gamma}+n_j]] = -\sum_{j=1}^{m}n_j\log(\gamma)+\sum_{j=1}^{m}\sum_{i=1}^{n_j}\log[1+\gamma(i-1)]+\sum_{j=1}^{m}\log[\Gamma(\frac{1}{\gamma})]$$

$$\Rightarrow \sum_{j=1}^{m}\sum_{i=1}^{n_j}\log[1+\gamma(i-1)] = \sum_{j=1}^{m}\log[\Gamma[\frac{1}{\gamma}+n_j]]+n\log(\gamma)-m\log[\Gamma(\frac{1}{\gamma})]$$

Hence, complete likelihood function for dynamic model:

$$\begin{aligned}l &= n\log(\lambda)+n\log(\beta)+(\beta-1)\sum_{j=1}^{m}\sum_{i=1}^{n_j}\log\left(t_{ij}\right)+\sum_{j=1}^{m}\log\left[\Gamma(\frac{1}{\gamma}+n_j)\right]\\ &\quad -m\log\left[\Gamma\left(\frac{1}{\gamma}\right)\right]+n\log(\gamma)+\gamma\lambda\sum_{j=1}^{m}\sum_{i=1}^{n_j}(t_{ij}^{\beta})-\lambda\tau^{\beta}\left[m+n\gamma\right]\end{aligned}$$

Partial derivative of $l(\lambda, \beta, \gamma)$ with respect to λ

$$\frac{\partial l(\lambda, \beta, \gamma)}{\partial \lambda} = \frac{n}{\lambda} + \gamma \sum_{j=1}^{m} \sum_{i=1}^{n_j} (t_{ij}^{\beta}) - \tau^{\beta} [m + n\gamma]$$

Second partial derivative of $l(\lambda, \beta, \gamma)$ with respect to λ

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \lambda^2} = \frac{-n}{\lambda^2}$$

Mixed partial derivative of $l(\lambda, \beta, \gamma)$ with respect to λ and then β

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \lambda \partial \beta} = \gamma \sum_{j=1}^{m} \sum_{i=1}^{n_j} (t_{ij}^{\beta} \log(t_{ij})) - \tau^{\beta} \log(\tau) [m + n\gamma]$$

Mixed partial derivative of $l(\lambda, \beta, \gamma)$ with respect to λ and then γ

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \lambda \partial \gamma} = \sum_{j=1}^{m} \sum_{i=1}^{n_j} (t_{ij}^{\beta}) - n\tau^{\beta}$$

Partial derivative of $l(\lambda, \beta, \gamma)$ with respect to β

$$\frac{\partial l(\lambda, \beta, \gamma)}{\partial \beta} = \frac{n}{\beta} + \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log(t_{ij}) + \lambda\gamma \sum_{j=1}^{m} \sum_{i=1}^{n_j} (t_{ij}^{\beta} \log(t_{ij})) - \lambda\tau^{\beta} \log(\tau) [m + n\gamma]$$

Second partial derivative of $l(\lambda, \beta, \gamma)$ with respect to β

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \beta^2} = \frac{-n}{\beta^2} + \lambda\gamma \sum_{j=1}^{m} \sum_{i=1}^{n_j} (t_{ij}^{\beta} [\log(t_{ij})]^2) - \lambda\tau^{\beta} [\log(\tau)]^2 [m + n\gamma]$$

Mixed partial derivative of $l(\lambda, \beta, \gamma)$ with respect to β and then λ

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \beta \partial \lambda} = \gamma \sum_{j=1}^{m} \sum_{i=1}^{n_j} (t_{ij}^{\beta} \log(t_{ij})) - \tau^{\beta} \log(\tau) [m + n\gamma]$$

which is the same expression as

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \lambda \partial \beta}$$

Mixed partial derivative of $l(\lambda, \beta, \gamma)$ with respect to β and then γ

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \beta \partial \gamma} = \lambda \sum_{j=1}^{m} \sum_{i=1}^{n_j} (t_{ij}^{\beta} \log(t_{ij})) - n\lambda \tau^{\beta} \log(\tau)$$

In mathematics, the trigamma function, denoted $\psi_1(x)$,is the second of the polygamma functions, and is defined as

$$\begin{aligned} \psi_1(x) &= \frac{d^2}{dx^2} \log \Gamma(x) \\ &= \frac{d}{dx} \psi(x) \end{aligned}$$

where $\psi(x) = \frac{d}{dx} \log \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the digamma function, which is the logarithmic derivative of the gamma function. As stated before, in matlab,digamma function at x, $\psi(x)$, is psi(x).

So, here after we will use thus facts in first and second derivative of $l(\lambda, \beta, \gamma)$ and $l(\beta, \gamma)$ with respect to γ

Partial derivative of $l(\lambda, \beta, \gamma)$ with respect to γ

$$\frac{\partial l(\lambda, \beta, \gamma)}{\partial \gamma} = -\frac{1}{\gamma^2} \sum_{j=1}^{m} \psi(\frac{1}{\gamma} + n_j) + \frac{m}{\gamma^2} \psi(\frac{1}{\gamma}) + \frac{n}{\gamma} + \lambda \left[\sum_{j=1}^{m} \sum_{i=1}^{n_j} t_{ij}^{\beta} - n\tau^{\beta} \right]$$

Second partial derivative of $l(\lambda, \beta, \gamma)$ with respect to γ

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \gamma^2} = \frac{1}{\gamma^4} \sum_{j=1}^{m} \left[\psi_1(\frac{1}{\gamma} + n_j) + 2\gamma\psi(\frac{1}{\gamma} + n_j) \right] - \frac{m}{\gamma^4} \left[\psi_1(\frac{1}{\gamma}) + 2\gamma\psi(\frac{1}{\gamma}) \right] - \frac{n}{\gamma^2}$$

Mixed partial derivative of $l(\lambda, \beta, \gamma)$ with respect to γ and then λ

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \gamma \partial \lambda} = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left[t_{ij}^{\beta} \right] - n\tau^{\beta}$$

which is the same expression as

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \lambda \partial \gamma}$$

Mixed partial derivative of $l(\lambda, \beta, \gamma)$ with respect to γ and then β

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \gamma \partial \beta} = \lambda \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left[t_{ij}^{\beta} \log[t_{ij}] \right] - n\lambda\tau^{\beta} [\log(\tau)]$$

which is the same expression as

$$\frac{\partial^2 l(\lambda, \beta, \gamma)}{\partial \beta \partial \gamma}$$

Since it is difficult to get the explicit solution of λ, β and γ the Newton-Raphson method will be used.The Newton-Raphson method converges relatively fast for most functions regardless of the initial value even if difficult to set the best initial value. The steps that we should follow in the aforementioned method:
Step1: Set initial value for all the three parameter λ_0,β_0 and γ_0.
Step2: Iterative formula:

$$\begin{bmatrix} \lambda_{i+1} \\ \beta_{i+1} \\ \gamma_{i+1} \end{bmatrix} = \begin{bmatrix} \lambda_i \\ \beta_i \\ \gamma_i \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\lambda^2} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\lambda\partial\beta} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\lambda\partial\gamma} \\ \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\beta\partial\lambda} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\beta^2} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\beta\partial\gamma} \\ \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\gamma\partial\lambda} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\gamma\partial\beta} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\gamma^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial l(\lambda,\beta,\gamma)}{\partial\lambda} \\ \frac{\partial l(\lambda,\beta,\gamma)}{\partial\beta} \\ \frac{\partial l(\lambda,\beta,\gamma)}{\partial\gamma} \end{bmatrix}$$

where the matrix H=$\begin{bmatrix} \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\lambda^2} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\lambda\partial\beta} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\lambda\partial\gamma} \\ \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\beta\partial\lambda} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\beta^2} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\beta\partial\gamma} \\ \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\gamma\partial\lambda} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\gamma\partial\beta} & \frac{\partial^2 l(\lambda,\beta,\gamma)}{\partial\gamma^2} \end{bmatrix}$
is called Hessian Matrix.

It might be easier to find solution from profile likelihood so

$$\frac{\partial l\left(\lambda,\beta,\gamma\right)}{\partial\lambda} = 0$$
$$\Rightarrow \lambda = \frac{n}{\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}}$$

put in to complete likelihood function and then the profile likelihood function

is:

$$l = n\log(n) + n\log(\beta) - n\log\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\right] + \left[\beta-1\right]\sum_{j=1}^{m}\sum_{i=1}^{n_j}\log\left[t_{ij}\right]$$
$$+\sum_{j=1}^{m}\log\left[\Gamma\left[\frac{1}{\gamma}+n_j\right]\right] - m\log\left[\Gamma[\frac{1}{\gamma}]\right] + n\log(\gamma) - n$$

Partial derivative of $l\left(\beta,\gamma\right)$ with respect to β

$$\frac{\partial l\left(\beta,\gamma\right)}{\partial\beta} = \frac{n}{\beta} + \sum_{j=1}^{m}\sum_{i=1}^{n_j}\log(t_{ij}) - \frac{n\left[\tau^{\beta}\log(\tau)\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}\log(t_{ij})t_{ij}^{\beta}\right]}{\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}}$$

Second partial derivative of $l\left(\beta,\gamma\right)$ with respect to β

$$\frac{\partial^2 l\left(\beta,\gamma\right)}{\partial\beta^2} = \frac{-n}{\beta^2}$$
$$-\frac{n\left[\tau^{\beta}\left[\log(\tau)\right]^2\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left[\log(t_{ij})\right]^2 t_{ij}^{\beta}\right]\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}\right]}{\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}\right]^2}$$
$$+\frac{n\left[\tau^{\beta}\log(\tau)\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}\log(t_{ij})\right]^2}{\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}\right]^2}$$

Mixed partial derivative of $l\left(\beta,\gamma\right)$ with respect to β and then γ

$$\frac{\partial^2 l\left(\beta,\gamma\right)}{\partial\beta\partial\gamma} = -\frac{n\left[n\tau^{\beta}\left[\log(\tau)\right] - \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left[\log(t_{ij})\right]t_{ij}^{\beta}\right]\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}\right]}{\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}\right]^2}$$
$$+\frac{n\left[\tau^{\beta}\log(\tau)\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}\log(t_{ij})\right]\left[n\tau^{\beta} - \sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}\right]}{\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j}t_{ij}^{\beta}\right]^2}$$

Partial derivative of $l(\beta,\gamma)$ with respect to γ

$$\frac{\partial l(\beta,\gamma)}{\partial \gamma} = -\frac{n\left[n\tau^{\beta} - \sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\right]}{\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\right]} - \frac{1}{\gamma^2}\sum_{j=1}^{m}\psi(\frac{1}{\gamma}+n_j) + \frac{m}{\gamma^2}\psi(\frac{1}{\gamma}) + \frac{n}{\gamma}$$

Second partial derivative of $l(\beta,\gamma)$ with respect to γ

$$\frac{\partial^2 l(\beta,\gamma)}{\partial \gamma^2} = \frac{n\left[n\tau^{\beta} - \sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\right]^2}{\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\right]^2} + \frac{1}{\gamma^4}\sum_{j=1}^{m}\left[\psi_1(\frac{1}{\gamma}+n_j) + 2\gamma\psi(\frac{1}{\gamma}+n_j)\right] - \frac{m}{\gamma^4}\left[\psi_1(\frac{1}{\gamma}) + 2\gamma\psi(\frac{1}{\gamma})\right] - \frac{n}{\gamma^2}$$

Mixed partial derivative of $l(\beta,\gamma)$ with respect to γ and then β

$$\frac{\partial^2 l(\beta,\gamma)}{\partial \gamma \partial \beta} = -\frac{n\left[n\tau^{\beta}\left[\log(\tau)\right] - \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left[\log(t_{ij})\right] t_{ij}^{\beta}\right]\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\right]}{\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\right]^2} + \frac{n\left[\tau^{\beta}\log(\tau)\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\log(t_{ij})\right]\left[n\tau^{\beta} - \sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\right]}{\left[\tau^{\beta}\left[m+n\gamma\right] - \gamma\sum_{j=1}^{m}\sum_{i=1}^{n_j} t_{ij}^{\beta}\right]^2}$$

which is the same expression as

$$\frac{\partial^2 l(\beta,\gamma)}{\partial \beta \partial \gamma}$$

Step in Newton Raphson's method for profile likelihood function is: Step1: Set initial value for all the three parameter β_0 and γ_0.

Step2: Iterative formula:

$$\begin{bmatrix} \beta_{i+1} \\ \gamma_{i+1} \end{bmatrix} = \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 l(\beta,\gamma)}{\partial \beta^2} & \frac{\partial^2 l(\beta,\gamma)}{\partial \beta \partial \gamma} \\ \frac{\partial^2 l(\beta,\gamma)}{\partial \gamma \partial \beta} & \frac{\partial^2 l(\beta,\gamma)}{\partial \gamma^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial l(\beta,\gamma)}{\partial \beta} \\ \frac{\partial l(\beta,\gamma)}{\partial \gamma} \end{bmatrix}$$

where the matrix H=$\begin{bmatrix} \frac{\partial^2 l(\beta,\gamma)}{\partial \beta^2} & \frac{\partial^2 l(\beta,\gamma)}{\partial \beta \gamma} \\ \frac{\partial^2 l(\beta,\gamma)}{\partial \gamma \partial \beta} & \frac{\partial^2 l(\beta,\gamma)}{\partial \gamma^2} \end{bmatrix}$
is called Hessian Matrix.

2.6 Interrelation Between Dynamic Behaviour And Frailty Model For Poisson Processes

It is generally agreed that frailty represents an unmeasured risk factor that eventually leads to wrong conclusions if not taken into account. Moreover, there is even a misunderstanding in the concept itself. That is, it is hard to differentiate between static and dynamic frailty.

So, the first and the critical point is a confirmation of whether there is frailty or not. Second, is this frailty static or dynamic? Sometimes the current frail may depend on the past. Thus, we are keenly interested to see the interrelation between static/fixed frailty for each individual and dynamic/stochastic processes that change over time (Aalen et al., 2008).

The idea of intensity functions and counting processes are vital for modelling and statistical analysis of recurrent events. The event intensity function gives the instantaneous probability of an event occuring at t, conditional on the process history. The intensity is defined formally as

$$\lambda(t|H(t)) = lim_{\Delta t \downarrow 0} \frac{Pr[\Delta N(t)=1|H(t)]}{\Delta t}$$

where $H(t) = [N(s) : 0 \leq s < t]$ denote the history of the process at time t (see e.g. Cook and Lawless, 2006).

Just referenced to (Aalen et al., 2008), in this study, individual intensity given the frailty variable a_j is

$$\lambda(t) = aw(t)$$

where $w(t) = \lambda\beta t^{\beta-1}$ is a common baseline intensity and fixed function, that is, independent of the past, while a_j, $j = 1, 2, ..., m$, are independent identically distributed random variables give in the multiplicative factor that determines the risk of an individual. Here we considered a_j to be gamma distributed with scale parameter δ and shape parameter $\frac{1}{\delta}$.

Hence, the conditional intensity of the fraility model,

$$\lambda(t) = w(t)\frac{\frac{1}{\delta}+N(t-)}{\delta+A(t)}$$

where $A(t) = \int_0^t w(u)du \equiv \lambda t^{\beta}$

Thus

$$\lambda(t) = \left[\frac{w(t)}{\delta[\delta + A(t)]}\right][1 + \delta N(t-)]$$

$$= \left[\frac{\lambda\beta t^{\beta-1}}{\delta[\delta + \lambda t^{\beta}]}\right][1 + \delta N(t-)]$$

Recall LEYP model:

$$\lambda(t) = w^*(t)[1 + \gamma N(t-)]$$

Hence,

$$Fraility \Leftrightarrow LEYP\text{: if } w^*(t) = \frac{w(t)}{\delta[\delta+A(t)]}$$

This bi-implication shows us that fraility models may alternatively be viewed as dynamic models. Hereafter, we are interested in confirming this theoretical observation by a simulation study.

2.7 Simulate m Systems With Dynamic ROCOF

As mentioned before we suppose that the data concerns on m systems with a consideration of these systems in a calender time interval [S,T] where S and

T are the starting and the ending time of observation. We might look the system at the start of the operation i.e S=0 and up to T=10.

We can use ordinary power law model $w(t) = \lambda\beta t^{\beta-1}$ to simulate failure time $T_1 = S_1$; to simulate T_2 in the interval $[S_1, \infty)$ we might use the intensity $w(t) = (1+\gamma)\lambda\beta t^{\beta-1}$; to simulate T_3 from the interval $[S_2, \infty)$ where $S_2 = S_1 + T_2$, we can consider the intensity $w(t) = (1+2\gamma)\lambda\beta t^{\beta-1}$ and so on.

How to simulate T_1? To generate the failure time T_1 we can see the following procedures

Step 1: Let us take $z_1(t) = \lambda\beta t^{\beta-1}$ the hazard rate of T_1. Its survival function $G_1(t) = P(T_1 > t) = e^{-\int_0^t z_1(u)du}$. By integrating the intensity function the survival function is $G_1(t) = e^{-\lambda t^\beta}$

Step 2: Draw a random variable $u_1 \sim u[0,1]$ and equate to the survival function $G_1(t) = e^{-\lambda t^\beta}$

Thus,

$$\begin{aligned} e^{-\lambda t^\beta} &= u_1 \\ \Rightarrow -\log u_1 &= \lambda T_1^\beta \\ \Rightarrow S_1 = T_1 &= \left(-\frac{\log u_1}{\lambda}\right)^{1/\beta} \end{aligned}$$

How to simulate T_2?

Claim: T_2 has hazard rate $z_2(t) = (1+\gamma)\lambda\beta(S_1+t)^{\beta-1}$, which is conditional on S_1. The survival function of T_2 conditional on S_1 is

$$\begin{aligned} G_2(t) &= P(T_2 > t) \\ &= e^{-\int_0^t z_2(u)du} \\ &= e^{-\int_0^t (1+\gamma)\lambda\beta(S_1+u)^{\beta-1}du} \\ &= e^{-(1+\gamma)\lambda\beta\int_0^t (S_1+u)^{\beta-1}du} \end{aligned}$$

$$= e^{-(1+\gamma)\lambda\beta\int_{S_1}^{S_1+t} x^{\beta-1}dx}$$

$$= e^{-(1+\gamma)\lambda\left[(S_1+t)^\beta - S_1^\beta\right]}$$

Let us draw a random variable $u_2 \sim u[0,1]$ and equate to the survival function $u_2 = e^{-(1+\gamma)\lambda\left[(S_1+T_2)^\beta - S_1^\beta\right]}$

This implies that

$$u_2 = e^{-(1+\gamma)\lambda\left[(S_1+T_2)^\beta - S_1^\beta\right]}$$

$$\Rightarrow \log u_2 = -(1+\gamma)\lambda\left[(S_1+T_2)^\beta - S_1^\beta\right]$$

$$\Rightarrow -\frac{\log u_2}{(1+\gamma)\lambda} = (S_1+T_2)^\beta - S_1^\beta$$

$$\Rightarrow S_1 + T_2 = \left(S_1^\beta - \frac{\log u_2}{(1+\gamma)\,\lambda}\right)^{1/\beta}$$

$$\Rightarrow T_2 = \left(S_1^\beta - \frac{\log u_2}{(1+\gamma)\lambda}\right)^{1/\beta} - S_1$$

Thus,

$$S_2 = \left(S_1^\beta - \frac{\log u_2}{(1+\gamma)\lambda}\right)^{1/\beta}$$

How to simulate T_3?
Claim: T_3 has hazard rate $z_3(t) = (1+2\gamma)\lambda\beta(S_2+t)^{\beta-1}$, which is conditional on S_2. The survival function of T_3 is

$$G_3(t) = P(T_3 > t)$$

$$= e^{-\int_0^t z_3(u)du}$$

$$= e^{-\int_0^t (1+2\gamma)\lambda\beta(S_2+u)^{\beta-1}du}$$

$$= e^{-(1+2\gamma)\lambda\beta\int_0^t (S_2+u)^{\beta-1}du}$$

$$= e^{-(1+2\gamma)\lambda\beta\int_{S_2}^{S_2+t} x^{\beta-1}dx}$$

$$= e^{-(1+2\gamma)\lambda\left[(S_2+t)^{\beta}-S_2^{\beta}\right]}$$

Let us draw a random variable $u_3 \sim u[0,1]$ and equate to the survival function $u_3 = e^{-(1+\gamma)\lambda\left[(S_2+T_2)^{\beta}-S_2^{\beta}\right]}$

This implies that

$$\begin{aligned}
u_3 &= e^{-(1+2\gamma)\lambda\left[(S_2+T_3)^{\beta}-S_2^{\beta}\right]} \\
\Rightarrow \log u_3 &= -(1+2\gamma)\lambda\left[(S_2+T_3)^{\beta}-S_2^{\beta}\right] \\
\Rightarrow -\frac{\log u_3}{(1+2\gamma)\lambda} &= (S_2+T_3)^{\beta}-S_2^{\beta} \\
\Rightarrow S_2+T_3 &= \left(S_2^{\beta}-\frac{\log u_3}{(1+2\gamma)\,\lambda}\right)^{1/\beta} \\
\Rightarrow T_3 &= \left(S_2^{\beta}-\frac{\log u_3}{(1+2\gamma)\lambda}\right)^{1/\beta}-S_2
\end{aligned}$$

Thus,

$$S_3 = \left(S_2^{\beta}-\frac{\log u_3}{(1+2\gamma)\lambda}\right)^{1/\beta}$$

Similarly,

$$S_4 = \left(S_3^{\beta}-\frac{\log u_4}{(1+3\gamma)\lambda}\right)^{1/\beta}$$

,

$$S_5 = \left(S_4^\beta - \frac{\log u_5}{(1+4\gamma)\lambda} \right)^{1/\beta}$$

and so on. Graphically,

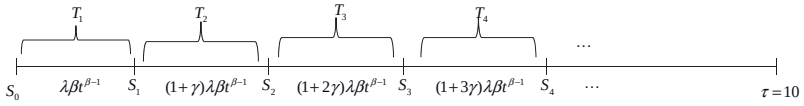

Figure 4: Relation between the interoccurrence times (T_i), caleander time S_i and intensity $w(t) = (1 + i\gamma)\lambda\beta t^{\beta-1}$ where $i = 0, 1, ...$

If $\gamma = 0$, then LEYP process is an ordinary power law process.

$$S_1 = T_1 = \left(-\frac{\log u_1}{\lambda} \right)^{1/\beta}$$

$$S_2 = \left(S_1^\beta - \frac{\log u_2}{\lambda} \right)^{1/\beta}$$

$$S_3 = \left(S_2^\beta - \frac{\log u_3}{\lambda} \right)^{1/\beta}$$

and so on but stop while we are passing τ. So we can simulate ordinary power law process from LEYP model by making $\gamma = 0$.

We may describe it graphically as:

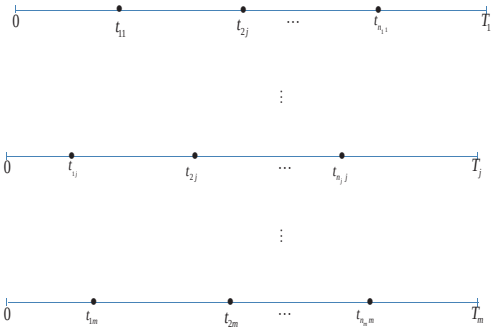

Figure 5: Observation of failure times of m systems.

Similarly, we can simulate power law with frailty from LEYP model. For process ♯1, draw an "a_1" from gamma distribution with expected value 1 and variance δ. Then the failure times are simulated as

$$S_1^{(1)} = T_1^{(1)} = \left(-\frac{\log u_1}{\lambda \cdot a_1}\right)^{1/\beta}$$

$$S_2^{(1)} = \left(S_1^{(1)\beta} - \frac{\log u_2}{\lambda \cdot a_1}\right)^{1/\beta}$$

$$S_3^{(1)} = \left(S_2^{(1)\beta} - \frac{\log u_3}{\lambda \cdot a_1}\right)^{1/\beta}$$

and so on.

For process ♯2, draw an "a_2" from gamma distribution.

$$S_1^{(2)} = T_1^{(2)} = \left(-\frac{\log u_1}{\lambda \cdot a_2}\right)^{1/\beta}$$

$$S_2^{(2)} = \left(S_1^{(2)\beta} - \frac{\log u_2}{\lambda \cdot a_2}\right)^{1/\beta}$$

$$S_3^{(2)} = \left(S_2^{(2)\beta} - \frac{\log u_3}{\lambda \cdot a_2}\right)^{1/\beta}$$

and so on. Graphically, provided $\beta = \frac{3}{2}$,

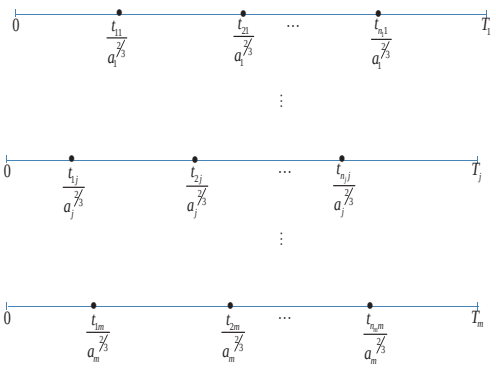

Figure 6: Observation of failure times of m system with frailties a

2.8 Maximum Likelihood Estimation

Although the method of maximum likelihood is an efficient method once we have an explicit likelihood function,it is a routine procedure for obtaining estimators for unknown parameters from a set of data. It's estimate for θ is a value of θ which maximize the likelihood function over the parameter space. It is a single parameter value which is most likely in light of what

have been observed.

Definition:

1. The likelihood function is the joint probability (density) function of observable random variables but it is viewed as the function of the parameters given the realized random variables.

   Mathematically,let $x_1, x_2, ..., x_n$ be a random sample of size n from the discrete pdf $p_X(x;\theta)$. The likelihood function, L(θ), is the product of the pdf evaluated at the n $x_i^{'}s$. That is,

   $$L(\theta) = \Pi_{i=1}^{n} p_X(x_i;\theta)$$

   Similarly,if $x_1, x_2, ..., x_n$ be a random sample of size n from a continuous pdf, $f_X(x;\theta)$. The likelihood function can be written as

   $$L(\theta) = \Pi_{i=1}^{n} f_X(x_i;\theta)$$

   where θ is an unknown parameter in both cases. Moreover, let θ_l is the value of the parameter such that $L(\theta_l) \geq L(\theta)$ for all possible values of θ. Then θ_l is maximum likelihood estimate(MLE) θ.

2. The function $l(\theta) = lnL(\theta)$ is the log likelihood function of $x_1, x_2, ..., x_n$.

3. The function $S(\theta) = \frac{\partial}{\partial x} l(\theta)$ is the score function of $x_1, x_2, ..., x_n$.

4. The function $I(\theta) = -\frac{\partial^2}{\partial x} l(\theta)$ is the information matrix of $x_1, x_2, ..., x_n$.

The Fisher information matrix is used to calculate the covariance matrices associated with maximum-likelihood estimates so that we can easily estimate the standard deviation of estimates.

2.9 Preliminary Analysis

The main objective of the preliminary analysis is to give a simple overview about simulation of a single process observed on the time interval [0,10].

2.9.1 Ordinary Power Law Model

This is simulation of a single process observed on the interval [0,10], where parameter values are $\lambda = 2$ and $\beta = 1.5$.

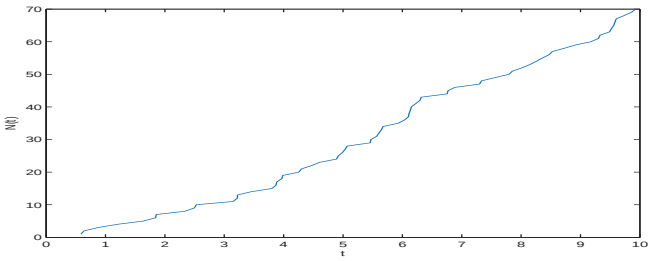

Figure 7: Random failures time t Vs number of failure N(t)

2.9.2 Frailty

These are simulations of single processes observed on the interval [0,10] for the same $\lambda = 2$ and $\beta = 1.5$ but varying δ values

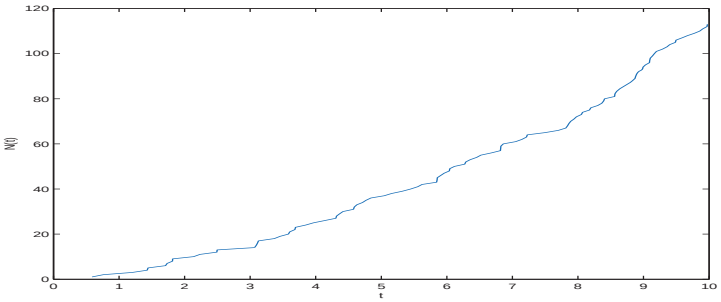

Figure 8: Random failures time t Vs number of failure N(t), δ=0.2

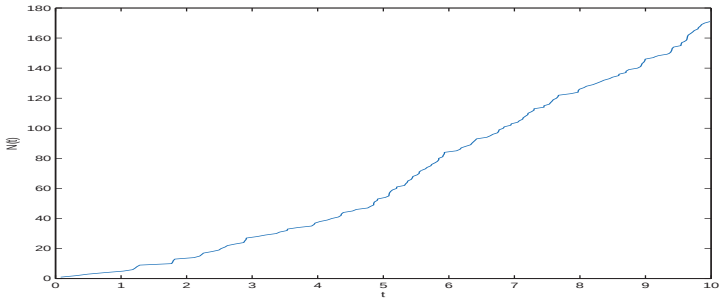

Figure 9: Random failures time t Vs number of failure N(t), δ=0.4

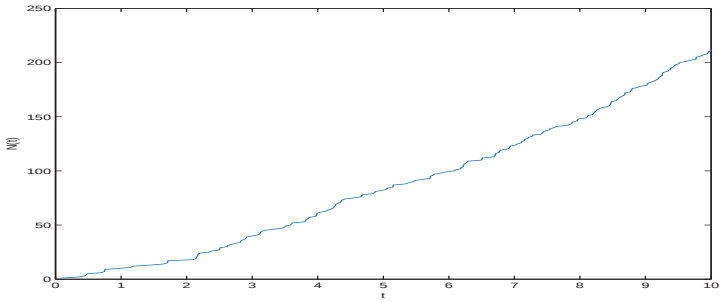

Figure 10: Random failures time t Vs number of failure N(t), δ=0.6

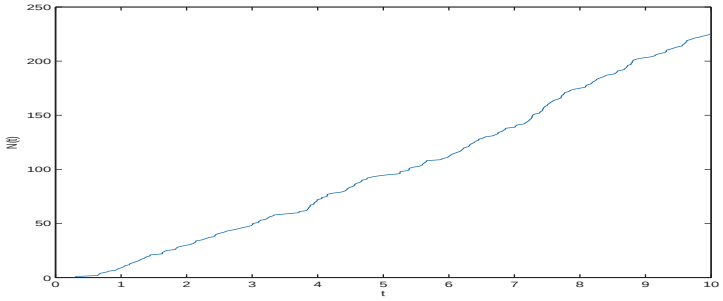

Figure 11: Random failures time t Vs number of failure N(t), δ=0.8

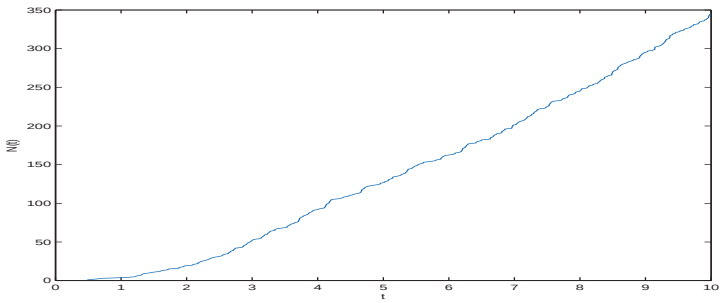

Figure 12: Random failures time t Vs number of failure N(t), δ=1

2.9.3 Dynamic Behaviour

These are simulations of single processes observed on the interval [0,10] for the same $\lambda = 2$ and $\beta = 1.5$ but varying γ values

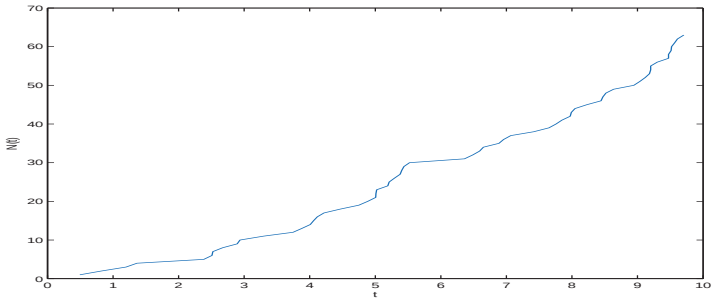

Figure 13: Random failures time t Vs number of failure N(t), γ=0.001

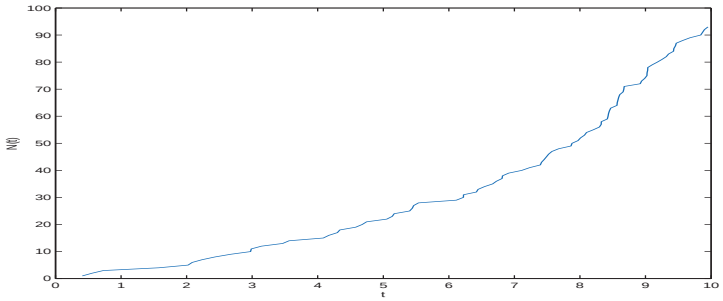

Figure 14: Random failures time t Vs number of failure N(t), γ=0.01

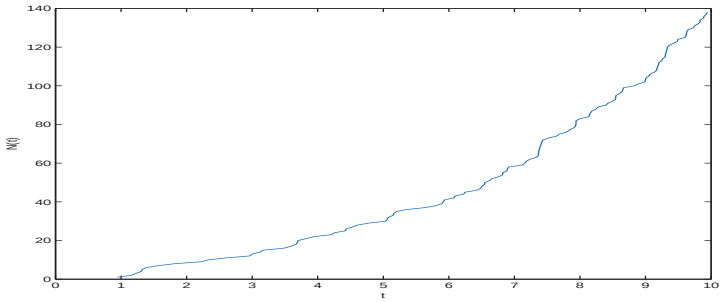

Figure 15: Random failures time t Vs number of failure N(t), γ=0.02

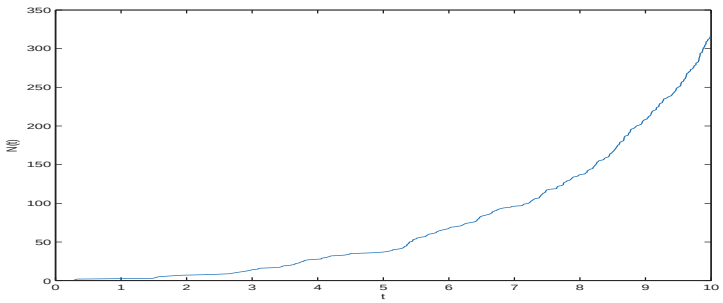

Figure 16: Random failures time t Vs number of failure N(t), γ=0.04

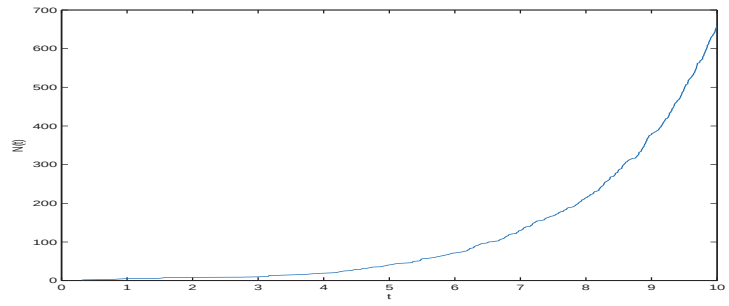

Figure 17: Random failures time t Vs number of failure N(t), γ=0.06

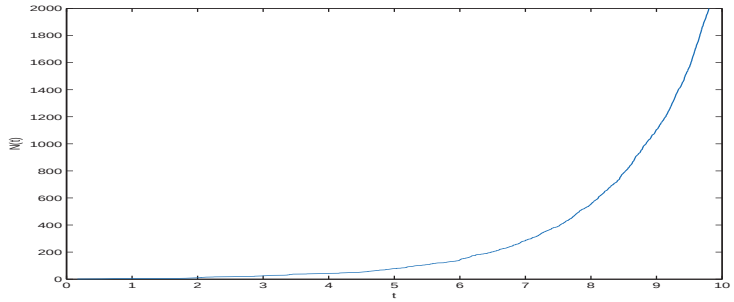

Figure 18: Random failures time t Vs number of failure N(t),γ=0.08

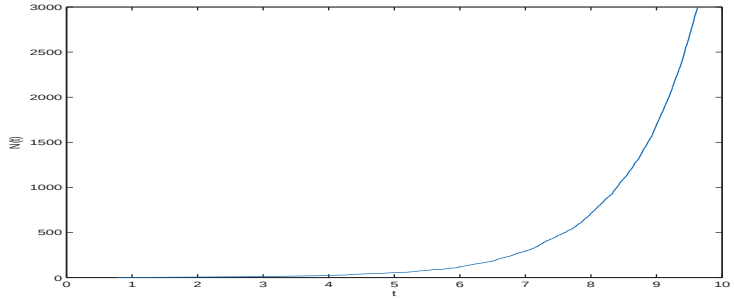

Figure 19: Random failures time t Vs number of failure N(t), γ=0.1

2.10 Simulation Study

2.10.1 Power Law Model

2.10.1.1 Maximum Likelihood Estimate

Maximum likelihood estimates of ordinary power law model for single simulation with a given value m=20, λ=2, β=1.5. The ML estimates are $\hat{\lambda}$=1.9926 and $\hat{\beta}$=1.4999. This resulted in the Fisher information

$$I\left(\hat{\lambda},\hat{\beta}\right)=\begin{bmatrix} 0.0407 & -0.0082 \\ -0.0082 & 0.0018 \end{bmatrix}$$

From the Fisher information matrix we can further derive the standard deviation of $\hat{\lambda}$ and $\hat{\beta}$ to be 0.2018 and 0.0423 in respective order. Its maximum likelihood,

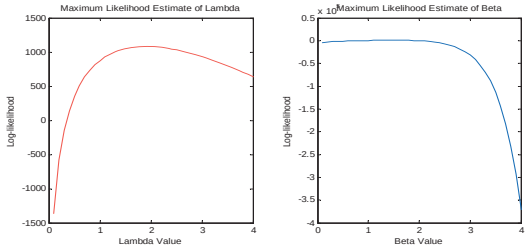

Figure 20: Maximum likelihood estimates of ordinary power law model

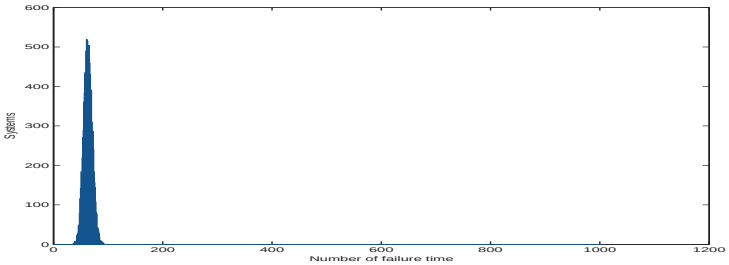

Figure 21: Histogram of Number of failure Vs Systems; λ=2, β=1.5, 10000 data sets and m=20 systems per data sets

Hereafter, average (Ave.)and standared deviation (St.D) are denoted as follows:

For estimate [$\hat{\lambda}$, $\hat{\beta}$, $\hat{\delta}$ and $\hat{\gamma}$]: Average (Ave.) is the sum of all estimates divided by number of data sets and standard deviation (St.D) is the average distance between the estimates and the mean(Average of estimates).

For number of failures [n]: Average (Ave.) is the sum of all number of failure per system divided by the product of number of system per data set ,and number of total data sets. Its standard deviation (St.D) is the square root of the quadratic distance between the number of failures per process and the mean(Average number of failures).

Data	m	True Value		n		Estimates			
10000	20	λ	β	Average	St.D	$\hat{\lambda}$		$\hat{\beta}$	
						Average	St.D	Average	St.D
		2	1.5	63.2896	7.9675	2.0034	0.2042	1.5012	0.0424
			1	19.9312	4.4597	2.0000	0.2495	1.0030	0.0504
			0.75	11.2457	3.3483	2.0015	0.2619	0.7526	0.0497

Table 1: Power law data and power law estimates

Data	m	True Value			n		Estimates			
10000	20	λ	β	δ	Average	St.D	$\hat{\lambda}$		$\hat{\beta}$	
							Average	St.D	Average	St.D
		2	1.5	0.2	62.5152	29.2594	1.9850	0.3041	1.5036	0.0488
				0.4	63.0461	40.7118	1.9918	0.3623	1.5023	0.0486
				0.6	62.6755	49.3998	1.9892	0.4078	1.5020	0.0485
				0.8	63.1969	56.7187	1.9962	0.4543	1.5018	0.0485
				1	61..5949	63.0386	1.9845	0.4952	1.5019	0.0494
			1	0.2	19.9349	9.9328	2.0050	0.3188	1.0024	0.0508
				0.4	19.9346	13.2303	1.9942	0.3754	1.0036	0.0504
				0.6	20.0902	16.4537	1.9949	0.4270	1.0033	0.0512
				0.8	20.0922	18.5500	1.9966	0.4700	1.0049	0.0520
				1	19.9214	20.3878	1.9906	0.5143	1.0085	0.0533
			0.75	0.2	11.3481	6.0628	1.9478	0.3118	0.7671	0.0469
				0.4	11.2743	7.7920	1.9453	0.3678	0.7680	0.0462
				0.6	11.5032	9.3936	1.9335	0.4249	0.7712	0.0479
				0.8	11.3861	10.5682	1.9247	0.4760	0.7745	0.0499
				1	11.5880	11.9995	1.9198	0.5246	0.7799	0.0523

Table 2: Fraility data and power law estimates

Table Summary
Case 1: $\beta > 1$
As δ increases: average number of failures per system are nearly constant but the standard deviation (St.D) increases; average of λ estimates are nearly constant but the standard deviation (St.D) increases; average and standared deviation (St.D)of β estimates are fairly constant.

Case 2: $\beta = 1$
As δ increase: similar to case 1.

Case 3: $\beta < 1$
As δ increase: average number of failures per system are very nearly constant but the standard deviation (St.D) increases; average of λ estimates are slightly decrease but standard deviation of λ estimates are increases; average and standared deviation (St.D)of β estimates are increases.

Data	m	True Value			n		Estimates			
10000	20	λ	β	γ	Average	St.D	$\hat{\lambda}$		$\hat{\beta}$	
							Average	St.D	Average	St.D
		2	1.5	0.001	65.0923	8.2762	1.9566	0.2013	1.5256	0.0428
				0.01	88.1634	12.6667	1.5068	0.1652	1.7700	0.0455
				0.02	126.9190	21.5008	0.9970	0.1171	2.1080	0.0483
				0.04	289.1689	59.6044	0.2704	0.0344	3.0315	0.0516
				0.06	726.9944	180.2163	0.0356	0.0045	4.3109	0.0495
			1	0.001	20.2509	4.5139	2.0015	0.2537	1.0073	0.0512
				0.01	22.1547	5.1818	1.9607	0.2539	1.0558	0.0517
				0.02	24.5958	6.0606	1.9192	0.2549	1.1105	0.0532
				0.04	30.8249	8.4493	1.7955	0.2505	1.2356	0.0551
				0.06	38.8179	11.3669	1.6233	0.2399	1.3812	0.0575
			0.75	0.001	11.2690	3.3855	1.9966	0.2703	0.7562	0.0515
				0.01	11.8937	3.5962	2.0140	0.2696	0.7743	0.0507
				0.02	12.6623	4.0317	2.0249	0.2805	0.7978	0.0523
				0.04	14.2409	4.7588	2.0406	0.2861	0.8462	0.0529
				0.06	16.0941	5.5937	2.0576	0.3010	0.8961	0.0550

Table 3: Dynamic data and power law estimates

Table Summary
Case 1: $\beta > 1$
As γ increase: average and standard deviation (St.D) of number of failure per system are highly increases; averages and standard deviation (St.D) of λ estimates are highly decrease; average of β estimates increases but the standard deviations (St.D) fairly constant.

Case 2: $\beta = 1$
As γ increase: average and standard deviation (St.D) of number of failure per system are increase; average of λ estimates are decrease and its standard deviation (St.D) estimates are fairly constant; average of β estimates are increase but the standard deviation (St.D) fairly constant.

Case 3: $\beta < 1$
As γ increase: average and standard deviation (St.D) of number of failure per system are slowly increases; average of λ estimates are fairly constant but the standard deviations (St.D) slowly increases; average of β estimates are slowly increase but standard deviation (St.D) fairly constant.

2.10.2 A Gamma Multiple(Frailty)Power Law Model

2.10.2.1 Maximum Likelihood Estimate

Here we are keenly interested to estimate parameters λ, β and δ.

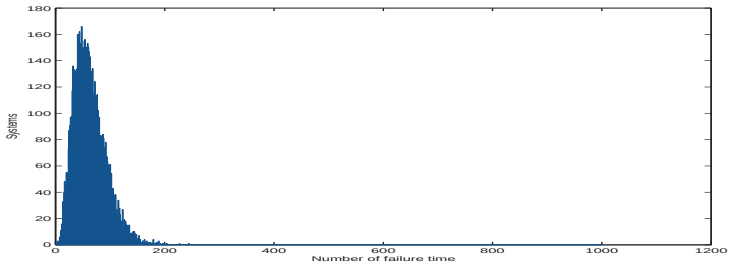

Figure 22: Histogram of Number of failure Vs Systems; λ=2, β=1.5, δ=0.2,10000 data sets and m=20 systems per data sets

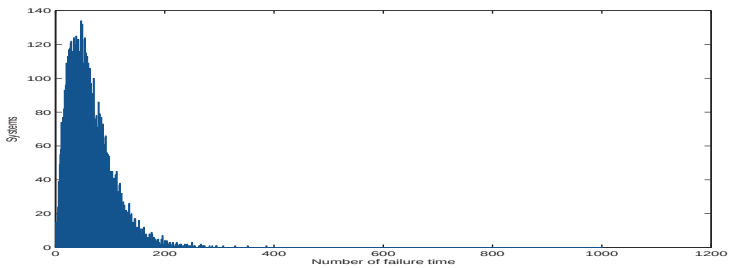

Figure 23: Histogram of Number of failure Vs Systems; λ=2, β=1.5, δ=0.4,10000 data sets and m=20 systems per data sets

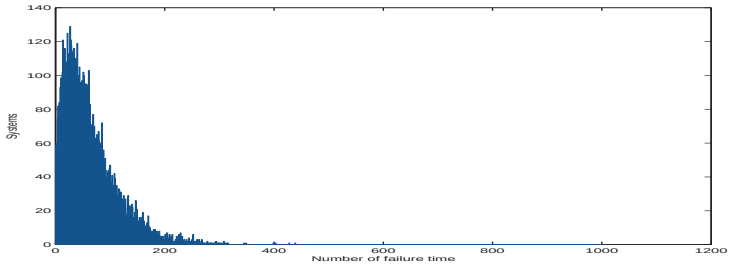

Figure 24: Histogram of Number of failure Vs Systems; λ=2, β=1.5, δ=0.6,10000 data sets and m=20 systems per data sets

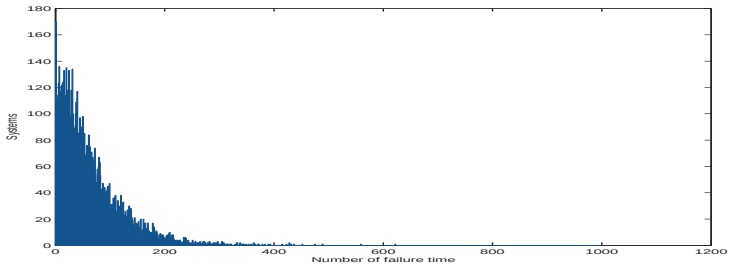

Figure 25: Histogram of Number of failure Vs Systems; λ=2, β=1.5, δ=0.8,10000 data sets and m=20 systems per data sets

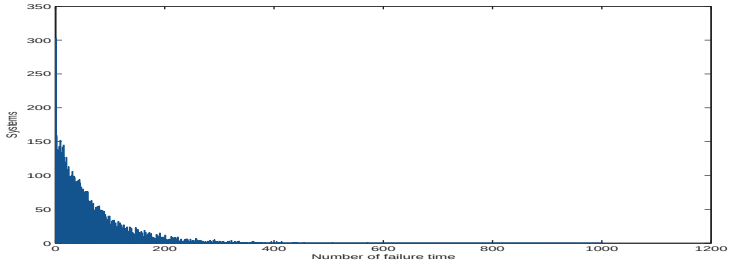

Figure 26: Histogram of Number of failure Vs Systems; λ=2, β=1.5, δ=1,10000 data sets and m=20 systems per data sets

Data	m	True Value			n		Estimates					
10000	20	λ	β	δ	Ave.	St.D	$\hat{\lambda}$		$\hat{\beta}$		$\hat{\delta}$	
							Ave.	St.D	Ave.	St.D	Ave.	St.D
		2	1.5	0.0001	63.2373	7.8956	2.0062	0.2024	1.5005	0.0423	0.002	0.0037
				0.01	63.4082	10.2008	2.0054	0.2071	1.5009	0.0422	0.0112	0.0097
				0.1	63.0648	21.4421	2.0075	0.2456	1.5005	0.0421	0.0942	0.0352
				0.2	63.2006	29.6606	2.0042	0.2853	1.5013	0.0425	0.1897	0.0651
				0.4	63.0469	40.1756	2.0033	0.3450	1.5009	0.0426	0.3795	0.1208
				0.6	63.3880	49.3983	2.0063	0.4033	1.5014	0.0431	0.5718	0.1750
				0.8	63.0149	56.1749	2.0050	0.4452	1.5013	0.0430	0.7511	0.2223
				1	63.4131	64.1655	2.0033	0.5023	1.5021	0.0432	0.9309	0.2610
			1	0.001	19.9770	4.6594	2.0057	0.2559	1.0030	0.0511	0.0071	0.0137
				0.2	19.9885	9.9238	1.9987	0.3231	1.0023	0.0509	0.1879	0.0774
				0.4	20.2335	13.4745	2.0052	0.3802	1.0026	0.0508	0.3742	0.1305
				0.6	19.8679	15.8587	1.9968	0.4285	1.0043	0.0514	0.5520	0.1769
				0.8	20.1814	18.2127	1.9923	0.4691	1.0051	0.0518	0.7158	0.2107
				1	20.4585	20.8251	1.9843	0.5149	1.0083	0.0524	0.8679	0.2421
			0.75	0.001	11.2590	3.3240	2.0083	0.2626	0.7521	0.0513	0.0123	0.0231
				0.2	11.8200	6.8717	2.0535	0.3142	0.7502	0.0493	0.1888	0.0924
				0.4	11.2779	7.8143	1.9923	0.3934	0.7557	0.0519	0.3620	0.1381
				0.6	11.1454	9.2048	1.9846	0.4330	0.7572	0.0522	0.5234	0.1763
				0.8	11.1906	10.4320	1.9702	0.4873	0.7624	0.0539	0.6685	0.2054
				1	11.1918	11.4700	1.959	0.5285	0.7663	0.0552	0.7917	0.2274

Table 4: Frailty data and frailty estimates

Table Summary
Case 1: $\beta > 1$
As δ increase: average number of failure per system are constant but its standard deviation (St.D) increases; average of λ estimates are fairly constant but the standard deviation (St.D) slowly increase: average and standard deviation (St.D) of β estimates are fairly constant; average and standard deviation (St.D) of δ estimates are increase.

Case 2: $\beta = 1$
As δ increase: similar as **Case 1**

Case 3: $\beta < 1$
As δ increase: average number of failure per system are fairly constant but the standard deviation (St.D) increases; average of λ estimates are decrease but the standard deviation (St.D) increases: average of β estimates are increase but the standard deviations (St.D) constant; average and standard deviation (St.D) of δ are increase.

Data	m	True V.		n		Estimates					
10000	20	λ	β	Ave.	St.D	$\hat{\lambda}$		$\hat{\beta}$		$\hat{\delta}$	
						Ave.	St.D	Ave.	St.D	Ave.	St.D
		2	1.5	63.1451	8.0541	2.0059	0.2057	1.5008	0.0430	0.0021	0.0036
			1	20.0288	4.4658	2.0051	0.2526	1.0019	0.0509	0.0064	0.0119
			0.75	11.2920	3.3512	1.9992	0.2653	0.7535	0.0507	0.0116	0.0215

Table 5: Power law data and frailty estimates

Data	m	True Value			n		Estimates					
1000	20	λ	β	γ	Ave.	St.D	$\hat{\lambda}$		$\hat{\beta}$		$\hat{\delta}$	
							Ave.	St.D	Ave.	St.D	Ave.	St.D
		2	1.5	0.00001	63.3176	8.0255	2.0030	0.2019	1.5017	0.0423	0.0020	0.0038
				0.001	65.2978	8.2883	1.9562	0.1972	1.5254	0.0425	0.0026	0.0042
				0.01	88.4093	12.8219	1.5084	0.1668	1.7694	0.0459	0.0111	0.0080
				0.02	127.0301	21.0328	0.9979	0.1159	2.1078	0.0480	0.0229	0.0119
				0.04	287.3850	61.2510	0.2698	0.0349	3.0325	0.0516	0.0387	0.0133
				0.06	719.8430	176.3796	0.0354	0.0044	4.3141	0.0479	0.0569	0.0194
			1	0.00001	19.9410	4.3837	2.0018	0.2592	1.0019	0.0513	0.0062	0.0114
				0.001	20.4110	4.6004	1.9959	0.2521	1.0097	0.0509	0.0069	0.0118
				0.01	21.9510	5.0593	1.9584	0.2518	1.0558	0.0514	0.0135	0.0174
				0.02	24.5330	6.0879	1.9214	0.2596	1.1113	0.0537	0.0238	0.0221
				0.04	30.8490	8.4581	1.8012	0.2496	1.2348	0.0551	0.0474	0.0296
				0.06	37.9590	11.3169	1.6259	0.2481	1.3816	0.0589	0.0682	0.0334
				0.08	50.1180	16.0797	1.4173	0.2131	1.5465	0.0587	0.0763	0.0305
				0.1	64.0240	22.7574	1.1807	0.1888	1.7359	0.0616	0.0943	0.0361
				0.2	269.6050	120.1785	0.2495	0.0397	3.0342	0.0576	0.1902	0.0602
			0.75	0.00001	11.2950	3.3210	2.0022	0.2644	0.7536	0.0490	0.0117	0.0215
				0.001	11.3640	3.5082	2.0046	0.2663	0.7550	0.0513	0.0126	0.0229
				0.01	11.8860	3.5045	2.0044	0.2715	0.7769	0.0520	0.0164	0.0253
				0.02	12.5950	4.1497	2.0310	0.2757	0.7970	0.0511	0.0261	0.0320
				0.04	14.3780	4.8219	2.0294	0.2925	0.8478	0.0541	0.0441	0.0410
				0.06	16.0900	5.8230	2.0615	0.3080	0.8961	0.0544	0.0668	0.0461
				0.08	18.5330	7.1101	2.0409	0.3171	0.9561	0.0574	0.0942	0.0534
				0.2	40.9700	19.5547	1.6950	0.2931	1.4002	0.0611	0.1937	0.0709
				0.4	219.4870	136.0986	0.5884	0.1085	2.5791	0.0532	0.3823	0.1157

Table 6: Dynamic data and frailty estimates

Table Summary
Case 1: $\beta > 1$
As γ increase: average and standard deviation (St.D) of number of failures per system are highly increases; average and standard deviation (St.D) of λ estimates are highly decreases; average of β estimates are increases but the standard deviation (St.D) fairly constant; average and standard deviation

(St.D) of δ estimates are increases.

Case 2: $\beta = 1$
As γ increase: average and standard deviation (St.D) of failures per system are increases; average and the standard deviation (St.D) decreases; average of β estimates are increases but the standard deviations (St.D) fairly constant; average and standard deviation (St.D) of δ estimates are increases.

Case 3: $\beta < 1$
As γ increase: average and standard deviation (St.D) of failure per system are increases; average and standard deviation (St.D) of λ estimates are decreases; average of β estimates increase but the standard deviations (St.D) fairly constant; average and standard deviation (St.D) of δ estimates are increases.

2.10.3 A Dynamic View of Power Law Model

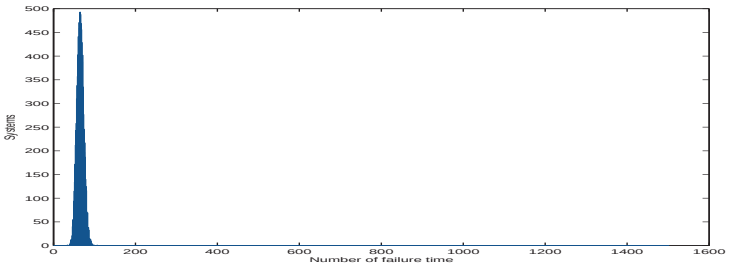

Figure 27: Histogram of Number of failure Vs Systems; λ=2, β=1.5,γ=0.001,10000 data sets and m=20 systems per data sets

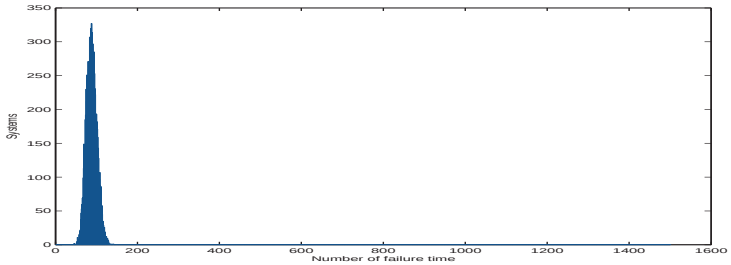

Figure 28: Histogram of Number of failure Vs Systems; λ=2, β=1.5,γ=0.01,10000 data sets and m=20 systems per data sets

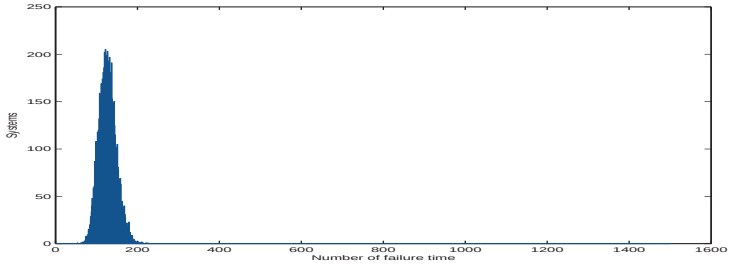

Figure 29: Histogram of Number of failure Vs Systems; λ=2, β=1.5,γ=0.02,10000 data sets and m=20 systems per data sets

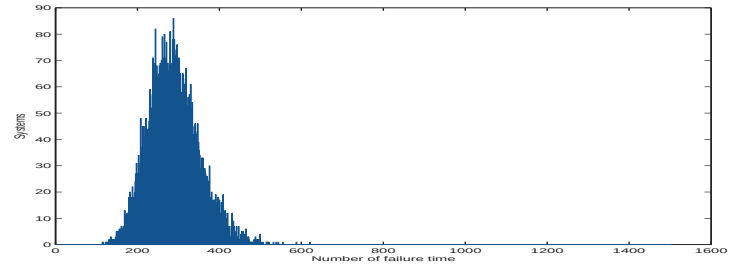

Figure 30: Histogram of Number of failure Vs Systems; λ=2, β=1.5,γ=0.04,10000 data sets and m=20 systems per data sets

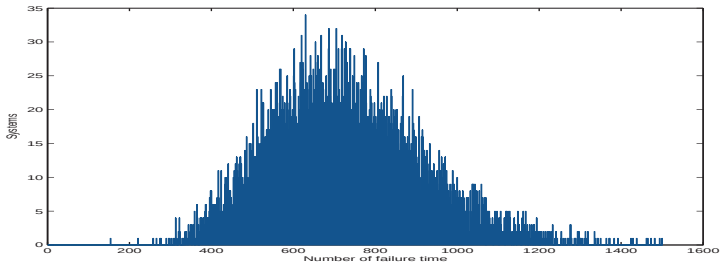

Figure 31: Histogram of Number of failure Vs Systems; λ=2, β=1.5,γ=0.06,10000 data sets and m=20 systems per data sets

2.11 Conclusion

In this paper, interrelation between fraility and dynamic models have been investigated. We have considered parameter δ as measure of fraility and γ as measure of dynamic behaviour. Moreover, these, parameters are considered as the main focus of the study and to see the difference from the baseline model. We were forced to use a smaller γ than δ to have a reasonable number of failures for dynamic behaviour. Unlike δ, as γ increases, a decrease in λ is seen, but the converse for β due to the fact that higher number of failures happen in system. Both features dynamic behaviour and frailty have great impact on analyses, avoiding wrong conclusions occurring if they are not taken into account. We have considered dynamic and fraility data sets and estimate their parameters by the fraility likelihood function. Often the true γ value and δ estimates are close to being equal. Hence, we can say that fraility models may be viewed as an alternative to dynamic models.

Literature and References

1. Aalen, O. , Borgan, O. and Gjessing, H. *Survival and Event History Analysis: A Process Point of View*. Springer, 2008.

2. Aalen, O. , Fosen, J. , Weedon-Fekjr, H., Borgan, O. and Husebye, E. , "Dynamic Analysis of Multivariate Failure Time Data", Biometrics, Vol. 60, No. 3 (Sep., 2004), pp. 764-773.

3. Babykina, G. and Couallier, V.,"Modelling Recurrent Events for Repairable Systems Under Worse Than Old Assumption", 2009.

4. Bhalotra, S. and Soest, A. V. , "Birth Spacing, Fertility and Neonatal Mortality in India: Dynamics, Frailty and Fecundity", Journal of Econometrics 143(2), pp. 274-290, April, 2008.

5. Borgan, O. , Fiaccone, R. L. , Henderson, R. , Barreto, M. L. , "Dynamic Analysis of Recurrent Event Data with Missing Observations, with Application to Infant Diarrhoea in Brazil", Scandinavian Journal of Statistics, Vol. 34, No. 1 (March 2007), pp. 53-69.

6. Coetzee, J. L. , "The role of NHPP models in the practical analysis of maintenance failure data", Reliability Engineering and System Safety 56 (1997) 161-168.

7. Cook, R. J. and Lawless, J. F. *The Statistical Analysis of Recurrent Events*. Springer, 2006.

8. Crow, L. H., "Estimation Procedures of the Duane Model", Proceedings US AMSAA Reliability Growth Symposium, aberdeen Proving Ground, MD, September 1972.

9. Crow, L. H.,"Evaluating the reliability of repairable systems", Proceedings Annual Reliability and Maintainability Symposium, IEEE, pp. 275-279, 1990.

10. Crow, L. H. , "New International Standards on reliability growth", Reliability and Maintainability Symposium. Annual Proceedings, 1991.

11. Crow, L. H., "Planning a Reliability Growth Program Utilizing Historical Data", Reliability and Maintainability Symposium, January, 2011.

12. Crow, L. H. and Shimi, I. N. , "Maximum Likelihood Estimation of Life-Distributions from Renewal Testing", The AAnnals of Mathematical Statistics, Vol. 43, No.6 (December, 1972), pp. 1827-1838.

13. Duane, J. H., "Learning curve approach tp reliability monitoring", IEEE Transactions on Aerospace 2, pp. 563-566,1964.

14. Elgmati, E. A. , "Dynamic Models versus Frailty Models for Recurrent Event Data", World Academy of Science, Engineering and Technology, issue: 0076, p.517-521, April 2013.

15. Engelhardt, M. and Lee, J. B. *Statistical Analysis of a Weibull process with left censored data*, 1992 kluwer academic publishers. Printed in the Netherlands.

16. Gamiz, M. L. , Kulasekera, K. B. , Limnios, N. , Lindqvist, B. H. *Applied Nonparametric Statistics in Reliability.* Springer, February, 2011.

17. Guilley, E. , Ghisletta, P. , Armi, F. , Berchtold, A. , Lalive C. , d'Epinay, Michel J. and Ribaupierre, A. , "Dynamics of Frailty and ADL Dependence in a Five-Year Longitudinal Study of Octogenarians, Research on Aging", originally published online 31 January 2008, DOI: 10.1177/0164027507312115.

18. Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*, Second edition, 2002.

19. Larsen, R. J. and Marx, M. L. *An Introduction to Mathematical Statistics and Its applications.* 2012, fifth edition.

20. Lawless, J. F. , "Regression Methods for Poisson Process Data. Journal of the American Statistical Association", Vol.82,No.399(Sep.,1987), pp.808-815.

21. Lee, E. T. and Wang, J. W. *Statistical methods for survival data analysis*, Third edition, 2003.

22. Le Gat, Y., "Extending the Yule process to model recurrent pipe failures in water supply networks", Urban Water Journal (2013), DOI: 10.1080/1573062X.2013.783088.

23. Meeker, W. Q. and Escobar, L. A. *Statistical Methods for Reliability Data*. July, 24, 1998.

24. Puts, M. T. E. , Lips, P. , Deeg, D. J. H. , "Static and dynamic measures of frailty predicted decline in performance-based and self-reported physical functioning", Journal of Clinical Epidemiology,vol 58, pp-1188-1198, 2005.

25. Rannestad, B., "Exact Statistical Inference in Non-homogeneous Poisson Processes, based on Simulation", Norwegian University of Science and Technology (NTNU), Department of Mathematical Sciences. July, 2007.

26. Rausand, M. and Hyland, A. *System Reliability Theory Models, statistical Methods, and Applications*, Second Edition, 2004.

27. Ruggeri, F. R. and Antonio P., "On the reliability of repairable systems: Methods and Applications", CNR-IMATI, Via A. M. Ampere 56, I-20131 Milano.

28. Ruggeri, F. and Sivaganesan, S. , "On modeling change points in Non-homogeneous Poisson processes", Statistical inference for stochastic processes (2005) 8:311-329 DOI 10.1007/s11203-005-6100-y.

29. Steenbergen, M. R. , "Maximum likelihood programming in R", January 2006.

30. Vaupel, J. , Manton K. and Stallard, E. , "The impact of heterogeneity in individual frailty on the dynamics of mortality", Demography, vol. 16, pp. 439-454, 1979.

31. Yan, J. , "Analysis of Multivariate Survival Data", Journal of the American Statistical Association, 100:469, 354-355, DOI: 10.1198/jasa.2005.s10, 2005.

Part III:

Report II

Report II

Localized/Shrinkage Kriging Predictors

Zeytu Gashaw Asfaw and Henning Omre

Preprint Series in Statistics no.1, 2014, Department of Mathematical Sciences, NTNU, Trondheim, Norway.

Abstract

The objective of the study is to improve the robustness and flexibility of spatial kriging predictors with respect to deviations from spatial stationarity assumptions. A predictor based on a non-stationary Gaussian random field is defined. The model parameters are inferred in an empirical Bayesian setting, using observations in a local neighborhood and a prior model assessed from the global set of observations. The localized predictor appears with a shrinkage effect and is coined a localized/shrinkage kriging predictor. The predictor is compared to traditional localized kriging predictors in a case study on observations of annual cumulated precipitation. A crossvalidation criterion is used in the comparision. The shrinkage predictor appears as uniformly preferable to the traditional kriging predictors. A simulation study on prediction in non-stationary Gaussian random fields is conducted. The results from this study confirms that the shrinkage predictor is favorable to the traditional ones. Moreover, the crossvalidation criterion is found to be suitable for selection of predictor. Lastly, the shrinkage predictor appears as particularly robust towards spatially varying expectation functions.

3 Localized/Shrinkage Kriging Prediction

3.1 Introduction

Consider a set of exact observations from a continuous regionalized variable. Focus is on prediction of the regionalized variable in an unobserved location with associated prediction variance. One option is to use traditional kriging prediction, see Journel and Huijbregts (1978) and Chiles and Delfiner (1999). If one assume a model with spatially constant expectation and variance with a shift-invariant spatial correlation function, then a global, ordinary kriging predictor will be a natural choice. This model assumption may be tested statistically, see Fuentes (2005).

A more flexible and robust spatial predictor can be defined by applying the ordinary kriging predictor locally. This entails using only observations in a specified finite neighborhood around the location of the variable to be predicted. This approach is termed local neighborhood kriging, see Chiles and Delfiner (1999), and it robustifies the predictor with respect to deviations from the assumption of spatially constant expectation and variance. Moreover, local neighborhood predictors can give huge computational gains in large scale problems.

The major challenge in using localized predictors is to specify the size of the neighborhoods, or the set of neighboring observations involved. Classical statistical trade-offs between bias and variance in the local predictor must be made. The spatial correlation structure may provide a screening effect by the neighboring observations, see Stein (2002), and this effect may be used to justify localization. The localized predictors can cause artifacts in the predicted regionalized variable as discontinuities when extreme observations are included or excluded in the neighborhood as it is shifted, see Gribov and Krivoruchko (2004).

The objective of this study is to improve the flexibility and robustness of the spatial predictor. We define a spatial model as a Gaussian random field with spatially varying expectation and variance. The spatial correlation function is shift invariant and known. Under these model assumptions the local neighborhood kriging predictor require expectation and variance to be assessed in the prediction and observation locations. In traditional kriging approaches

this inference is made by a sliding neighborhood maximum-likelihood estimator.

We define a new localized predictor inspired by the empirical Bayes approach discussed in Efron and Morris (1973). We phrase the inference of the spatial expectation and variance in a Bayesian setting along the lines of Røislien and Omre (2006). The conjugate prior models are assessed empirically from the global set of observations. The resulting local kriging predictor appear with shrinkage caused by the global prior model. We term the predictor as localized/shrinkage kriging.

The report is organized as follows : Section 2.2 contain a list of notation. In Section 2.3 general random field models are defined and discussed. Model parameter inference in these random field models are discussed in Section 2.4, while Section 2.5 contain definitions of the localized predictors. In Section 2.6 a presentation of the evaluation criteria is included. Section 2.7 contain a demonstration and evaluation of the predictors on a couple of real data examples, while Section 2.8 presents the empirical simulation study on Gaussian random fields. Lastly, Section 2.9 contains the conclusions from the study.

3.2 Notation

The folowing notation is used:

$\mathcal{L}_{\mathbf{D}}$	grid over $\mathbf{D}$
n	number of grid nodes in $\mathbf{D}$
$\mathcal{L}_o$	locations of sampled observations
n_o	number of sampled observation
k	number of closest observation
$\mathbf{r}$	vector of values in grid $\mathcal{L}_{\mathbf{D}}$
$\mathbf{r}_o$	vector of values of observations in $\mathcal{L}_o$
r_+	value at location x_+
$\mathbf{r}_{+o}^k$	vector of values in k-closest observations to location x_+
i_n	unit $[n \times 1]$vector
I_n	unit diagonal $[n \times n]$ matrix
μ_r	expected value
σ_r^2	variance value
Σ_{rr}	covariance matrix

Ω_r	correlation matrix
Γ_r	diagonal standard deviation matrix
H	selection matrix
ν	degree of freedom for hierarichical representation

3.3 Random field models

A random field (RF) is a generalization of a stochastic process, taking references on some topological space. Due to the complexity of natural phenomena and the actual problem, several kind of random fields are defined, among them Gaussian RF, Poisson RF and Markov RF. The former is of concern in the current study. We consider Gaussian RF $\{r(x); x \in \mathbf{D} \subset \Re\}$, where x is a reference location running over the domain $\mathbf{D}$ as a subset of $\Re^m$, with $r(x)$ being the random variable of interest.

3.3.1 General Gaussian Random Field Model

A Gaussian RF is defined by the Gaussian probability density functions. The Gaussian RF is a preferable model for continuous, or almost continuous, spatial variables due to its simplicity in inferences and analytical tractability.

The definition of a Gaussian RF is:

A RF $\{r(x); x \in \mathbf{D} \subset \Re^m\}$ is denoted a Gaussian RF if

$$\mathbf{r} = \begin{bmatrix} r(x_1) \\ \vdots \\ r(x_n) \end{bmatrix} \sim N_n(\boldsymbol{\mu}_r, \Sigma_{rr}) \tag{1}$$

for $\forall$ conf $(x_1, \cdots, x_n) \in \mathbf{D}^n$, $\forall n \geq 1$

and the corresponding pdf can be written as:

$$f(\mathbf{r}) = (2\pi)^{-\frac{n}{2}} |\Sigma_{rr}|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}(\mathbf{r} - \boldsymbol{\mu}_r)^T \Sigma_{rr}^{-1} (\mathbf{r} - \boldsymbol{\mu}_r)\right\},$$

where

$$\boldsymbol{\mu}_r = \begin{bmatrix} \mu_r(x_1) \\ \vdots \\ \mu_r(x_n) \end{bmatrix},$$

and

$$\Sigma_{rr} = \Gamma_r \Omega_{rr} \Gamma_r^T ,$$

with

$$\Gamma_r = \begin{bmatrix} \sigma_r(x_1) & 0 & \dots & 0 \\ 0 & \sigma_r(x_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r(x_n) \end{bmatrix}$$

$$\Omega_{rr} = \begin{bmatrix} \rho_r(x_1, x_1) & \rho_r(x_1, x_2) & \dots & \rho_r(x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_r(x_n, x_1) & \rho_r(x_n, x_2) & \dots & \rho_r(x_n, x_n) \end{bmatrix}$$

The model parameters for Gaussian RFs are:

$$\{\mu_r(x) = E[r(x)]; \quad x \in \mathbf{D}\} \text{ - spatial expectation field.}$$

$$\left\{\sigma_r^2(x) = Var[r(x)]; \quad x \in \mathbf{D}\right\} \text{ - spatial variance field.}$$

$$\left\{\rho_r(x', x'') = Corr\left[r(x'), r(x'')\right]; \quad x', x'' \in \mathbf{D}^2\right\} \text{ - spatial correlation field.}$$

$$\left\{\phi_r(x', x'') = Cov[r(x'), r(x'')] = \sigma_r(x')\sigma_r(x'')\rho_r(x', x''); \quad x', x'' \in \mathbf{D}^2\right\}$$

- spatial covariance field.

Hence, the model parametrization for a Gaussian RF is: $\{\mu_r(x), \sigma_r^2(x); x \in \mathbf{D}\}$ and $\left\{\rho_r(x', x''); x', x'' \in \mathbf{D}^2\right\}$. The requirements for the model parameters are:

- $\sigma_r(x) \geq 0$ for $x \in \mathbf{D}$.
- $-1 \leq \rho_r(x', x'') \leq 1$ for $x', x'' \in \mathbf{D}^2$

- Ω_{rr} - non-negative definite matrix.

Thus, a correlation field must be a positive semi-definite field to ensure that Ω_{rr} is valid. A correlation field $\rho_r(x^{'}, x^{''})$ is called positive semi-definite if the associated quadratic form is non-negative,then:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\rho_r(x_i, x_j) \geq 0$$

$\forall$ conf$(x_1, ..., x_n) \in \mathbf{D}^n$, $\forall n > 1$, $\forall \boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)^T \in \Re^n$

If in addition the quadratic form $\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\rho_r(x_i, x_j) = \mathbf{0}$ only for $\boldsymbol{\alpha} = 0\mathbf{i}_n$, then the correlation field $\rho_r(x^{'}, x^{''})$ is called positive definite. Expectation, variance and correlation fields determine all stochastic properties of a Gaussian RF.

Consider a regular grid over $\mathbf{D}$, denote it $\mathcal{L}_\mathbf{D}$ and let the number of grid nodes be n. Define the discretized Gaussian RF $\mathbf{r} = \{r(x); x \in \mathcal{L}_\mathbf{D}\}$.

Further let the expectation $[n\times 1]$ vector $\boldsymbol{\mu}_r = \{\mu_r(x); x \in \mathcal{L}_\mathbf{D}\}$, the standard deviation $[n\times n]$ matrix Γ_r be diagonal with elements $\{\sigma_r(x); x \in \mathcal{L}_\mathbf{D}\}$ and the correlation $[n \times n]$ matrix Ω_{rr} have elements $\{\rho_r(x^{'}, x^{''}); x^{'}, x^{''} \in \mathcal{L}_\mathbf{D} \times \mathcal{L}_\mathbf{D}\}$. Moreover, the covariance $[n \times n]$ matrix is $\Sigma_{rr} = \Gamma_r\Omega_{rr}\Gamma_r^T$.

Let $\mathbf{r}_o$ be a $[n_o \times 1]$ vector of observations which occur at grid locations hence at a subset of $\mathcal{L}_\mathbf{D}$,

$$\mathbf{r}_o = \begin{bmatrix} r(x_{o1}) \\ r(x_{o2}) \\ \vdots \\ r(x_{on_o}) \end{bmatrix} = \begin{bmatrix} r_{o1} \\ r_{o2} \\ \vdots \\ r_{on_o} \end{bmatrix}$$

Hence, $[\mathbf{r}_o \mid \mathbf{r}] = H\mathbf{r}$, where H is a binary selection $[n_o \times n]$ matrix having one on all sampled location and zero in unsampled location.

Consider the combined vector of values at grid nodes and observations:

$$\begin{bmatrix} \mathbf{r} \\ \mathbf{r}_o \end{bmatrix} \sim N_{n+n_o} \left[\begin{bmatrix} \boldsymbol{\mu}_r \\ H\boldsymbol{\mu}_r \end{bmatrix}, \begin{bmatrix} \Sigma_{rr} & \Sigma_{rr}H \\ H\Sigma_{rr} & H\Sigma_{rr}H^T \end{bmatrix} \right]$$

Hence, the conditional Gaussian RF is

$$[\mathbf{r} \mid \mathbf{r}_o] \sim N_n(\boldsymbol{\mu}_{\mathbf{r}|\mathbf{r}_o}, \Sigma_{\mathbf{rr}|\mathbf{r}_o}). \tag{2}$$

where

$$\boldsymbol{\mu}_{\mathbf{r}|\mathbf{r}_o}=\mathrm{E}[\mathbf{r} \mid \mathbf{r}_o] = \boldsymbol{\mu}_r + \Sigma_{rr}H^T \left[H\Sigma_{rr}H^T\right]^{-1} (\mathbf{r}_o - H\boldsymbol{\mu}_r).$$

$$\Sigma_{\mathbf{rr}|\mathbf{r}_o}=\mathrm{Var}[\mathbf{r} \mid \mathbf{r}_o] = \Sigma_{rr} - \Sigma_{rr}H^T \left[H\Sigma_{rr}H^T\right]^{-1} H\Sigma_{rr}$$

Consider prediction of the value in an arbitrary location $x_+ \in \mathbf{D}$, and denote it $r_+ = r(x_+)$. The avaliable observations are $\mathbf{r}_o$.

$$\begin{bmatrix} r_+ \\ \mathbf{r}_o \end{bmatrix} \sim N_{1+n_o} \left[\begin{bmatrix} \mu_+ \\ H\boldsymbol{\mu}_r \end{bmatrix}, \begin{bmatrix} \sigma_+^2 & \sigma_+\omega_{o+}^T H\Gamma_r H^T \\ \sigma_+ H\Gamma_r H\omega_{o+} & H\Gamma_r\Omega_{rr}\Gamma_r H^T \end{bmatrix} \right]$$

$$= N_{1+n_o} \left[\begin{bmatrix} \mu_+ \\ \boldsymbol{\mu}_o \end{bmatrix}, \begin{bmatrix} \sigma_+^2 & \sigma_+\omega_{o+}^T\Gamma_o \\ \sigma_+\Gamma_o\omega_{o+} & \Gamma_o\Omega_{oo}\Gamma_o^T \end{bmatrix} \right]$$

where $\mu_+ = \mu(x_+)$, $\sigma_+^2 = \sigma^2(x_+)$ and

$$\boldsymbol{\mu}_o = \begin{bmatrix} \mu(x_{o1}) \\ \vdots \\ \mu(x_{on_o}) \end{bmatrix} = H\boldsymbol{\mu}_r$$

$$\omega_{o+} = \begin{bmatrix} \rho(x_{o1}, x_+) \\ \vdots \\ \rho(x_{on_o}, x_+) \end{bmatrix}$$

$$\Gamma_o = \begin{bmatrix} \sigma(x_{o1}) & 0 & \dots & 0 \\ 0 & \sigma(x_{o2}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma(x_{on_o}) \end{bmatrix} = H\Gamma_r H^T$$

$$\Omega_{oo} = \begin{bmatrix} \rho(x_{o1}, x_{o1}) & \rho(x_{o1}, x_{o2}) & \dots & \rho(x_{o1}, x_{on_o}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(x_{on_o}, x_{o1}) & \rho(x_{on_o}, x_{o2}) & \dots & \rho(x_{on_o}, x_{on_o}) \end{bmatrix} = H\Omega_{rr}H^T$$

and let $\Sigma_{oo} = \Gamma_o \Omega_{oo} \Gamma_o^T$.

The conditional Gaussian random variable in location x_+ given the observation vector $\mathbf{r}_o$ is:

$$[r_+ \mid \mathbf{r}_o] \sim N_1 \left[\mu_{+|o}, \sigma^2_{+|o}\right]$$

with

$$\begin{aligned} \mu_{+|o} &= \mu_+ + \sigma_+ \omega_{o+}^T \Gamma_o \left[\Gamma_o \Omega_{oo} \Gamma_o\right]^{-1} \left[\mathbf{r}_o - \boldsymbol{\mu}_o\right] \\ &= \mu_+ + K_G[\mathbf{r}_o - \boldsymbol{\mu}_o] \end{aligned}$$

$$\begin{aligned} \sigma^2_{+|o} &= \sigma^2_+ - \sigma_+ \omega_{o+}^T \Gamma_o \left[\Gamma_o \Omega_{oo} \Gamma_o\right]^{-1} \Gamma_o \omega_{o+} \sigma_+ \\ &= \sigma^2_+ [1 - \omega_{o+}^T \Gamma_o \left[\Gamma_o \Omega_{oo} \Gamma_o\right]^{-1} \Gamma_o \omega_{o+}] \\ &= \sigma^2_+ - K_G \Gamma_o \Omega_{oo} \Gamma_o K_G^T \end{aligned}$$

where $K_G^T = \sigma_+ \omega_{o+}^T \Gamma_o \left[\Gamma_o \Omega_{oo} \Gamma_o\right]^{-1}$ is a weight $[n_o \times 1]$ vector.
Note that the weight vector, termed generalized kriging weights, are functions of the variances in all the locations involved in the prediction and the correlation structure.

Simulation of a Gaussian RF
Consider a conditional discretized Gaussian RF represented by $[\mathbf{r} \mid \mathbf{r}_o]$ with parameters as in Equation (1). We may want to generate a sample of $[\mathbf{r} \mid \mathbf{r}_o]$ from the conditional model. This simulated surface can be generated as follows:

1. Cholesky factorization of $\Sigma_{rr|r_o} = LL^T$ where L and L^T are lower/upper triangular $[n \times n]$ matrices.
2. Sample $\boldsymbol{z} \sim N_n(\mathbf{0}, I_n)$
3. Compute $\boldsymbol{v} = L\boldsymbol{z}$
4. Compute $[\mathbf{r} \mid \mathbf{r}_o] = \boldsymbol{\mu}_{r|r_o} + \boldsymbol{v}$
5. Return $[\mathbf{r} \mid \mathbf{r}_o]$

3.3.2 Stationary Gaussian RF Model

Stationarity is a property of a regionalized variable that has shift invariant statistical properties within the area of interest. Stationary random field is a random field whose joint probability distribution does not change when shifted in location. Prediction in stationary Gaussian RF is termed either simple or ordinary kriging dependent on wheather the expectation is known or must be estimated.

A RF $\{r(x); x \in \mathbf{D}\}$ is defined to be stationary if it satisfies:

$$\{\mu_r(x) = E[r(x)] = \mu_r; \quad x \in \mathbf{D}\}$$
$$\left\{\sigma_r^2(x) = Var[r(x)] = \sigma_r^2; \quad x \in \mathbf{D}\right\}$$
$$\left\{\rho_r(x^{'}, x^{''}) = Corr\left[r(x^{'}), r(x^{''})\right] = \rho_r(x^{'} - x^{''}); \quad x^{'}, x^{''} \in \mathbf{D}^2\right\}$$
$$\left\{\phi_r(x^{'}, x^{''}) = Cov\left[r(x^{'}), r(x^{''})\right] = \sigma_r^2\rho_r(x^{'} - x^{''}); \quad x^{'}, x^{''} \in \mathbf{D}^2\right\}$$

A Gaussian RF having stationary model parameters is said to be stationary Gaussian RF. The predictor in $r_+ = r(x_+)$, in a stationary Gaussian RF is defined by the expression:

$$\begin{bmatrix} r_+ \\ \boldsymbol{r}_o \end{bmatrix} \sim N_{1+n_o}\left[\begin{bmatrix} \mu_r \\ \mu_r \boldsymbol{i}_{n_o} \end{bmatrix}, \begin{bmatrix} \sigma_r^2 & \sigma_r^2\omega_{o+}^T \\ \omega_{o+}\sigma_r^2 & \sigma_r^2\Omega_{oo} \end{bmatrix}\right]$$

hence,

$$[r_+ \mid \boldsymbol{r}_o] \sim N_1\left[\mu_{r|o}, \sigma^2_{r|o}\right]$$

$$\mu_{r|o} = \mu_r + \sigma_r^2 \omega_{o+}^T \left[\sigma_r^2 \Omega_{oo}\right]^{-1} \left[\boldsymbol{r}_o - \mu_r i_{n_o}\right]$$

$$= \mu_r + K_S \left[\boldsymbol{r}_o - \mu_r i_{n_o}\right]$$

$$\sigma^2_{r|o} = \sigma_r^2 - \sigma_r^2 \omega_{o+}^T \left[\sigma_r^2 \Omega_{oo}\right]^{-1} \omega_{o+} \sigma_r^2$$

$$= \sigma_r^2 \left[1 - K_S \Omega_{oo} K_S^T\right]$$

where $K_S^T = \omega_{o+}^T \Omega_{oo}^{-1}$, is a weight $[n_o \times 1]$ vector. Note that the weight vector, termed stationary (simple) kriging weights, is dependent on the correlation structure only, not on the variance.

3.3.3 Hierarchical Stationary Gaussian Random Field

For a stationary Gaussian RF to be fully specified, the model parameters μ_r, σ_r^2 and $\rho_r(.)$ need to be known. In the hierarchical representation we let μ_r and σ_r^2 be represented by random variables m and s^2 while $\rho_r(.)$ is considered to be known. By conditioning on $[m, s^2]$, the Gaussian RF is fully specified.

Consider the discrete representation of stationary Gaussian RF. Let $\mathbf{r}$ conditional on the random parameters $[m, s^2]$ be distributed as

$$[\mathbf{r} \mid m, s^2] \sim N_n(m i_n, s^2 \Omega_{rr})$$

where, m and s^2 are univariate random variables and Ω_{rr} is a known positive definite correlation $[n \times n]$ matrix.

Moreover, assume the following prior model for $[m\ ,\ s^2]$:

$$[m \mid s^2] \sim N_1(\mu_m, \tau_m s^2) \tag{3}$$

$$s^2 \sim IG(\xi_s, \gamma_s) \tag{4}$$

with $\mu_m \in \Re$, $\tau_m \in \Re_+$ and $IG(\xi_s, \gamma_s)$ representing the inverse gamma pdf

$$f(s^2) = \frac{1}{\Gamma(\xi_s)} \gamma_s^{\xi_s} [s^2]^{-(\xi_s+1)} exp\left\{-\gamma_s [s^2]^{-1}\right\},\ s^2 > 0.$$

where $\Gamma(x)$ is the gamma function, $\xi_s \in \Re_+$ is a shape parameter and $\gamma_s \in \Re_+$ is a scale parameter.

Hence, it can be demonstrated, see Røislien and Omre (2006) that:

$$\mathbf{r} \sim T_n(\mu_m i_n, \Sigma_{rr}, \nu)$$

represent a T-dist RF defined by the multivariate T-distribution:

$$f(\mathbf{r}) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{n}{2}}} \left|\Sigma_{rr}\right|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu}\left[\mathbf{r} - \mu_m i_n\right]^T \Sigma_{rr}^{-1} \left[\mathbf{r} - \mu_m i_n\right]\right]^{-\frac{\nu+n}{2}}$$

where $\nu \in \Re_+$ is the degrees of freedom which is defined by $\nu = 2\xi_s$. This definition specifies a spherical-symmetric pdf centered at $\mu_m i_n$ with Ω_{rr} controlling scale and multivariate dependence, while ν controls the tail behavior(Mardia et al., 1979).

Multivariate Gaussian and Cauchy distributions are special cases of multivariate T-distributions. That is,

$$T_n(\mu_m i_n, \Sigma_{rr}, \nu) \xrightarrow{\nu \to \infty} N_n(\mu_m i_n, \Sigma_{rr}) \text{ - Gaussian distribution.}$$

$$T_n(\mu_m i_n, \Sigma_{rr}, 1) = C_n(\mu_m i_n, \Sigma_{rr}) \text{ - Cauchy distribution.}$$

This hierarchical representation can be interpreted in a Bayesian setting with $[m, s^2]$ being random hyperparameters.

Consider a set of observations $\mathbf{r}_o$ as previously defined. The posterior model for the model parameters $[m, s^2 \mid \mathbf{r}_o]$ can then be determined. From the definition of hierachical stationary Gaussian RF one has:

$$[\mathbf{r}_o \mid m, s^2] \sim N_{n_o}(m i_{n_o}, s^2 \Omega_{oo})$$

and using the prior model for $[m \mid s^2]$, in Equation (3), one obtains:

$$\begin{bmatrix} \mathbf{r}_o \\ m \end{bmatrix} \mid s^2 \sim N_{n_o+1} \left[\begin{bmatrix} \mu_m i_{n_o} \\ \mu_m \end{bmatrix}, \begin{bmatrix} \tau_m s^2 i_{n_o} i_{n_o}^T + s^2 \Omega_{oo} & \tau_m s^2 i_{n_o} \\ \tau_m s^2 i_{n_o}^T & \tau_m s^2 \end{bmatrix} \right]$$

Consequently,

$$[m \mid \mathbf{r}_o, s^2] \sim N_1 \left[\mu_{m|o}, \sigma^2_{m|o} \right]$$

with

$$\begin{aligned} \mu_{m|o} &= \mu_m + \tau_m s^2 i_{n_o}^T \left[\tau_m s^2 i_{n_o} i_{n_o}^T + s^2 \Omega_{oo} \right]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}) \\ &= \mu_m + \tau_m i_{n_o}^T \left[\tau_m i_{n_o} i_{n_o}^T + \Omega_{oo} \right]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}). \end{aligned}$$

$$\sigma^2_{m|o} = \tau_m s^2 - \tau_m s^2 i_{n_o}^T \left[\tau_m s^2 i_{n_o} i_{n_o}^T + s^2 \Omega_{oo} \right]^{-1} \tau_m s^2 i_{n_o}$$

Note that $\mu_{m|o} = E[m \mid \mathbf{r}_o, s^2] = E[m \mid \mathbf{r}^o]$ and hence independent of s^2. Similarly, the marginal pdf of $[\mathbf{r}_o \mid s^2]$ is,

$$[\mathbf{r}_o \mid s^2] \sim N_{n_o} \left[\mu_m i_{n_o}, \tau_m s^2 i_{n_o} i_{n_o}^T + s^2 \Omega_{oo} \right]$$

and using the prior model for s^2, in Equation (4), one obtains, see Appendix A:

$$[s^2 \mid \mathbf{r}_o] \sim IG(\xi_{s|o}, \gamma_{s|o})$$

with

$$\xi_{s|o} = \xi_s + \tfrac{n_o}{2}$$

$$\gamma_{s|o} = \gamma_s + \tfrac{1}{2} \left[[\mathbf{r}_o - \mu_m i_{n_o}]^T \left[\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T \right]^{-1} [\mathbf{r}_o - \mu_m i_{n_o}] \right]$$

Note that from the characteristics of the inverse Gamma distribution we have:

$$\mu_{s|o} = E[s^2 \mid \mathbf{r}_o] = \tfrac{\gamma_{s|o}}{\xi_{s|o}-1}, \quad \xi_{s|o} > 1$$

$$\sigma^2_{s|o} = Var[s^2 \mid \mathbf{r}_o] = \tfrac{\gamma^2_{s|o}}{(\xi_{s|o}-1)^2(\xi_{s|o}-2)}, \quad \xi_{s|o} > 2$$

3.4 Model Parameter Inference

Both the stationary Gaussian RF model, see Section 3.2, and the hierarchical, stationary Gaussian RF model, see Section 3.3, depends on a set of model parameters. These model parameters must be assessed from the avaliable observations $\mathbf{r}_o$ in order to make the respective models operable.

In the study we also use localized estimators for the model parameters. Consider location $x_+ \in \mathbf{D}$ and define the $[k \times 1]$ vector of observations:

$$\mathbf{r}^k_{+o} = G^k_+ \mathbf{r}_o$$

where G^k_+ is a binary $[k \times n_o]$ matrix which selects the k observations located closest to x_+. The selection may also include some symmetry criteria.

3.4.1 Stationary Gaussian RF model

The actual set of model parameters are $[\mu_r, \sigma^2_r, \rho_r(\tau)]$. We consider the spatial correlation function $\rho_r(\tau)$ to be known, hence the expected value μ_r and variance value σ^2_r must be assessed from $\mathbf{r}_o$.

We choose to use a maximum likelihood criterion in the assessment, and the log-likelihood function is:

$$\begin{aligned} l(\mu_r, \sigma^2_r; \mathbf{r}_o) = &-\frac{n_o}{2}\log(2\pi) - \frac{n_o}{2}\log(\sigma^2_r) \\ &- \frac{1}{2}\log|\Omega_{oo}| - \frac{1}{2}[\sigma^2_r]^{-1}\left[(\mathbf{r}^o - \mu_r i_{n_o})^T \Omega^{-1}_{oo}(\mathbf{r}_o - \mu_r i_{n_o})\right] \end{aligned}$$

Hence the maximum likelihood estimates are:

$$\hat{\mu}_r = [i^T_{n_o}\Omega^{-1}_{oo}\mathbf{r}_o][i^T_{n_o}\Omega^{-1}_{oo}i_{n_o}]^{-1}$$

$$\hat{\sigma^2_r} = \tfrac{1}{n_o}(\mathbf{r}_o - \hat{\mu}_r i_{n_o})^T \Omega^{-1}_{oo}(\mathbf{r}_o - \hat{\mu}_r i_{n_o})$$

The corresponding localized estimators of $[\mu_r, \sigma^2_r]$ centered at location $x_+ \in \mathbf{D}$ based on $\mathbf{r}^k_{+o} = G^k_+ \mathbf{r}_o$ are:

$$\hat{\mu}^k_+ = [i^T_k[G^k_+\Omega_{oo}[G^k_+]^T]^{-1}G^k_+\mathbf{r}_o][i^T_k[G^k_+\Omega_{oo}[G^k_+]^T]^{-1}i_k]^{-1}$$

$$\hat{\sigma}^{k2}_+ = \tfrac{1}{k}(G^k_+\mathbf{r}_o - \hat{\mu}^k_+ i_k)^T[G^k_+\Omega_{oo}[G^k_+]^T]^{-1}(G^k_+\mathbf{r}_o - \hat{\mu}^k_+ i_k)$$

Define also the expectation $[n_o \times 1]$ vector centered at the observation locations:

$$\hat{\boldsymbol{\mu}}_o^k = \begin{bmatrix} \hat{\mu}_1^k \\ \vdots \\ \hat{\mu}_{n_o}^k \end{bmatrix},$$

and the corresponding standared deviation diagonal $[n_o \times n_o]$ matrix

$$\hat{\Gamma}_o^k = \begin{bmatrix} \hat{\sigma}_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\sigma}_{n0}^k \end{bmatrix}$$

3.4.2 Hierarchical, stationary Gaussian RF model

The actual set of model parameters is $[\mu_m, \tau_m, \xi_s, \gamma_s, \rho_r(\tau)]$. We consider the spatial correlation function $\rho_r(\tau)$ to be known, hence the prior model parameters for expectation $[\mu_m, \tau_m]$ and for variance $[\xi_s, \gamma_s]$ must be assessed from $\mathbf{r}_o$.

We choose to make this assessment in an empirical Bayes setting based on the observations $\mathbf{r}_o$. The k-closest localization is used to define a set of localizations centered at the observations over the domain $\mathbf{D}$. This set is considered to be a super-population from which the k-closest prior model is assessed.

The estimates for the Gaussian prior model parameters for expectation are:

$$\hat{\mu}_m^k = \tfrac{1}{n_o} i_{n_o}^T \hat{\boldsymbol{\mu}}_o^k$$

$$\hat{\sigma}_m^{k2} = \tfrac{1}{n_o} [\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k i_{n_o}]^T [\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k i_{n_o}]$$

$$\hat{\sigma}_{r|.}^{k2} = \tfrac{1}{n_o} Tr[\hat{\Gamma}_o^k]$$

$$\hat{\tau}_m^k = \tfrac{\hat{\sigma}_m^{k2}}{\hat{\sigma}_{r|.}^{k2}}$$

The corresponding localized estimators for the posterior expectation m centered at location $x_+ \in \mathbf{D}$ based on observations $\mathbf{r}_{+o}^k = G_+^k \mathbf{r}_o$ is:

$$\begin{aligned} \hat{m}_+^k &= E[m \mid s^2, G_+^k \mathbf{r}_o] \\ &= \hat{\mu}_m^k + \hat{\tau}_m^k i_k^T \left[\hat{\tau}_m^k i_k i_k^T + [G_+^k \Omega_{oo} [G_+^k]^T\right]^{-1} \left[G_+^k \mathbf{r}_o - \hat{\mu}_m^k i_k\right] \end{aligned}$$

which is independent of s^2.

Define also the expectation $[n_o \times 1]$ vector centered at the observation locations:

$$\hat{\boldsymbol{m}}_o^k = \begin{bmatrix} \hat{m}_1^k \\ \vdots \\ \hat{m}_{n_o}^k \end{bmatrix},$$

The estimates for the inverse gamma prior model parameters for variance are more complicated. Note first that the prior expectation and variance for $\xi_s > 2$ are :

$$\mu_s = E[s^2] = \frac{\gamma_s}{\xi_s - 1}$$

$$\sigma_s^2 = Var[s^2] = \frac{\gamma_s^2}{[\xi_s - 1]^2[\xi_s - 2]}$$

Consquently,

$$\xi_s = \frac{\mu_s^2}{\sigma_s^2} + 2$$

$$\gamma_s = \mu_s \left[\frac{\mu_s^2}{\sigma_s^2} + 1\right]$$

Define the $[n_o \times 1]$ vector defined for a k-neighborhood

$$\mathbf{s}^2 = \begin{bmatrix} (r_{o1} - \hat{\mu}_m^k)^2 \\ \vdots \\ (r_{on_o} - \hat{\mu}_m^k)^2 \end{bmatrix}$$

The two first moments are estimated by:

$$\hat{\mu}_s = \frac{1}{n_o} i_{n_o}^T \mathbf{s}^2$$

$$\hat{\sigma}_s^2 = \frac{1}{n_o} [\mathbf{s}^2 - \hat{\mu}_s i_{n_o}]^T [\mathbf{s}^2 - \hat{\mu}_s i_{n_o}]$$

The prior model estimates $\hat{\xi}_s$ and $\hat{\gamma}_s$ are obtained by inserting $\hat{\mu}_s$ and $\hat{\sigma}_s^2$ into the expressions above.

The corresponding localized estimator for the posterior variance s^2 centered at location $x_+ \in \mathbf{D}$ based on observations $\mathbf{r}^k_{+o} = G^k_+ \mathbf{r}_o$ is:

$$\begin{aligned}
\hat{s}^{k2}_+ &= E[s^2 \mid G^k_+ \mathbf{r}_o] \\
&= \frac{\hat{\gamma}_{s|o}}{\hat{\xi}_{s|o} - 1} \\
&= \frac{\hat{\gamma}_s + \frac{1}{2}\left[[G^k_+ \mathbf{r}_o - \hat{\mu}^k_m i_k]^T \left[[G^k_+ \Omega_{oo} [G^k_+]^T] + \hat{\tau}^k_m i_k i^T_k \right]^{-1} [G^k_+ \mathbf{r}_o - \hat{\mu}^k_m i_k] \right]}{\hat{\xi}_s + \frac{k}{2} - 1}
\end{aligned}$$

Define also the diagonal standard deviation $[n_o \times n_o]$ matrix centered at the observation locations

$$\hat{S}^k_o = \begin{bmatrix} \hat{s}^k_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{s}^k_{n0} \end{bmatrix}$$

3.5 Prediction Models

The objective of the study is to define improved spatial predictors, and we consider localized predictors which only utilizes observations in a neighborhood of the location in focus for prediction.Two model types are defined: localized/stationary [Loc/Stat] model and localized/non-stationary [Loc/Nonstat] model. For each model type we consider a traditional [Trad] predictor and a shrinkage [Shr] predictor.

Focus is on predicting $r(x_+) = r_+$ in arbitrary location $x_+ \in \mathbf{D}$. The prediction is based on the observation $[n_o \times 1]$ vector $\mathbf{r}_o = [r(x_{o1}), \cdots, r(x_{on_o})] = [r_{o1}, \cdots, r_{on_o}]$. Define also the binary, selection $[k \times n_o]$ matrix G^k_+ which selects the k closest observations to location x_+. Note that G^k_+ may also include some symmetry criteria.

3.5.1 Localized/Stationary Model

The predictor is based on the stationary Gaussian RF model, see Section 3.2, with the model parameters assessed in two different localized ways.

3.5.1.1 Traditional Predictor

The Loc/Stat/Trad predictor of r_+ with associated prediction variance is defined as:

$$\hat{r}^k_{STP+} = \hat{\mu}^k_+ + [G^k_+\omega_{o+}]^T[G^k_+\Omega_{oo}[G^k_+]^T]^{-1}\left[G^k_+\mathbf{r}_o - \hat{\mu}^k_+ i_k\right]$$

$$\hat{\sigma}^{k2}_{STP+} = \hat{\sigma}^{k2}_+\left[1 - [G^k_+\omega_{o+}]^T[G^k_+\Omega_{oo}[G^k_+]^T]^{-1}G^k_+\omega_{o+}\right]$$

with the parameter estimators defined in Section 4.1.

The expectation μ_+ and the variance σ^2_+ are estimated by maximum likelihood in a neighborhood of x_+, hence the predictor appears like a localized ordinary kriging predictor. This corresponds to the traditional approach to localized spatial interpolation, see Chiles and Delfiner (1999). The challenge is to define the size of the neighborhood to obtain a suitable bias-variance trade-off. The neighborhood must be small to adopt to possible spatially varying expectation /variance functions and large to contain enough observations to provide stable estimates.

3.5.1.2 Shrinkage Predictor

The Loc/Stat/Shr predictor of r_+ with associated prediction variance is defined as:

$$\hat{r}^k_{SSP+} = \hat{m}^k_+ + [G^k_+\omega_{o+}]^T[G^k_+\Omega_{oo}[G^k_+]^T]^{-1}\left[G^k_+\mathbf{r}_o - \hat{m}^k_+ i_k\right]$$

$$\hat{\sigma}^{k2}_{SSP+} = \hat{s}^{k2}_+\left[1 - [G^k_+\omega_{o+}]^T[G^k_+\Omega_{oo}[G^k_+]^T]^{-1}G^k_+\omega_{o+}\right]$$

with the parameters estimators defined in Section 4.2.

The expectation m_+ and the variance s^2_+ are estimated in a Bayesian setting as the posterior expectations given the observations in a neighborhood of x_+. Hence the prior model acts like a regulizer when estimating the local expectation and variance. The prior models for m_+ and s^2_+ are assessed from the avaliable observations in an empirical Bayesian setting. This makes it possible to use smaller neighborhoods which hopefully provides predictors with less bias.

3.5.2 Localized/Non-stationary Model

The predictor is based on the general Gaussian RF model, see Section 3.1, with the model parameters assessed in two different localized ways.

3.5.2.1 Traditional Predictor

The Loc/Non-stat/Trad predictor of r_+ with associated prediction variance is defined as:

$$\hat{r}^k_{NTP+} = \hat{\mu}^k_+ + \hat{\sigma}^k_+ [G^k_+ \hat{\Gamma}^k_o \omega_{o+}]^T \left[G^k_+ \hat{\Gamma}^k_o \Omega_{oo} \hat{\Gamma}^k_o [G^k_+]^T \right]^{-1} G^k_+ [\mathbf{r}_o - \hat{\boldsymbol{\mu}}^k_o]$$

$$\hat{\sigma}^{k2}_{NTP+} = \hat{\sigma}^{k2}_+ \left[1 - [G^k_+ \hat{\Gamma}^k_o \omega_{o+}]^T \left[G^k_+ \hat{\Gamma}^k_o \Omega_{oo} \hat{\Gamma}^k_o [G^k_+]^T] \right]^{-1} G^k_+ \hat{\Gamma}^k_o \omega_{+o} \right]$$

with the parameter estimators defined in Section 4.1.

The expectation and variance is locally and uniquely estimated for each observation according to the General Gaussian RF model in Section 3.1. The number of parameter estimates is $2(n_o + 1)$, expectation and variance for x_+ and all observation locations. Hence the predictor is very sensitive to the estimate precision, which favors large neighborhoods. Large neighborhoods will however introduce larger bias in the predictor, which is unfavorable.

3.5.2.2 Shrinkage Predictor

The Loc/Non-stat/Shr predictor of r_+ with associated prediction variance is defined as:

$$\hat{r}^k_{NSP+} = \hat{m}^k_+ + \hat{s}^k_+ [G^k_+ \hat{S}^k_o \omega_{o+}]^T \left[G^k_+ \hat{S}^k_o \Omega_{oo} \hat{S}^k_o [G^k_+]^T \right]^{-1} G^k_+ [\mathbf{r}_o - \hat{\boldsymbol{m}}^k_o]$$

$$\hat{\sigma}^{k2}_{NSP+} = \hat{s}^{k2}_+ \left[1 - [G^k_+ \hat{S}^k_o \omega_{o+}]^T \left[G^k_+ \hat{S}^k_o \Omega_{oo} \hat{S}^k_o [G^k_+]^T \right]^{-1} G^k_+ \hat{S}^k_o \omega_{o+} \right]$$

with the parameter estimators defined in Section 4.2.

The expectation and variance for both x_+ and all observation locations are assessed in a Bayesian setting as conditional expectations given the observations in a neighborhood. Hence the empirical prior model for expectation

and variance act as regulizers in the inference. Note that this regularization will influence each individual kriging weight under this model. Hence this can be seen as a truly shrinkage kriging predictor. This regularized approach makes it possible to use smaller neighborhoods which hopefully entails less biased predictions.

3.5.3 CrossValidation Calibrated Predictor

Consider a specific predictor with associated prediction variance in arbitrary location $x_+ \in \mathbf{D}$:

$$\hat{r}_{\wedge +} = \hat{\mu}_+ = \hat{E}[r(x_+) \mid \mathbf{r}_o]$$

$$\hat{\sigma}^2_{\wedge +} = \hat{Var}[r(x_+) \mid \mathbf{r}_o]$$

based on the observations $\mathbf{r}_o = [r(x_{o1}), ..., r(x_{on_o})]$.

Define the crossvalidation predictions:

$$\hat{r}_{\wedge oi} = \hat{E}[r(x_{oi}) \mid \mathbf{r}_{o(-i)}]; \quad i = 1, ..., n_o$$

$$\hat{\sigma}^2_{\wedge oi} = \hat{Var}[r(x_{oi}) \mid \mathbf{r}_{o(-i)}]$$

where $\mathbf{r}_{o(-i)}$ entails $\mathbf{r}_o$ with observation i removed.

The corresponding normalized crossvalidation errors are:

$$\Delta_{\wedge i} = \frac{r(x_{oi}) - \hat{r}_{\wedge oi}}{\hat{\sigma}_{\wedge oi}}; \quad i = 1, ..., n_o$$

and define the two first moments:

$$\Delta_\wedge = \frac{1}{n_o} \sum_{i=1}^{n_o} \Delta_{\wedge i}$$

$$\kappa^2_\wedge = \frac{1}{n_o} \sum_{i=1}^{n_o} [\Delta_{\wedge i} - \Delta_\wedge]^2$$

Note that both $\Delta_\wedge$ and $\kappa^2_\wedge$ can be calculated, and that for a reliable predictor we want $\Delta_\wedge \approx 0$ and $\kappa^2_\wedge \approx 1$. Actually, we can adjust the predictor with associated prediction variance such that the global normalized crossvalidation statistics are exactly 0 and 1.

Define the crossvalidation calibrated (CVC) predictor with associated prediction variance at an arbitrary location $x_+ \in \mathbf{D}$:

$$\tilde{r}_{\wedge+} = \hat{r}_{\wedge+} + \hat{\sigma}_{\wedge+}\Delta_{\wedge}$$

$$\tilde{\sigma}^2_{\wedge+} = \kappa^2_{\wedge}\hat{\sigma}^2_{\wedge+}$$

where the associated normalized crossvalidation errors will have the two first moments equal to 0 and 1 respectively.

Hence we are able to construct a predictor which always reproduce the favored values for the global crossvalidation statistics. The CVC predictor can be seen as a globally centered and scale corrected version of the original predictor. These calibrations are particularly beneficial for localized predictors which often lack global references.

3.6 Evaluation Criteria

The test criteria are based on the CVC predictors with associated prediction variances, see Section 5.3:

$$\tilde{r}_{\wedge+} = \tilde{E}[r(x_+) \mid \mathbf{r}_o]$$

$$\tilde{\sigma}^2_{\wedge+} = \tilde{Var}[r(x_+) \mid \mathbf{r}_o]$$

The corresponding crossvalidation predictions are:

$$\tilde{r}_{\wedge oi} = \tilde{E}[r(x_{oi}) \mid \mathbf{r}_{o(-i)}]; \quad i = 1, ..., n_o$$

$$\tilde{\sigma}^2_{\wedge oi} = \tilde{Var}[r(x_{oi}) \mid \mathbf{r}_{o(-i)}]$$

The CVC predictions will be globally centered and scaled with respect to the normalized cross-validation errors. This means however that large deviations in predictions may be compensated by a large estimated prediction variance.

A prediction criterion which favors precise predictions is mean squared cross-validation error non-normalized:

$$PMSE = \frac{1}{n_o}\sum_{i=1}^{n_o}[r(x_{oi}) - \tilde{r}_{\wedge oi}]^2$$

We favor small values for this prediction criterion of course.

The CVC prediction variances are also globally scaled. This scaling does not ensure close agrement between large observed deviation in predictions and large estimated prediction variances, however. This agreement is indicated by the normalized crossvalidation errors being close to one, not only unity in average.

A prediction variance criterion which favors agreement between observed prediction deviations and estimated prediction variances is :

$$VMSE = \frac{1}{n_o} \sum_{i=1}^{n_o} \left[\left[\frac{r(x_{oi}) - \tilde{r}_{\wedge oi}}{\tilde{\sigma}_{\wedge oi}} \right]^2 - 1 \right]^2$$

We favor small values for this prediction variance criterion of course.

3.7 Case Studies

Two case studies are presented: US precipitation and Gamma-log data. The former consists of observations in a number of locations in the US. The latter contain observations very densely located along a vertical subsurface profile.

3.7.1 US precipitation

We consider a data set of yearly accumulated precipitation in locations in an area in the US, see Figure 1. The study area contains 1001 locations with observations. The data is a subset of a much larger spatio-temporal data base http://www.image.ucar.edu/GSP/Data/US.monthly.met/, see also Craig et al (2003), and we use data from 1997 in an subarea in the south-east US.

By inspecting the data in Figure 1.b there appears to be a slight increase in values in the south-east direction, but the observation density is very high. The empirical spatial correlation function is displayed in Figure 2, and we fit a generalized exponential correlation function:

$$\rho_r(\tau) = exp\left\{ - \left| \frac{\tau}{3.5} \right|^{1.4} \right\} \quad ; \quad \tau \geq 0$$

which represents a fairly smooth precipitation surface. This correlation function is used throughout the study.

We aim at demonstrating and evaluating the predictors defined in Section 5. The evaluation criteria presented in Section 6 are used.

The Global ordinary kriging predictor, using all 1001 observations under a stationary model with unknown expectation/variance, is used as reference. In Figure 3 the results from the corresponding CVC predictor is presented. The crossvalidation predictions in the observation locations and the associated prediction standard deviations are displayed. These predictions results from predictions based on the global data set with the observation in the actual location removed. We observe that the predictions appear with similar patterns as the observations in Figure 1, and that the standard deviations are fairly constant across the area. The standard deviations are only dependent on the location configuration of the observations used in the predictor, not on the actually observed values. This explains the somewhat higher values along the boundary since the location configurations are unfavourably, asymmetric along the boundary. In Figure 4 the corresponding normalized crossvalidation errors and the resulting histogram are displayed. Recall that the use of the CVC predictor ensures that these errors are globally centered to zero and scaled to one. From the figure we observe some larger errors in areas with high-value predictions and that the histogram appears as somewhat peaked with heavy tails. These effects may indicate that the variance of the observations does vary across the area.

We study localized kriging predictors with a $k = 10$ neighborhood in some detail, see Figure 5 through 13. In Figure 5 the CVC predictions from the Loc/Stat/Trad predictor defined in Section 5.1.1 are displayed. The format of the figure correspond to Figure 3. The global predictions in Figure 3.a and the Loc/Stat/Trad predictions in Figure 5.a look fairly similar, which is not surprising since there is a large number of observations and a fairly smooth spatial correlation model is used. The standard deviations for global and Loc/Stat/Trad predictors do differ significantly, though, see Figure 3.b and 5.b respectively. The former does only depend on location configuration while the latter also depend on the locally estimated variance. One concern is, however, that estimated variances based on $k = 10$ observations may be unstable. We observe that there is larger dispersion in Figure 5.b than in Figure 3.b. The resulting normalized crossvalidation errors in Figure 6 for the Loc/Stat/Trad predictor appears as homogeneous across the area. More so than the corresponding errors for the Global predictor in Figure 4. By

comparing the histograms from the Loc/Stat/Trad and Global predictors in Figure 6.b and Figure 4.b respectively, we observe that the former has much lighter tails.

The Loc/Stat/Shr predictor is defined in Section 5.1.2 and we study a $k = 10$ neighborhood version. The empirical Bayes approach used in the predictor include prior models on the expectation and variance inferred from the global data set. These prior models are displayed in Figure 13. The results from the CVC Loc/Stat/Shr predictor are presented in Figure 7. These results are compared to the corresponding results for the Global and Loc/Stat/Trad predictors presented in Figure 3 and 5, respectively. The crossvalidation predictions appear as fairly similar for all predictors, while the standard deviations differs significantly. The Loc/Stat/Shr results seem to lie in between the results for the two other predictors. The former predictor does actually shrink the localized estimates towards the global ones, hence this results are not surprising. The normalized crossvalidation error results are displayed in Figure 8. The results do not deviate much from the corresponding results in Figure 6. Note, however, that the histogram of the errors in Figure 8.b are very symmetrical with very light tails.

The Loc/Non-stat/Trad predictor is defined in Section 5.2.1. This predictor has weighting of the observations that depends on the spatially varying variance estimates. The results from the CVC Loc/Non-stat/Trad predictor with $k = 10$ are displayed in Figure 9 and 10. The crossvalidation predictions do not deviate much from the other predictors. The standard deviations are even more dispersed than the ones obtained from the Loc/Stat/Trad predictor in Figure 5. The normalized crossvalidation errors from the Loc/Non-stat/Trad predictor in Figure 10 have very little dispersion, hence the histogram is very compact almost without tails.

The Loc/Non-stat/Shr predictor is defined in Section 5.2.2 and it is based on the empirical Bayes approach and uses the prior models in Figure 13. The CVC prediction results are presented in Figure 11 and 12, and the crossvalidation predictions are similar to the other predictors. The prediction standard deviations appear as a shrunk version of to the ones for the Loc/Non-stat/Trad predictor in Figure 9. By inspecting the normalized crossvalidation errors, in Figure 12, they appear with little dispersion, and the histogram is compact with fairly light tails. Recall that for the Loc/Non-stat/Shr predic-

tor shrinkage is enforced on the observation weighting itself.

In Table 1, the evaluation criteria defined in Section 6 are displayed for the various predictors. The results for the Global predictor, corresponding to Loc/Stat/Trad $k = 1000$, are presented in the leftmost column. The mean normalized error (MNE) and mean squared normalized error (MSNE) represents statistics of the normalized crossvalidation errors prior to crossvalidation calibration (CVC). We observe that the errors are correctly centered at zero, but somewhat over-dispersed. The evaluation criteria prediction mean square error (PMSE) and variance mean square error (VMSE) are based on the CVC predictors, see Section 5.3. These CVC predictors are globally centered to zero and scaled to unity. The PMSE criterion represents prediction quality while the VMSE represents prediction variance quality, and for both criteria small values are favored. We consider the prediction criteria PMSE as more important than the prediction variance criterion VMSE. The criteria have no absolute scale and are only suitable for comparision between various predictors.

In Table 1, column two, the evaluation criteria for the Loc/Stat/Trad $k = 10$ predictor are displayed. The non-calibrated centering appear as very good, while the variance is too large. The former follows from the localized predictors all being unbiased, while the latter is caused by the localized prediction variances lacking a global reference. The criteria PMSE and VMSE are based on the corresponding CVC predictors which are globally calibrated. The prediction quality, represented by PMSE, appears to be slightly better for the Loc/Stat/Trad predictor than for the Global predictor. The prediction variance quality, represented by VMSE, however, appears as significantly better for the former than for the latter. Hence the Loc/Stat/Trad $k = 10$ predictor dominates the Global predictor in this study.

In Table 1, column three, the evaluation criteria for the Loc/Stat/Shr $k = 10$ predictor are presented. From the MNE and MSNE values we observe good centering and over dispersion in the non-caliberated normalized crossvalidation errors. The PMSE and VMSE values based on the corresponding CVC predictor, are very encouraging. The prediction quality appears as slightly better than for the Loc/Stat/Trad and Global predictors, while the prediction variance quality seems to be significantly better than for the two other predictors. Hence the localized/shrinkage kriging predictor appears to dom-

inate the global and localized kriging predictors in this study.

Table 1, column four and five, contain results from the Loc/Non-stat/Trad $k = 10$ and Loc/Non-stat/Shr $k = 10$ predictors, respectively. Both these predictors are well centered, and are highly over-dispersed prior to global crossvalidation caliberation. The criteria PMSE and VMSE based on the corresponding CVC predictors give some mixed signals. The PMSE representing prediction quality appear as poorer than for the global and localized, stationary predictors. This lack of precision in the predictions may be explained by the large number of model parameters that are implicitly estimated. Recall that the observation weights are based on locally estimated variances. The quality of the prediction variances, represented by PMSE, appear as very favourable compared to the other predictors. It is somewhat surprising that the Loc/Non-stat/Trad predictor performs better than the Loc/Non-stat/Shr one for prediction variance assessment for $k = 10$, however.

To summarize, localized predictors appear favorably to the global predictor. This effect is most likely caused by lack of stationarity in both expectation and variance of the phenomen under study. For a phenomenon that is less smooth and with more sparse observations the localized predictors are expected to be even more favorable. The localized, shrinkage predictors stabilizes the corresponding localized, traditional predictors and provide very encouraging prediction results in the study. The previous results are all based on a localization with a $k = 10$ neighborhood. In Table 2 through 4 corresponding results for $k = 4, 8, 16$ respectively, are presented.

In Table 2, results are displayed from the Global predictor, Loc/Stat/Trad $k = 1000$, and the other predictors with $k = 4$. By Comparing prediction quality PMSE and prediction variance quality VMSE we observe that such a small neighborhood provide very unstable local estimates of expectation and variance. The corresponding predictors have poor performance compared to the Global predictor. Note, however, the improvements in prediction variance quality VMSE by using shrinkage.

Table 3 contain results from the $k = 8$ neighborhoods. The results are very similar to ones for $k = 10$, in Table 1. If we compare each predictor for different k-neighborhoods, we observe that precision quality PMSE is best

for $k = 8$ while precision variance quality VMSE is best for $k = 10$. There appears to be some kind of trade-off between the criteria PMSE and VMSE.

Table 4 presents results for $k = 16$ neighborhoods. The localized, shrinkage predictors perform very favorably for this case, particularly the Loc/Stat/Shr predictor which has high prediction quality PMSE and very favorable prediction variance quality VMSE. This may indicate the neighborhood somewhat larger than $k = 10$ should be used. In fact, one may optimize the size of the neighborhood with respect to a loss criterion combining PMSE and VMSE. One may also perform predictor selection along these lines.

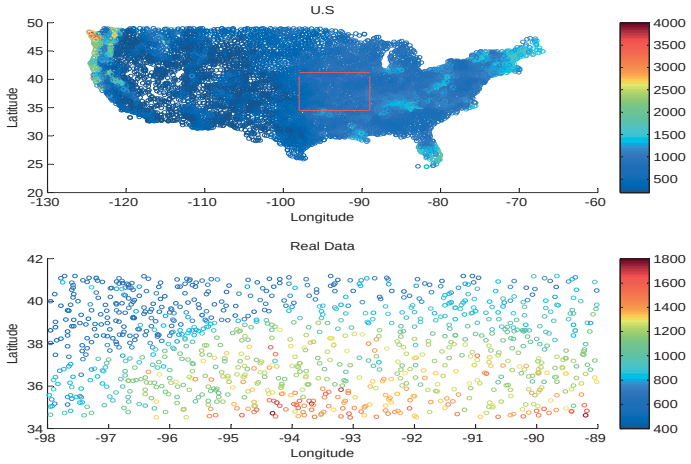

Figure 32: Annual accumulated prepecitation observation in the US and sub-area studied.

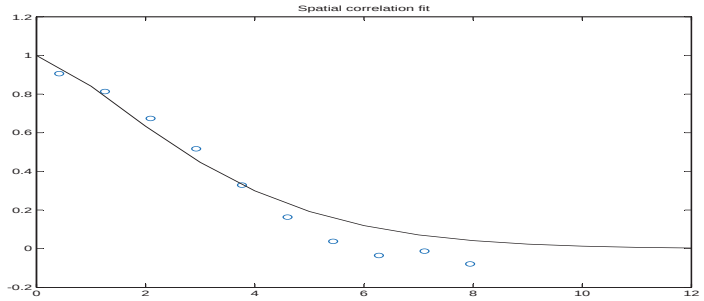

Figure 33: Spatial correlation function with estimated values.

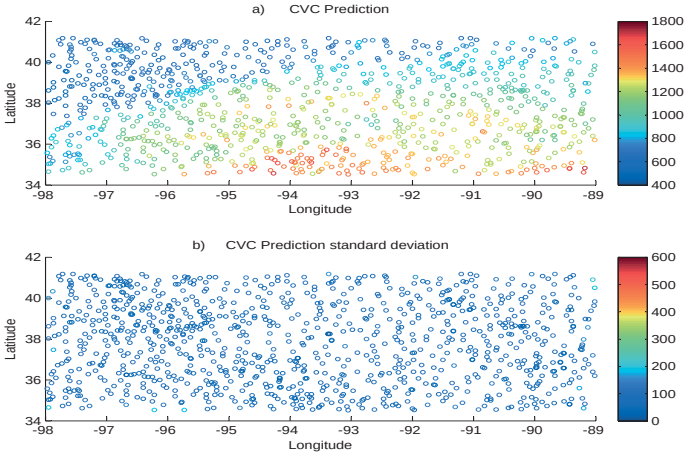

Figure 34: Global CVC predictor-ordinary kriging.

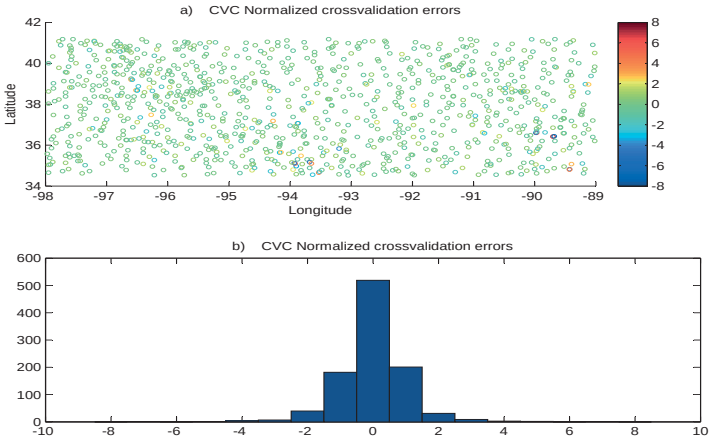

Figure 35: Global crossvalidation errors.

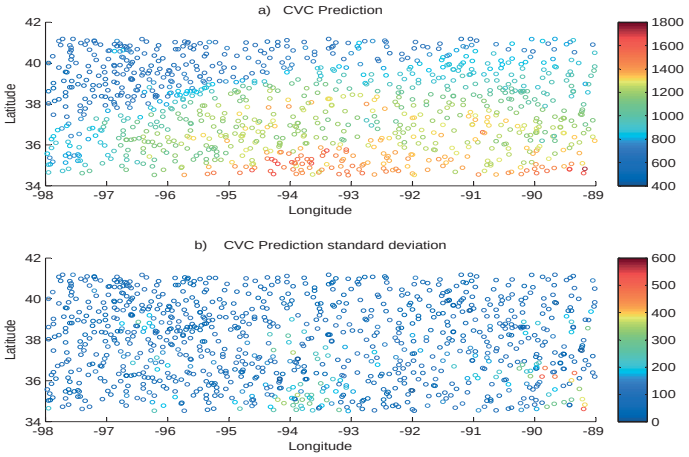

Figure 36: Loc/Stat/Trad/10 CVC predictor.

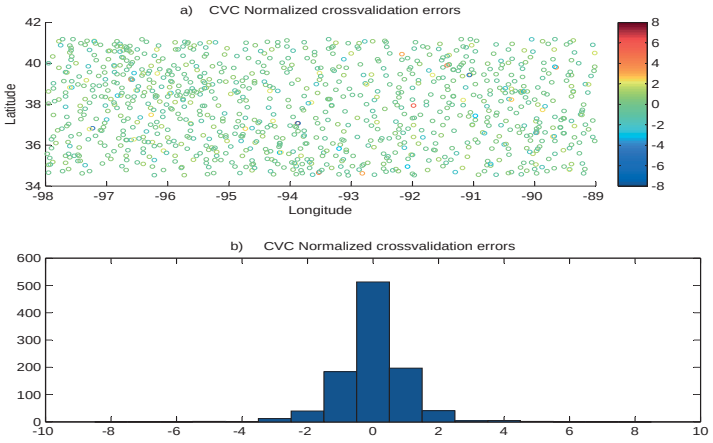

Figure 37: Loc/Stat/Trad/10 crossvalidation errors.

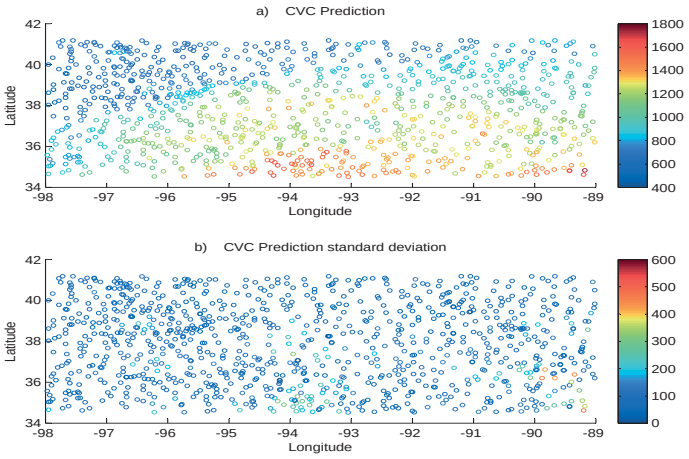

Figure 38: Loc/Stat/Shr/10 CVC predictor.

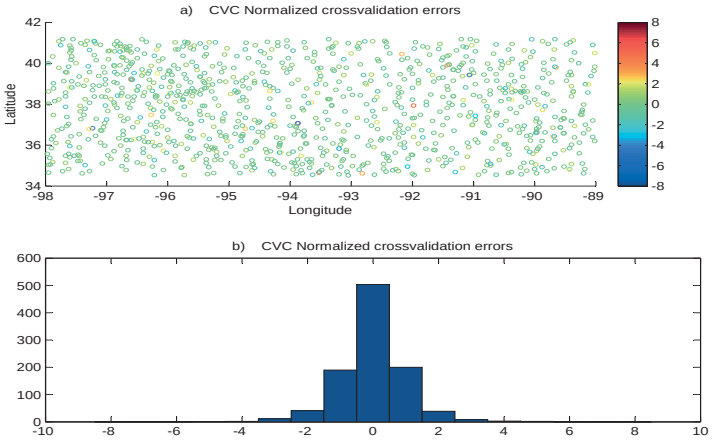

Figure 39: Loc/Stat/Shr/10 crossvalidation errors.

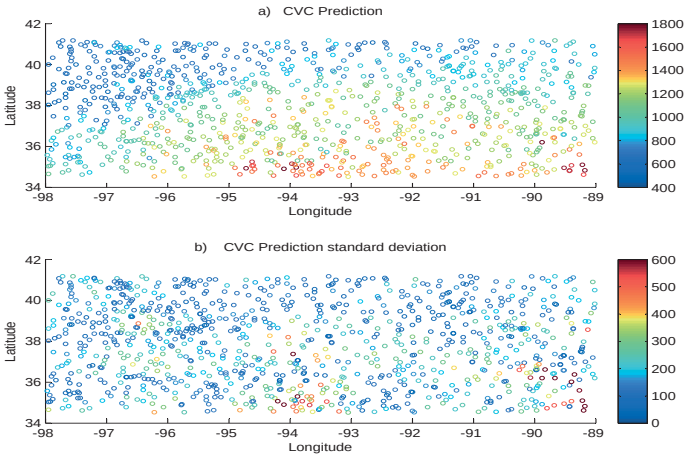

Figure 40: Loc/Non-stat/Trad/10 CVC predictor.

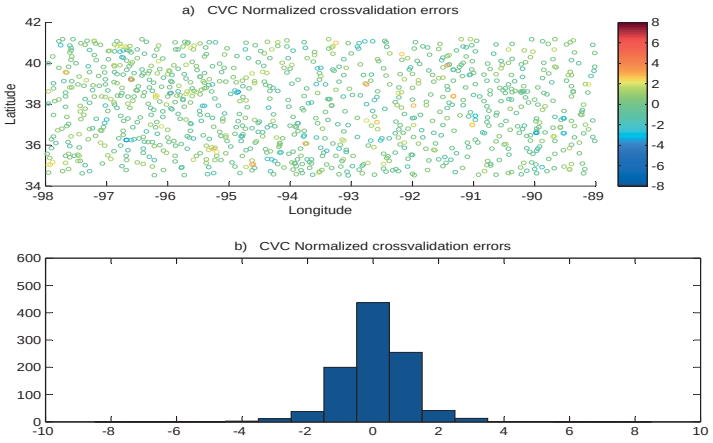

Figure 41: Loc/Non-stat/Trad/10 crossvalidation errors.

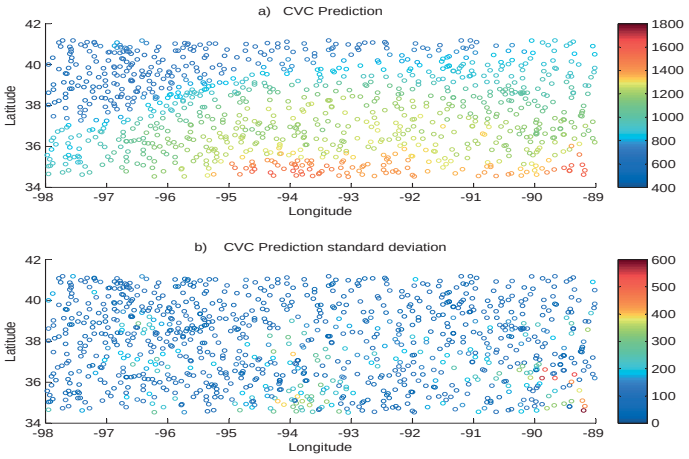

Figure 42: Loc/Non-stat/Shr/10 CVC predictor.

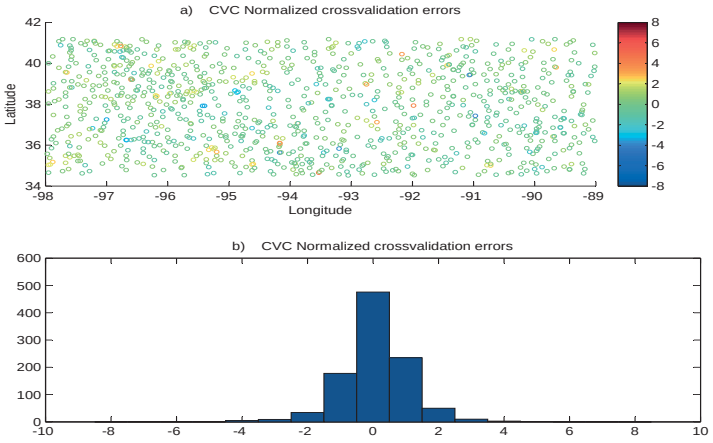

Figure 43: Loc/Non-stat/Shr/10 crossvalidation errors.

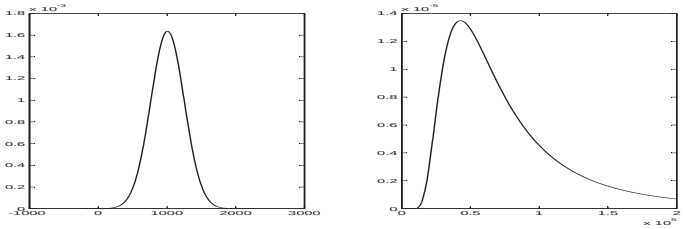

Figure 44: Priors model for expectation and variance.

Model	Localized/Stationary			Localized/Non-stationary	
	Traditional		Shrinkage	Traditional	Shrinkage
Test D.	$k = 1000$	$k = 10$	$k = 10$	$k = 10$	$k = 10$
MNE	1.0148e-17	-5.3792e-18	5.9892e-17	1.5084e-16	-1.1579e-16
MSNE	1.5399	2.9740	3.3928	9.2487	5.1702
PMSE	$6.8758e+03$	$6.8745e+03$	$6.8654e+03$	$2.2181e+04$	$9.3555e+03$
VMSE	9.8749	5.9746	5.2027	4.0787	4.3475

Table 7: Precipitation crossvalidation: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE)and Variance mean squared error (VMSE).

Model	Localized/Stationary			Localized/Non-stationary	
	Traditional		Shrinkage	Traditional	Shrinkage
Test D.	$k = 1000$	$k = 4$	$k = 4$	$k = 4$	$k = 4$
MNE	1.0148e-17	2.3291e-18	-1.8677e-16	1.7080e-17	5.6565e-17
MSNE	1.5399	12.4782	6.8602	14.9119	7.0732
PMSE	$6.8758e+03$	$7.2351e+03$	$7.1938e+03$	$1.0271e+04$	$6.2025e+03$
VMSE	9.8749	76.5237	14.7642	39.6900	10.3703

Table 8: Precipitation crossvalidation: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE)and Variance mean squared error (VMSE).

Model	Localized/Stationary			Localized/Non-stationary	
	Traditional		Shrinkage	Traditional	Shrinkage
Test D.	$k = 1000$	$k = 8$	$k = 8$	$k = 8$	$k = 8$
MNE	1.0148e-17	-1.5639e-17	3.9207e-17	-1.2422e-16	-1.7258e-16
MSNE	1.5399	3.6322	3.9626	8.7274	5.3399
PMSE	$6.8758e+03$	$6.8421e+03$	$6.8275e+03$	$1.8969e+04$	$8.3440e+03$
VMSE	9.8749	8.0073	5.9935	4.2336	4.5284

Table 9: Precipitation crossvalidation: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE)and Variance mean squared error (VMSE).

Model	Localized/Stationary			Localized/Non-stationary	
	Traditional		Shrinkage	Traditional	Shrinkage
Test D.	$k = 1000$	$k = 16$	$k = 16$	$k = 16$	$k = 16$
MNE	1.0148e-17	1.7191e-17	1.6304e-17	-5.1907e-17	1.8367e-16
MSNE	1.5399	2.2864	2.5668	9.4840	4.9365
PMSE	$6.8758e+03$	$6.8581e+03$	$6.8549e+03$	$2.5527e+04$	$1.1038e+04$
VMSE	9.8749	5.8588	4.8849	4.3886	4.0764

Table 10: Precipitation crossvalidation: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE)and Variance mean squared error (VMSE).

3.7.2 Gamma-log Data

In Figure 14 a Gamma ray data set from a vertical subsurface well is displayed. The data set locations are numbered as $[1, 2, ..., 600]$. We split the data in an observations set $[1, 30, ..., 570, 600]$ with $n_o = 21$, and a control set containing the 579 remaining data. The two sets are presented in Figure 14. Contrary to the US precipitation study, we have control data here, while we only operate in one dimension with a limited number of observations.

We use the observation set of size $n_o = 21$ in a cross-validation study and we also do prediction into the locations of the control data set. The spatial correlation function

$$\rho_r(\tau) = exp\left\{-\tfrac{1}{30}\tau^{1.5}\right\};\ \tau \geq 0$$

is used throughout the study. The predictors in Section 5 are evaluated by the evaluation criteria in Section 6.

In Figure 15 results from both the Global predictor and the localized predictor are displayed. The predictions look fairly similar, while the prediction variances differ. The localized predictors have prediction variances that vary with location. In the shrinkage predictors empirical prior models for expectation and variance are used, and these prior models are displayed in Figure 16. The former being Gaussian and the latter Inverse-Gamma. Note how the prediction variances in the shrinkage predictor shrink the localized ones towards the global ones.

Table 5 contains values of the evaluation criteria for the four predictors for different values of $k = 4, 8, 12$ by using crossvalidation within the observation data set of size $n_o = 21$. Note that all normalized crossvalidation errors are reasonably centered, MNE close to zero, but under-dispersed, MSNE greater than unity. The CVC predictions that normalize with respect to this under-dispersion provide the base for the prediction criterion PMSE and the prediction variance criterion VMSE. Both are favored to be small. Note that the shrinkage predictors make significantly better predictions than their traditional counterparts, ie smaller PMSE. For the prediction variance criterion VMSE we observe the same picture. The shrinkage predictors appears with smallest VMSE. The best shrinkage predictors have very small neighborhoods ± 2 which indicates fast changing model characteristics in the Gamma ray data.

Table 6 contain values of the evaluation criteria based on the differences between the control set and the different CVC predictors. The prediction criterion PMSE is almost equal for all predictors, while the prediction variance criterion VMSE appears as favourable for the traditional predictors. The latter constitutes a surprising result. Note, however, that these results are based on a very limited number of observations.

To summarize, if crossvalidation within the $n_o = 21$ observations is used in the calculations of the criteria, then we obtain results similar to the ones on the US precipitation data. The shrinkage predictors are clearly favorable

to their traditional counterparts. If, however, the criteria is based on a control set of data, shrinkage predictors do not appear as favourable. We do not understand why, but it may just be by chance in a limited dataset.

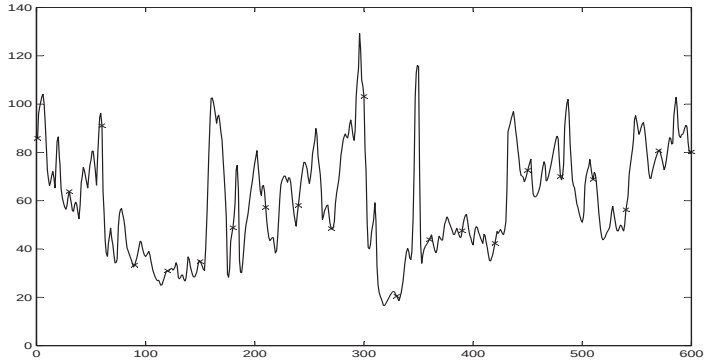

Figure 45: Gamma ray observations-with observations (*) and control values.

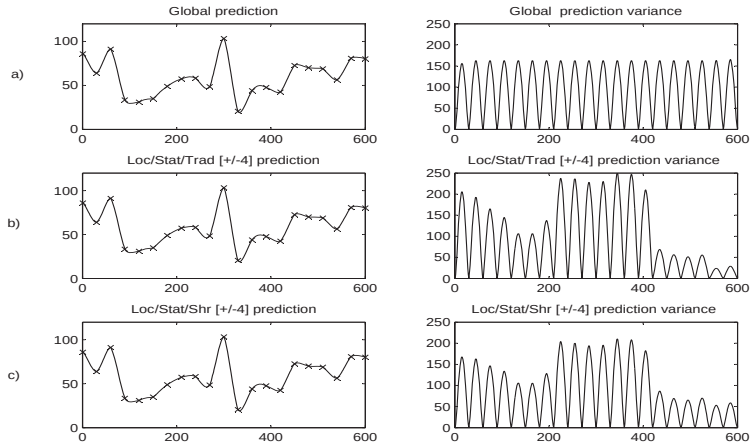

Figure 46: Gamma ray predictions and prediction variances.

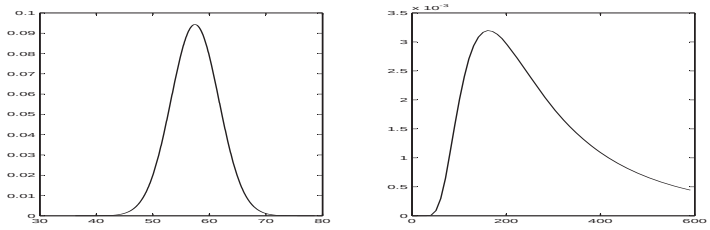

Figure 47: Prior model for expectation and variance.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	±6	±4	±2	±6	±4	±2	±6	±4	±2	±6	±4	±2
MNE	0.0	0.2	0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.1	0.1
MSNE	3.2	4.5	3.0	2.3	2.5	1.9	4.1	6.0	4.2	2.5	2.7	1.8
PMSE	639.7	705.2	703.3	589.6	594.1	561.1	921.5	1318.4	1247.7	662.6	704.3	595.7
VMSE	5.3	6.7	4.5	4.3	5.0	3.5	3.8	4.3	2.8	3.5	3.8	2.5

Table 11: Gamma ray crossvalidation: Mean normalized error (MNE), Mean Squared normalized error (MSNE) Prediction Mean squared error (PMSE) and Variance Mean Squared error (VMSE).

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	±6	±4	±2	±6	±4	±2	±6	±4	±2	±6	±4	±2
PMSE	336.71	331.40	330.40	340.98	335.22	333.86	346.16	340.86	335.60	345.93	342.59	358.97
VMSE	9.6033	4.9232	72.3516	19.4386	15.2300	50.1785	5.7387	3.1133	42.3435	16.4979	13.6487	57.3204

Table 12: Gamma ray control values: Prediction Mean squared error (PMSE) and Variance Mean Squared error (VMSE) .

3.8 Empirical Study

The prediction models defined in the previous section can be evaluated based on one realization only. We define a test design and several evaluation criteria and make the evaluation for a variety of Gaussian random field models for which the exact solutions are analytically obtainable.

3.8.1 Test Design and Criteria

We define a variety of 1**D** Gaussian RF: $\{r(x); x \in \mathbf{D} \in \Re^1\}$, discretised as $\mathbf{r} = \{r(x); x \in \mathcal{L}_\mathbf{D}\}$ where $\mathcal{L}_\mathbf{D} = \{1, 2, ..., 199, 200\}$ and hence $n = 200$. The observations are obtained as $\mathbf{r}_o = \{r(x); x \in \mathcal{L}_\mathbf{o}\}$ where $\mathcal{L}_\mathbf{o} = \{1, 10, ..., 190, 200\}$ and hence $n_o = 21$. We use the CVC predictors defined in Section 5 based on $\mathbf{r}_o$ to obtain the predictions $\tilde{\mathbf{r}} = [\tilde{r}_1, ..., \tilde{r}_{200}]^T$ and the associated prediction variances $\tilde{\boldsymbol{\sigma}}^2 = [\tilde{\sigma}_1^2, ..., \tilde{\sigma}_{200}^2]^T$.

Note that all the Gaussian RF are analytically tractable when the model parameters are known, hence the optimal predictions $\mathbf{r}^* = [r_1^*, ..., r_{200}^*]^T$ and associated prediction variances $\boldsymbol{\sigma}^{*2} = [\sigma_1^{*2}, ..., \sigma_{200}^{*2}]^T$ are avaliable.

The evaluation for a single realization is based on both comparison with the correct predictions and predictions variances and cross-validation in the 15 centrally located observations.

The evaluation criteria in the comparison with correct results are:

$$PMSC = \frac{1}{126} \sum_{\substack{i=30 \\ i \notin \mathcal{L}_\mathbf{o}}}^{170} [\tilde{r}_i - r_i^*]^2$$

$$VMSC = \frac{1}{126} \sum_{\substack{i=30 \\ i \notin \mathcal{L}_\mathbf{o}}}^{170} \left[\frac{\sigma_i^{*2}}{\tilde{\sigma}_i^2} - 1 \right]^2$$

We randomize over the model by averaging over 1000 realizations to obtain APMSC and AVMSC. Note that this is the ultimate criteria for goodness for the predictor and prediction variances.

The evaluation criteria in the cross-validation are PMSE and VMSE as in previous sections, but in addition we randomize over the model by averaging over 1000 realizations to obtain the criteria APMSE and AVMSE. These are the criteria we need to use when only one set of observations are avaliable.

This study is targeted at identifying a good spatial predictor and suitable number of neighborhood observations involved in the predictor. Thus, $k = \pm 2, \pm 4, \pm 6$ observations located closest to prediction location x_+ is considered.

3.8.2 Test Cases

We generate four different test cases with varying expectation and variance fields but with one identical correlation field defined as:

$$Corr[r(x^{'}), r(x^{''})] = \rho_r(x^{'}, x^{''}) = exp\left\{-\tfrac{1}{5}\left|x^{'} - x^{''}\right|^{1.5}\right\}$$

Case I-Test

This case defines a regular stationary Gaussian random field with constant expectation and variance. We use $\boldsymbol{\mu}_r(x) = 10$ and $\boldsymbol{\sigma}_r^2(x) = 20$, see Figure 17.a. In Figure 17.b, the optimal prediction and prediction variance are analytically obtained from the correct model for one realization of the field. The prediction and prediction variance in Figure 17.c are obtained from a stationary model with globally estimated model parameters. These results correspond to ordinary kriging. Figure 17.d displays Loc/Stat/Trad/ ± 4 predictions made according to CVC predictors defined in Section 5, while Figure 17.e displays the results from the corresponding Loc/Stat/Shr/± 4 CVC predictor.

Figure 18 displays the prior models for expectation and variance for one realization in the Loc/Shr predictors defined in Section 5. The parametric prior models for expectation and variance are infered in an empirical Bayes framework as defined in Section 5.

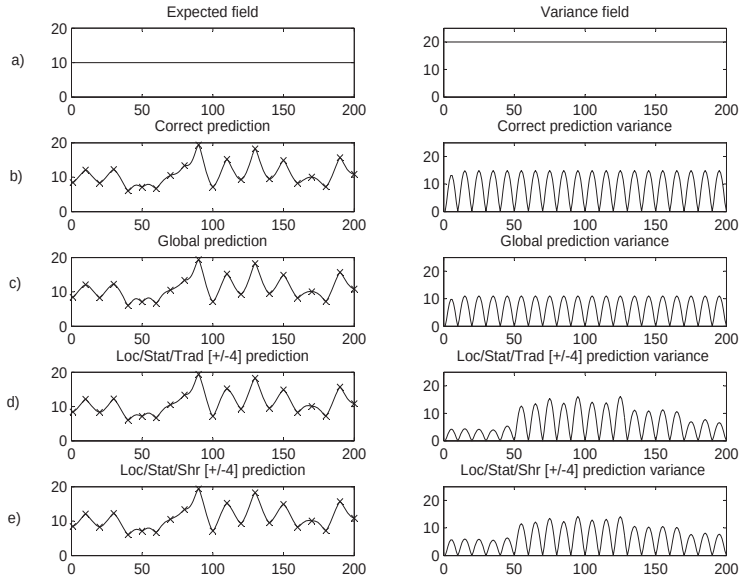

Figure 48: Case I - Predictions and prediction variances for one realization.

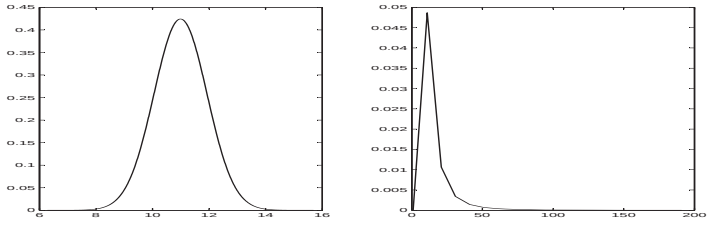

Figure 49: Case I - Prior model for expectation and variance for one realization.

The values of evaluation criteria are listed in Table 7. In Figure 19, histograms of the deviations between one realization and the corresponding predictors, for all the evaluated predictors are displayed.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	±6	±4	±2	±6	±4	±2	±6	±4	±2	±6	±4	±2
APMSC	0.2666	0.3484	1.3569	0.2027	0.2181	0.2696	0.3809	0.6220	2.2520	0.2318	0.2860	0.5083
AVMSC	0.4139	0.6684	43.3676	0.4149	0.5338	1.1251	0.4004	0.6353	37.1616	0.4099	0.5245	1.1335
APMSE	21.6177	22.5767	25.0502	18.5324	18.1406	16.3947	22.1637	23.4164	29.8467	18.6158	18.2514	16.5170
AVMSE	1.8692	1.9430	2.6963	1.6603	1.5907	1.3328	1.8588	1.9239	2.3750	1.6532	1.5778	1.2846

Table 13: Case I Deviation from correct predictions and crossvalidation: Average prediction mean squared correct (APMSC), Average variance mean squared correct (AVMSC), Average prediction mean squared error (APMSE) and Average variance mean squared error (AVMSE).

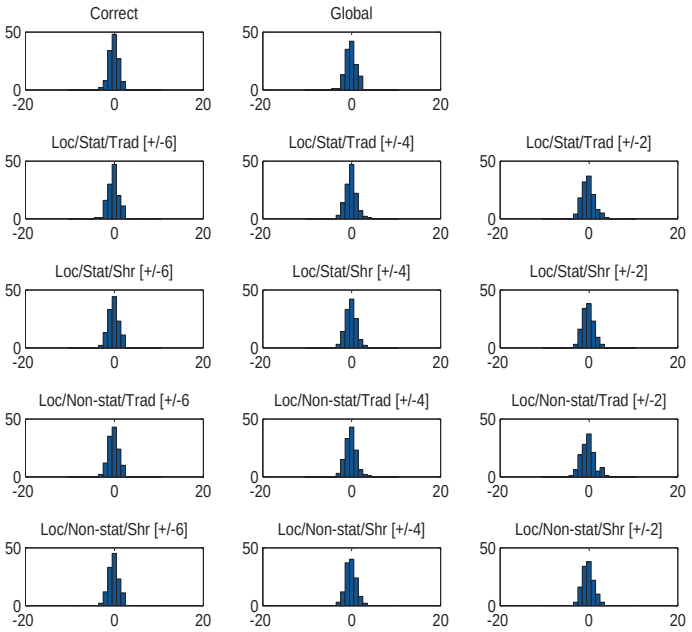

Figure 50: Case I - Histogram for normalized error for one realization.

Case I-Discussion

In Figure 17, the predictions from all CVC predictors appear fairly similar. This is not surprising since the model is stationary in both expectation and variance and the correlation function is known. The prediction variances, however, vary considerably between the predictors. The Global predictor appears with a somewhat under-estimated variance, but otherwise identical to the correct prediction variance. The localized predictors appear with locally varying variances. Note that the shrinkage predictor has variances that are in between the localized traditional and the global traditional one, and hence justifies its shrinkage label. The shrinkage is caused by the prior models on expectation and variance, see Figure 18, which are inferred in an empirical Bayesian setting.

Table 7 contain the values for the evaluation criteria APMSC/AVMSC and APMSE/AVMSE, defined at the begining of Section 8. The two former are

based on comparisions with the correct predictions which are analytically obtainable while the two later relates to crossvalidation. Recall that we favor all four criteria to be as small as possible. For each pair of criteria, the P-criterion is most important since it reflects prediction quality while the V-criterion reflects prediction variance quality. The results for the ultimate criteria APMSC/AVMSC are not surprising, all predictors improve with increasing neighborhood and the shrinkage predictors are almost consistently favorable to the corresponding traditional ones. For a model with stationarity in both expectation and variance, global predictors are optimal of course which favors large neighborhoods. Moreover, shrinkage does reduce local estimation variability which obviously improve the localized predictors. This effect is clearly observable for small neighborhoods, where shrinkage provides dramatic improvements in localized predictors.

In Table 7, also the corresponding crossvalidation criteria APMSE/AVMSE are listed, which we must rely on with only one set of observations avaliable. Note that also these criteria consistently favor the shrinkage predictors relative to their traditional counterparts. The best shrinkage predictor with respect to the APMSE/AVMSE criterions appear with the smallest neighborhood, however. This is unfortunate, and it may be caused by overfitting to the observations.

The normalized error histograms for one realization for all CVC predictors are displayed in Figure 19. The histograms are fairly similar, but for small neighborhoods, ± 2, we can observe that the errors are somewhat regularized by the shrinkage effect.

Case II-Test

This case defines a non-stationary Gaussian random field with varying expectation and variance. We use $\boldsymbol{\mu}_r(x) = 10\sin(\pi\frac{x}{200})$ and $\boldsymbol{\sigma}_r^2(x) = 20\sin(\pi\frac{x}{200})$, see Figure 20.a. In Figure 20.b, the optimal prediction and prediction variance are analytically obtained from the correct model. The prediction and prediction variance in Figure 20.c are obtained from a stationary model with globally estimated model parameters. These results correspond to ordinary kriging. Figure 20.d displays Loc/Stat/Trad/± 4 predictions made according to CVC predictors defined in Section 5, while Figure 20.e displays the results from the corresponding Loc/Stat/Shr/± 4 CVC predictor.

Figure 21 displays the prior models for Loc/Shr predictors, and these models capture the gross variability over the domain. Table 8 and Figure 22 correspond to Table 7 and Figure 19, for Case I.

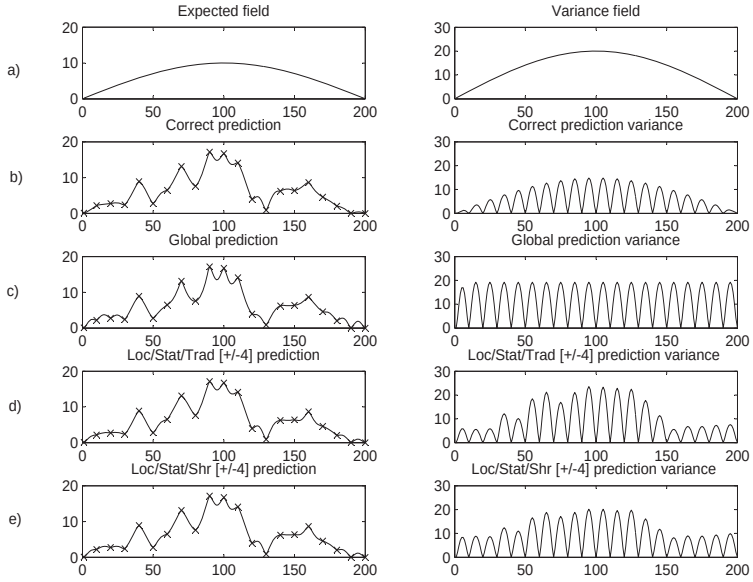

Figure 51: Case II - Predictions and prediction variances for one realization.

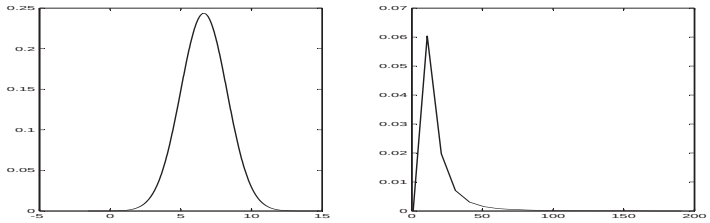

Figure 52: Case II - Prior model for expectation and variance for one realization.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	±6	±4	±2	±6	±4	±2	±6	±4	±2	±6	±4	±2
APMSC	0.3665	0.3357	1.0100	0.3968	0.2838	0.2563	0.4815	0.6669	2.1658	0.4259	0.4305	0.6264
AVMSC	0.5287	1.0939	159.3760	0.3927	0.5944	1.2491	0.5108	1.0369	138.6639	0.3860	0.5723	1.2327
APMSE	18.2536	17.8311	19.8851	16.9502	15.6486	13.8070	18.6699	18.4643	23.7843	17.0732	15.8705	14.1365
AVMSE	1.8936	2.0277	2.9888	1.7271	1.7306	1.4011	1.8908	2.0086	2.4491	1.7277	1.7187	1.3483

Table 14: Case II Deviation from correct predictions and crossvalidation: Average prediction mean squared correct (APMSC), Average variance mean squared correct (AVMSC), Average prediction mean squared error (APMSE) and Average variance mean squared error (AVMSE).

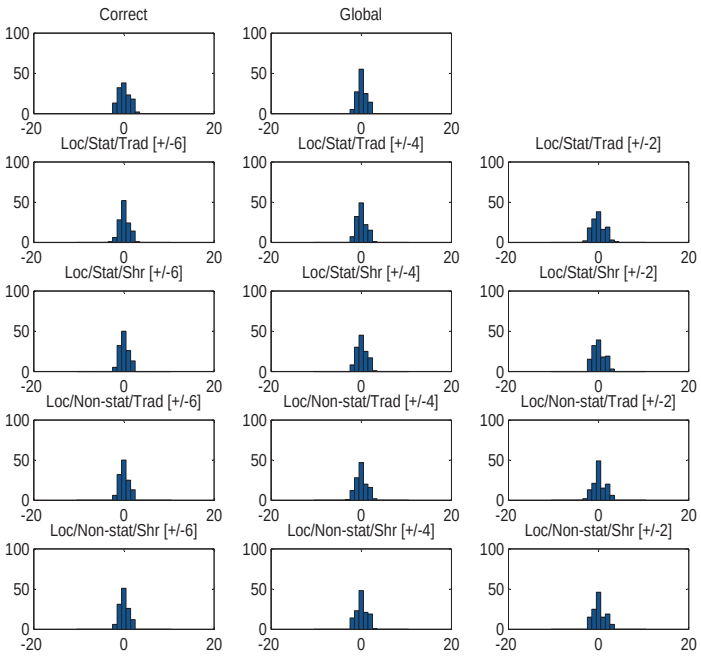

Figure 53: Case II - Histogram for normalized error for one realization.

Case II-Discussion

In Figure 20, the predictions vary considerably. The global predictor is based on a stationary model and predictions between observations are clearly biased towards the global average of the observations. The localized predictors are fairly similar and do not deviate much from the correct prediction. The prediction variances for the global predictor reflect the stationarity assumptions made and only capture the localization configuration of the observations. The prediction variances from the localized predictors capture much more of the non-stationarity of the model. Note also that the shrinkage predictor provides prediction variances in between the local traditional and global traditional predictors. The shrinkage is caused by the prior models displayed in Figure 21.

Table 8 contain the values of the evaluation criteria. Consider first the APMSC for Loc/Stat/Trad CVC predictors with varying neighborhoods. These are prediction quality relative to the correct predictions. Observe the pattern, best for ± 4 and poorer for ± 6 and ± 2. We observe a bias/variance crossing point, since too large neighborhood provides biased estimates of the non-stationary expectation and variance, while too small neighborhood provides large estimation variance due to few observations. Consider now the APMSC criterion for corresponding shrinkage CVC predictor, Loc/Stat/Shr. This predictor regularizes the estimates of the expectation and variances and appear as clearly favorable for small neighborhoods. The improvements are so large that the bias/variance crossing point is moved to ± 2 for the shrinkage predictor. The prediction variance quality is reflected by AVMSC relative to the correct predictions. The stability of variance estimates are of course very poor, and use of shrinkage predictors has a dramatic positive effect for small neighborhood predictors. A joint assessment based on both prediction and prediction variance quality, would probably make us choose the Loc/Stat/Shr/± 4 CVC predictor among the stationary predictors.

Note, however, that for models which are non-stationary in both expectation and variance, localized, non-stationary predictors have the potential of being better than stationary ones. By inspecting the APMSC/AVMSC criteria for Loc/Non-stat/Trad and Loc/Non-stat/Shr predictors for varying neighborhoods, we observe that they are less favorable than the stationary ones. We belive this is caused by the need to estimate a large number of

parameters.

In practice, with one set of observations we need to select a predictor based on the crossvalidation criteria APMSE/AVMSE. Based on these criteria the shrinkage predictors appear as consistently favorable to the corresponding traditional ones. As they actually are. Moreover, we would select a stationary, shrinkage predictor prior to a non-stationary one. However, we would select the Loc/Stat/Shr ± 2 predictor since the prediction variance criterion appears as under-estimated for the crossvalidation criterion. Hence we select a too small neighborhood in the predictor.

In Figure 22, normalized error histograms for one realization are displayed. Observe the favorable shape of the global histogram, but this is caused by normalization based on severely over-estimated prediction variances. Note also that the histograms for the correct predictor and Loc/Stat/Shr with ± 4 and ± 2 are very similar. Lastly, a small regularization effect on the histograms can be seen from shrinkage.

Case III-Test

This case defines a non- stationary Gaussian random field with varying expectation and constant variance. We use $\boldsymbol{\mu}_r(x) = 10\sin(\pi\frac{x}{200})$ and $\boldsymbol{\sigma}_r^2(x) = 20$, see Figure 23.a. In Figure 23.b, the optimal prediction and prediction variance are analytically obtained from the correct model. The prediction and prediction variance in Figure 23.c are obtained from a stationary model with globally estimated model parameters. These results correspond to ordinary kriging. Figure 23.d displays Loc/Stat/Trad/± 4 predictions made according to CVC predictors defined in Section 5, while Figure 23.e displays the results from the corresponding Loc/Stat/Shr ± 4 CVC predictor.

Figure 24 displays the prior models for Loc/Shr predictors, which represents the gross variability over the domain. Table 9 and Figure 25 correspond to Table 7 and 8, and Figure 19 and 22 for Case I and II, respectively.

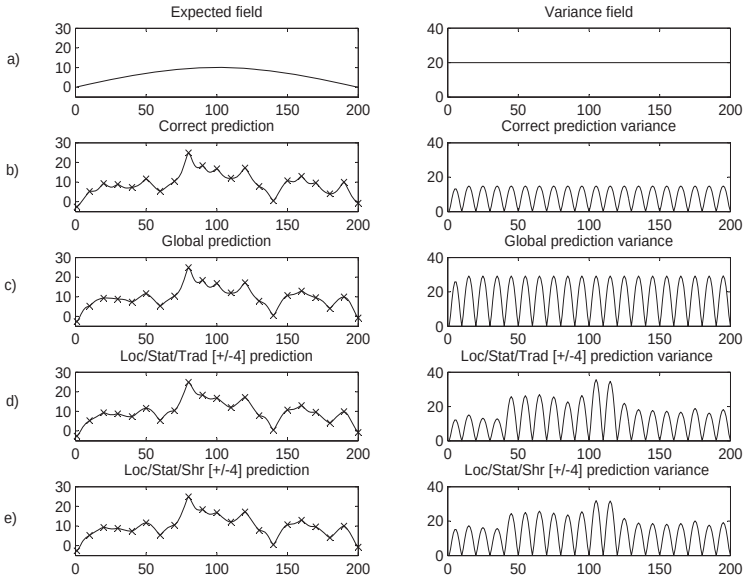

Figure 54: Case III - predictions and prediction variances for one realization.

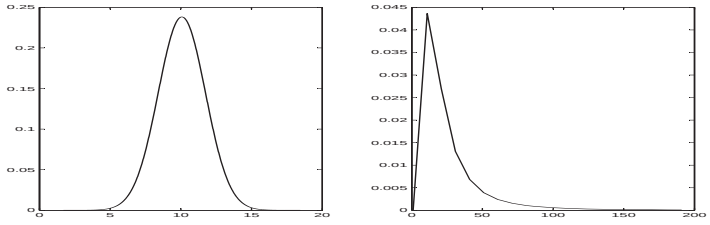

Figure 55: Case III - Prior model for expectation and variance for one realization.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	±6	±4	±2	±6	±4	±2	±6	±4	±2	±6	±4	±2
APMSC	0.4339	0.4027	1.5498	0.4628	0.3354	0.3152	0.5800	0.7993	3.0182	0.5014	0.4956	0.7195
AVMSC	0.5783	1.0568	687.1633	0.3919	0.5517	1.0074	0.5566	0.9922	590.5700	0.3847	0.5302	0.9891
APMSE	22.5294	22.3758	24.8330	20.6543	19.2866	17.0429	23.0463	23.1767	29.0106	20.7936	19.5368	17.3665
AVMSE	1.8404	1.9649	3.3623	1.6693	1.6335	1.3537	1.8366	1.9477	2.5569	1.6691	1.6236	1.3147

Table 15: Case III Deviation from correct predictions and crossvalidation: Average prediction mean squared correct (APMSC), Average variance mean squared correct (AVMSC), Average prediction mean squared error (APMSE) and Average variance mean squared error (AVMSE).

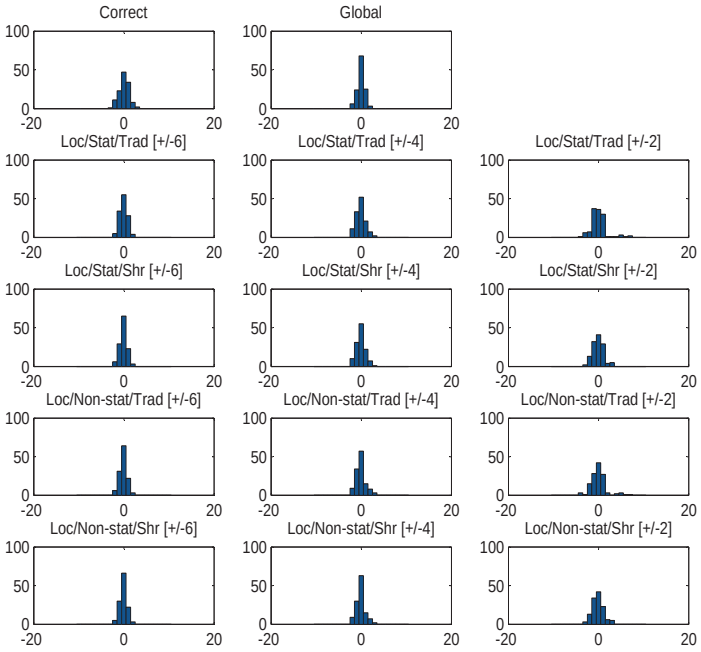

Figure 56: Case III - Histogram for normalized error for one realization.

Case III-Discussion

In Figure 23, the predictions vary somewhat, since the global predictions are biased towards the global average of the observations. The prediction variances are different because they merge variance and expectation curvature. Note the upwards bias in the variance of the global stationary model. The localized predictors appear with localized variance estimates and the shrinkage predictor is somewhat closer to the global prediction variance. The prior models in Figure 24 causes this shrinkage.

Table 9, contain the results for the evaluation criteria APMSC/AVMSC and they exposes patterns very similar to the results from the model with non-stationarity in both expectation and variance, see Figure 8. The stationary, traditional predictor makes bias/variance trade-offs in the predictor, while the stationary, shrinkage improve on the predictor by regularization of the estimation and hence reducing variance. This shrinkage make small neighborhood predictors more favorable. The prediction variances are more unstable, and we probably end up recommending the Loc/Stat/Shr/$\pm$4 CVC predictor. The non-stationary predictors are less favorable due to their dependence on a large number of parameters.

In practice, with only one set of observations avaliable, we must select a predictor based on the crossvalidation criteria APMSE/AVMSE. We would correctly favor shrinkage predictors for traditional ones and stationary predictors for non-stationary ones. However, we would select a too small neighborhood, as Loc/Stat/Shr $\pm$2 would be the favorable predictor.

In Figure 25, the normalized error histograms for one realization for all the CVC predictors are presented. Note in particular the localized predictors with a $\pm$2 neighborhood. The histograms for the shrinkage predictors are clearly regularized compared to the histograms for the traditional ones.

Case IV-Test

This case defines a Gaussian random field with constant expectation and varying variance. We use $\boldsymbol{\mu}_r(x) = 10$ and $\boldsymbol{\sigma}_r^2(x) = 20\sin(\pi\frac{x}{200})$, see Figure 26.a. In Figure 26.b, the optimal prediction and prediction variance are analytically obtained from the correct model. The prediction and prediction

variance in Figure 26.c are obtained from a stationary model with globally estimated model parameters. These results correspond to ordinary kriging. Figure 26.d displays Loc/Stat/Trad/± 4 predictions are made according to CVC predictors defined in Section 5, while Figure 26.e displays the results from the corresponding Loc/Stat/Shr/± 4 CVC predictor.

Figure 27 displays the prior models for Loc/Shr predictors, and these models capture the gross variability over the domain. Table 10 and Figure 28 correspond to Table 7, 8 and 9, and Figure 19, 22 and 25, for Case I, II and III, respectively.

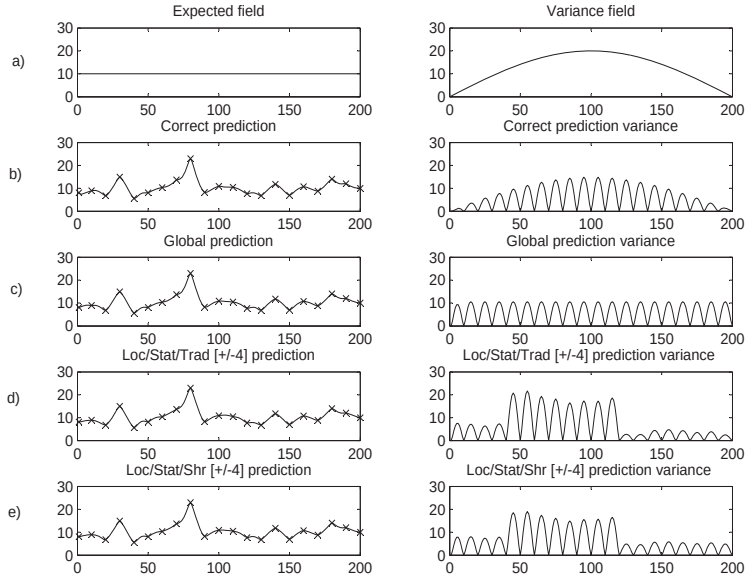

Figure 57: Case IV - Predictions and prediction variances for one realization.

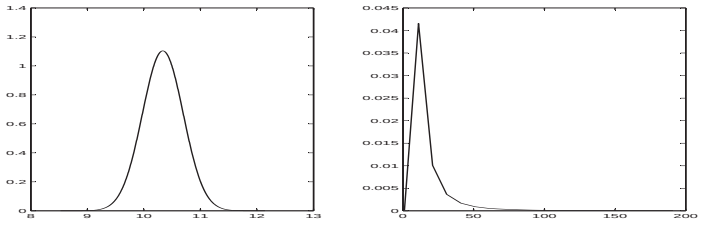

Figure 58: Case IV - Prior model for expectation and variance for one realization.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	±6	±4	±2	±6	±4	±2	±6	±4	±2	±6	±4	±2
APMSC	0.1891	0.2810	1.0623	0.1357	0.1510	0.2023	0.2576	0.4462	1.6301	0.1510	0.1886	0.4097
AVMSC	0.3821	0.7524	33.0924	0.4556	0.6696	1.9722	0.3689	0.7104	26.8692	0.4506	0.6606	1.9826
APMSE	16.9439	17.5722	19.1127	14.5025	14.1418	12.7449	17.3596	18.2110	22.9114	14.5563	14.2143	12.9016
AVMSE	1.8534	1.8717	2.7330	1.6950	1.5993	1.3174	1.8438	1.8514	2.3343	1.6881	1.5868	1.2709

Table 16: Case IV Deviation from correct predictions and crossvalidation: Average prediction mean squared correct (APMSC), Average variance mean squared correct (AVMSC), Average prediction mean squared error (APMSE) and Average variance mean squared error (AVMSE).

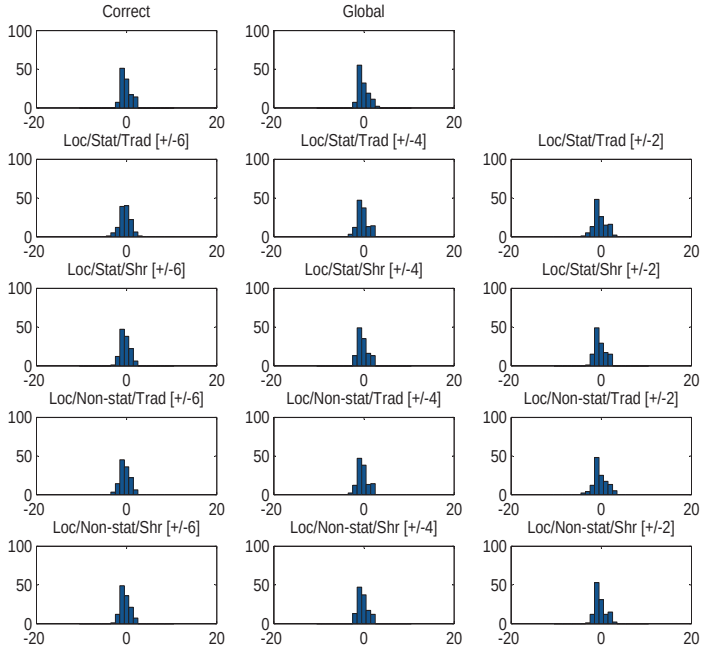

Figure 59: Case IV - Histogram for normalized error for one realization.

Case IV-Discussion

In Figure 26, all predictions appear as very similar. The prediction variances are of course very different since the global predictor estimate one global variance. The shrinkage predictor regularizes the prediction variances of the traditional predictor. The prior models are presented in Figure 27.

Table 10 contain the values of the evaluation criteria APMSC/AVMSC. Large neighborhoods are clearly favorable for all CVC predictors both for prediction and prediction variance criteria. Moreover, shrinkage predictors are clearly favorable to their traditional counterparts, and stationary are favorable to non-stationary. If we select predictor based on the crossvalidation criteria APMSE/AVMSE, we would select Loc/Stat/Shr/$\pm$2, hence correct predictor type but with too small neighborhood. All histograms in Figure 28 are

very similar.

Summary

For a model with stationarity in both expectation and variance the global stationary predictor is of course favorable. Only the global expectation and variance must be estimated. If, however, localized predictors are required, for example to reduce computational demands, then localized shrinkage CVC predictors are clearly favorable.

For models with non-stationary expectation and variances across the domain of study, localized predictors will often appear as clearly favorable to global ones. Moreover, localized, shrinkage CVC predictors seem to be favorable to localized, traditional ones since it regularizes the expectation and variance estimates and hence can operate with larger localization hence smaller neighborhoods. The critical factor appears to be non-stationarity in expectation. If the expectation vary across the domain of study, global predictors can be severely biased towards the average of the observations and the variance estimates are upwards biased by the interplay of expectation and variance. Localized, shrinkage CVC predictors appear as robust towards non-stationary expectations. Non-stationary variances across the domain of study appears to have less influence on the predictors.

In practice, with only one set of observations avaliable, we must rely on crossvalidation criteria in the selection of optimal CVC predictors. We will correctly select localized, stationary, shrinkage predictors, but we will tend to select too small neighborhoods.

3.9 Conclusion

Spatial prediction is usually made under Gaussian assumptions by kriging. In order to robustify the predictor towards lack of spatial stationarity, localized kriging, which includes only observations in a neighborhood, is frequently used. Bias/variance trade-off must be made in order to specify the local neighborhood. We introduce a shrinkage version of kriging in a Bayesian setting, where estimates in a local neighborhood are regularized by a global prior model. This prior model is assessed in an empirical Bayes tradition from the complete, global set of observations.

Two versions of the shrinkage kriging predictor are defined. One stationary version perform regularization only of the model parameter estimates, while the other non-stationary version make regularized estimates of both parameter estimates and kriging weights. Further we define crossvalidation calibrated (CVC) predictors which empirically calibrates the predictor for centering and scale.

The localized/shrinkage CVC predictors are compared to localized/traditional kriging CVC predictors in an empirical simulation study. The experimental design include 1**D** Gaussian random fields with varying expectation and variance trends. Deviation measures between the predictors specified above and the correct predictions and prediction variances which are analytically assessible under the model specification are compared. The major conclusions are:

- localized/shrinkage kriging appears to be almost uniformily superior to the corresponding localized/traditional kriging - for relatively small neighborhoods.
- localized/shrinkage kriging with a regulizer only on model parameter estimates appears superior to the version that also regularizes the kriging weight. In the latter a large number of parameters need to be inferred and uncertainty related to this estimation detoriate the predictor.
- cases with both non-stationary expectation and variance make the localized/shrinkage predictor more favorable to the localized/traditional kriging predictors. Large curvature in the expected trend seems to be the important factor.

- optimal size of the neighborhood is not specifically studied since only one correlation function and one observation design is evaluated. The optimal neighborhood shall normally increase with more smoothness in expectation and variance trends and longer range in the correlation function.

A set of criteria for evaluating the localized CVC predictors based on crossvalidation are developed. This set of criteria can be computed from the set of observations only. The characteristics of this criteria set are explored in the empirical study, and the major conclusions are:

- the crossvalidation criteria can be used to identify the favorable predictor type most frequently being localized, stationary, shrinkage predictors.
- the optimal neighborhood of the favorable predictor seems to be underestimated, hence too small neighborhoods are selected.

We evaluated the localized predictors on two different real datasets: Observations of annually cumulated precipitation in locations in an sub-area of the US, and Gamma ray recordings along a vertical well through the subsurface.

The findings are:

- in the precipitation study, 1001 observations are avaliable, and crossvalidation based evaluation criteria are used. Localized CVC predictors are found to be clearly favorable to the global ordinary kriging predictor. Since a model non-stationary in both expectation and variance appears as most representative for the observation, this conclusion make sense. The localized, shrinkage predictors are uniformily favorable to the corresponding localized traditional ones for the neighborhoods being studies. Moreover, the stationary, shrinkage predictors appear as favorable to the non-stationary shrinkages ones. Lastly, a neighborhood of about 10 observations are found to be suitable.
- in the Gamma ray study, one observation set of size 21 and a control set of 579 are used. By using the crossvalidation based criteria, we conclude that the localized, stationary, shrinkage predictors are clearly favorable to the other predictors. The evaluation results from the control set of data, can not confirm this conclusion, though. We have no clear explanation for this result.

We conclude that whenever a localized predictor is prefered - either due to non-stationaries in the expectation and variance fields, or due to need for computational efficiency - one should use a localized, stationary, shrinkage CVC predictor. Since this predictor is also reliable for stationary models, the recommendation may as well be to always use localized, stationary, shrinkage CVC predictors whenever there are a fair number of obervations.

References

[1] Chiles, J.P and Delfiner,J.P; *Geostatistics: Modeling spatial uncertainty(wiley series in probability and statistics)*, Wiley,New York), 695p, 1999.

[2] Efron, B. and Morris, C.; "Stein's estimation rule and its competitors-An empirical bayes approach", Journal of the American Statistical Association, Vol. 68, No. 341 (Mar., 1973), 117-130.

[3] Fuentes M. A.; "formal test for nonstationarity of spatial stochastic processes", Journal of Multivariate Analysis. 2005; 96:3055.

[4] Gribov, A. and Krivoruchko, K. ; "Geostatistical Mapping with Continuous Moving Neighborhood", Published in Mathematical Geology, Volume 36, Number 2, February 2004.

[5] http://www.image.ucar.edu/GSP/Data/US.monthly.met/

[6] Johns, C. J., Nychka, D., Kittel, T. G. F., and Daly, C. (2003). Infilling sparse records of spatial fields. Journal of the American Statistical Association, 98(464):796806.

[7] Journel, A.G and Huijbregts, C.J; *Mining Geostatistics*, Academic Press Inc, London , 1978.

[8] Mardia, K. V., Kent, J. T., and Bibby, J. M.; "Multivariate analysis," Academic press, London, 518p, 1979.

[9] Stein, M. L. ; "The Screening Effect in Kriging", The Annals of Statistics, Vol. 30, No. 1 (Feb., 2002), pp. 298-323

[10] Røislien, J. and Omre, H.;"T-distributed random fields: A parametric model for heavy-tailed well-log data," Mathematical Geosciences, 38(7), 821-849, 2006.

Appendices

A Posterior pdf

The conditional pdf for $[s^2 \mid \mathbf{r}_o]$ is:

$$\begin{aligned}
f(s^2 \mid \mathbf{r}_o) &= const \times f(\mathbf{r}_o \mid s^2) f(s^2) \\
&= const \times [2\pi]^{-\frac{n_o}{2}} \left| s^2 \left[\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T \right] \right|^{-\frac{1}{2}} \\
&\times exp\left[-\frac{1}{2} \left[(\mathbf{r}_o - \mu_m i_{n_o})^T \left[s^2 \left[\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T \right] \right]^{-1} (\mathbf{r}^o - \mu_m i_{n_o}) \right] \right] \\
&\times \frac{\gamma_s^{\xi_s}}{\Gamma(\xi_s)} \left[s^2 \right]^{-(\xi_s+1)} exp\left[-\gamma_s [s^2]^{-1} \right] \\
&= const' \times [s^2]^{-(\xi_s + \frac{n_o}{2} + 1)} \\
&\times exp\left\{ -\left[\gamma_s + \frac{1}{2} \left[(\mathbf{r}_o - \mu_m i_{n_o})^T \left[\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T \right]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}) \right] \right] [s^2]^{-1} \right\} \\
&= IG(\xi_{s|o}, \gamma_{s|o})
\end{aligned}$$

with

$$\xi_{s|o} = \xi_s + \tfrac{n_o}{2}$$

$$\gamma_{s|o} = \gamma_s + \tfrac{1}{2} \left[(\mathbf{r}_o - \mu_m i_{n_o})^T \left[\Omega_{oo} + \tau_m i_{n_o} i_{n_o}^T \right]^{-1} (\mathbf{r}_o - \mu_m i_{n_o}) \right]$$

Part IV:

Appendix

Paper I.A

Unobserved heterogeneity in the power law nonhomogeneous Poisson process

Zeytu Gashaw Asfaw [1] and Bo Henry Lindqvist *

Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim, Norway

Abstract

A study of possible consequences of heterogeneity in the failure intensity of repairable systems is presented. The basic model studied is the nonhomogeneous Poisson process with power law intensity function. When several similar systems are under observation, the assumption that the corresponding processes are independent and identically distributed is often questionable. In practice there may be an unobserved heterogeneity among the systems. The heterogeneity is modeled by introduction of unobserved gamma distributed frailties. The relevant likelihood function is derived, and maximum likelihood estimation is illustrated. In a simulation study we then compare results when using a power law model without taking into account heterogeneity, with the corresponding results obtained when the heterogeneity is accounted for. A motivating data example is also given.

Key words: Repairable system, Likelihood function, Frailty, Monte Carlo simulation.

1 Introduction

In the reliability literature, systems are generally classified as either non-repairable or repairable (see, e.g., Ascher and Feingold [1]). Non-repairable systems are those that do not get repaired when they fail. Thus, non-repairable system can fail only once, and a lifetime model such as the Weibull distribution provides the distribution of the time to failure of such systems.

* Corresponding author.
Email address: bo@math.ntnu.no (Bo Henry Lindqvist).
[1] Email address: zeytu.gashaw@math.ntnu.no (Zeytu Gashaw Asfaw)

On the other hand, repairable systems are those systems (machines, industrial plants, software, etc.) which, in the event of a failure, can be restored to satisfactory operation by any action, including parts replacements or changes to adjustable settings. But, to what extent can the system perform after being returned back to its regular operation? We may have that the system's performance is in the same state that the system had at the start of the operation, which means an "as good as new" condition. Or, its performance may be returned to the same state as before the failure, which means an "as bad as old" condition.

The latter case is usually referred to as a "minimal repair", modeled by a nonhomogeneous Poisson process (NHPP). Minimal repair thus means that a failed system is restored just back to a functioning state, and after repair the system continues as "if nothing had happened". This implies that the likelihood of system failure, right after a failure and subsequent repair, is the same as it was immediately before the failure. Note that repair times in this kind of modeling are assumed to be negligible.

NHPP models, which are the main concern of this paper, are useful due to their flexible assumption that events are occurring randomly in time, with rates which may vary with time. This is in contrast to the more established homogeneous Poisson process (HPP), where the rate of events is constant in time.

The present paper is concerned with the problem of predicting the behaviour of a system based on failure data from several similar systems. There is a well established theory for analysis of data for NHPPs. But as discussed for example in Lindqvist [5], there may be unobserved heterogeneity between the monitored systems which, if overlooked, may lead to non-optimal or possibly completely wrong decisions. An intuitve way of interpreting heterogeneity is to imagine an unknown covariate, with values that may vary between systems, and leading to an unexpected variation in the failure intensity of the different processes (see, e.g., Slimacek and Lindqvist [8]). Still it is believed that heterogeneity has been neglected in many reliability applications, and it is the purpose of the present paper, through a simulation study and a real data set, to point to some of the consequences that may result from not including heterogeneity in a model for repairable systems.

A striking example of heterogeneity is given by some data presented by Bhattacharjee et al. [2], presenting failure data for motor operated closing valves in safety systems at two boiling water reactor plants in Finland. Failures of the type "External Leakage" were considered for 104 valves with a follow-up time of 9 years. The data shows an apparently unnormal variation in the number of failures per valve, suggesting a heterogeneity between valves. In their analysis, Bhattacharjee et al. (2003) stressed the importance of taking hetero-

geneity into consideration and concluded that even very simple models may describe the heterogeneous behavior successfully. For illustration, their data are given in Section 5 together with a statistical analysis using the approach of the present paper.

The paper is structured as follows. In Section 2 we give the formal definition of the NHPP and derive the likelihood function for the case of a power law intensity model, which will be the basic model considered in the paper. Section 3 introduces heterogeneity between systems, in the form of individual unobserved multiplicative "frailties" defined for each system, assumed independent and, for simplicity of exposition, gamma distributed with unit expectation. The likelihood function of data from several systems under heterogeneity is developed, leading to explicit expressions for the power law parameter estimates in the case where each process is observed on the same time frame. Section 4 is devoted to a simulation study; first giving an algorithm for simulation of data, and then performing a comprehensive simulation study with the aim of illustrating the main messages of the paper. The data from Bhattacharjee et al. [2] are analysed in Section 5, while some concluding remarks are given in the final Section 6.

2 The classical power law process

2.1 Characteristics of an NHPP model

An NHPP model is fully characterized by the intensity function $w(t)$, commonly denoted ROCOF (Rate of ocurrence of failures), see e.g. Rausand and Høyland [7]. It is furthermore convenient to introduce the cumulative rate function $W(t) = \int_0^t w(s)ds$, later called the CROCOF (cumulative ROCOF).

As is well known, the number of failures experienced in a time interval from 0 to t, $N(t)$, is Poisson-distributed with parameter $W(t)$, for any t, so that in particular $E[N(t)] = W(t)$ and $Var[N(t)] = W(t)$.

2.2 The power law NHPP

For illustration, we shall in this paper concentrate on the most celebrated parameterization of the NHPP, namely the power law model. One reason for its popularity is that the ROCOF as a function of t is of the same form as the hazard rate of a Weibull distribution. Hence the time to first failure of the power law NHPP is Weibull distributed. Because of this, the power law model

is sometimes denoted the Weibull process.

The CROCOF of the power law is given by (see e.g. Rausand and Høyland [7])

$$W(t) = \lambda t^{\beta}, \text{ for } \lambda > 0, \beta > 0. \tag{1}$$

Thus, by differentitation, the ROCOF of the power law process is

$$w(t) = W'(t) = \lambda\beta t^{\beta-1}.$$

This intensity function was introduced in Crow [3] as a stochastic model for the Duane reliability growth postulate. The parameter β in the power law model gives information about the system as follows; if $0 < \beta < 1$, then the system is *improving* (happy); if $\beta > 1$, then the system is *deteriorating* (sad); and if $\beta = 1$ the model reduces to an HPP.

2.3 Maximum likelihood estimation in the power law NHPP

Suppose that data are available from m independent systems governed by NHPPs with the same intensity function $w(t)$, where system j is observed in the time interval $[S_j, T_j]$, $j = 1, 2, ..., m$, with events observed at times $t_{1j}, t_{2j}, ..., t_{n_j j}$.

The likelihood function of these data is given by (see, e.g., Meeker and Escobar [6])

$$L = \prod_{j=1}^{m} \left\{ \prod_{i=1}^{n_j} w(t_{ij}) \right\} e^{-[W(T_j)-W(S_j)]}, \tag{2}$$

which is the product of the individual likelihoods of each of the m systems. The log-likelihood function, which is usually easier to work with, is hence

$$l = \log(L) = \sum_{j=1}^{m} \left[\left\{ \sum_{i=1}^{n_j} \log w(t_{ij}) \right\} - [W(T_j) - W(S_j)] \right].$$

For the power law model, with the parametization given in (1), the log-likelihood function is given by

$$l = n \log \lambda + n \log \beta + (\beta - 1) \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij} - \lambda \sum_{j=1}^{m} \left[T_j^{\beta} - S_j^{\beta} \right] \tag{3}$$

where $n = \sum_{j=1}^{m} n_j$

For simplicity we shall in the following assume that $S_j = 0$ and $T_j = \tau$ for $j = 1, \ldots, m$. Thus all the m processes are observed on the time interval from 0 to a fixed time τ. As we shall see, this simplifies several results, while the main ideas prevail. Using the standard method of finding the maximum likelihood estimators (MLEs) of $\hat{\lambda}$ of λ and $\hat{\beta}$ of β by setting the partial derivatives of the log-likelihood function with respect to each parameter equal to zero, we get from (3),

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - m\tau^{\beta} = 0,$$

which implies

$$\hat{\lambda} = \frac{n}{m\tau^{\hat{\beta}}} \tag{4}$$

Similarly we get the equation

$$\frac{\partial l}{\partial \beta} = \frac{n}{\beta} + \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij} - \lambda m \tau^{\beta} \log \tau = 0,$$

which by using (4) leads to

$$\hat{\beta} = \frac{n}{n \log \tau - \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log t_{ij}}.$$

This gives an explicit solution for $\hat{\beta}$, which can afterwards be substituted in the expression (4) for $\hat{\lambda}$.

We now consider the Fisher information matrix for the computation of variances and covariances of the MLEs. The Fisher information matrix is used to measure the amount of information that the observed data carries about the unknown parameters. It is defined as

$$I(\lambda, \beta) = E \begin{bmatrix} -\frac{\partial^2 l(\lambda,\beta)}{\partial \lambda^2} & -\frac{\partial^2 l(\lambda,\beta)}{\partial \lambda \partial \beta} \\ -\frac{\partial^2 l(\lambda,\beta)}{\partial \lambda \partial \beta} & -\frac{\partial^2 l(\lambda,\beta)}{\partial \beta^2} \end{bmatrix},$$

and a straightforward computation in our case implies that

$$I(\lambda,\beta) = E\begin{bmatrix} \frac{n}{\lambda^2} & m\tau^\beta \log\tau \\ m\tau^\beta \log\tau & \frac{n}{\beta^2} + \lambda m\tau^\beta (\log\tau)^2 \end{bmatrix}. \tag{5}$$

The only random element in the above matrix is n, so we need to substitute $E(n) = m\lambda\tau^\beta$ to get the final expression

$$I(\lambda,\beta) = m\tau^\beta \begin{bmatrix} \frac{1}{\lambda} & \log\tau \\ \log\tau & \lambda(\frac{1}{\beta^2} + (\log\tau)^2) \end{bmatrix}.$$

Standard theory for maximum likelihood tells us that the maximum likelihood estimates for λ and β are, for large samples, approximately normally distributed centered at the true parameter values and with variances given as the diagonal elements, respectively, of the inverse of the Fisher information matrix. It is hence possible to estimate those variances by inverting the estimated matrix $I\left(\hat{\lambda},\hat{\beta}\right)$.

Alternatively, one may invert the so called observed Fisher information matrix, which is the matrix (5) where one does not take the expectation, but instead uses the observed value of n, and substitute maximum likelihood estimates for the parameters.

3 Heterogeneity in the power law model

3.1 Heterogeneous NHPPs

Consider an NHPP with intensity function $w(t)$. With the inclusion of heterogeneity, this model is modified to assuming that the intensity is given by

$$w_a(t) = a\ w(t),$$

where $w(t)$ is the basic ("baseline") intensity function, and a is an unobserved positive constant, which may vary from system to system. More precisely, a is assumed to be a positive random variable with mean 1 and variance $\delta \geq 0$. The idea is that in the case of m systems, each system has its own value of a, i.e., $a_1, a_2, ..., a_m$, which are assumed to be independent draws from this distribution. The a_j are in survival analysis usually called "frailties" (Vaupel, Manton and Stallard [9]).

Although there are several potential distributions for the frailties a_j, we shall here apply the most commonly used one, namely the gamma distribution. The popularity of this distribution as a frailty distribution is due to both mathematical convenience and often good fit to actual data. There is, however, no physical justification to prefer gamma frailties instead of other models.

The density of the two-parameter gamma distribution is generally given as

$$f(a) = \frac{a^{k-1}e^{-\frac{a}{\theta}}}{\theta^k \Gamma(k)}$$

for $a > 0$, where $k > 0$ is the shape parameter and $\theta > 0$ is the scale parameter. The corresponding expected value and variance are, respectively, $k\theta$ and $k\theta^2$. Since we require $E(a) = 1$ and $Var(a) = \delta$, we use $k = 1/\delta$ and $\theta = \delta$. The density of a hence becomes

$$h(a) = \frac{a^{\frac{1}{\delta}-1}e^{-\frac{a}{\delta}}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}} \tag{6}$$

Figure 1 shows several densities of gamma distributions with expected value 1.

The likelihood function for data from m systems modeled by NHPPs, was given in (2). We now study the changes needed when including a frailty a. We use a similar argument as before, but now with $Var(a) = \delta$ as an additional parameter.

Now the likelihood function for system j, for given value of the frailty, a_j, is

$$L_j(a_j) = \left\{ \prod_{i=1}^{n_j} w(t_{ij}) \right\} a_j e^{-a_j[W(T_j)-W(S_j)]}.$$

Since a_j is an unobservable random variable, the contribution to the full likelihood from this system is obtained by unconditioning with respect to a_j, which in practice will mean to compute the expected value of $L_j(a_j)$ with respect to the distribution of a_j. Since, furthermore, a_j is gamma distributed with expected value 1, and hence has probability density function (6), the expected value of $L_j(a_j)$ is

$$\begin{aligned} L_j &= E\left[L_j(a_j)\right] \\ &= \int L_j(a_j)h(a_j)da_j \end{aligned}$$

$$= \int \left\{ \prod_{i=1}^{n_j} w(t_{ij}) \right\} a_j e^{-a_j[W(T_j)-W(S_j)]} \frac{a_j^{\frac{1}{\delta}-1} e^{\frac{-a_j}{\delta}}}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}} da_j$$
$$= \frac{\prod_{i=1}^{n_j} w(t_{ij})}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}} \int_0^\infty a_j^{r_j-1} e^{-a_j s_j} da_j$$

where $r_j = n_j + \frac{1}{\delta}$ and $s_j = W(T_j) - W(S_j) + \frac{1}{\delta}$.

Now it is easy to show that $\int_0^\infty a^{r-1}e^{-sa}da = \Gamma(r)/s^r$ for all $r, s > 0$, so we get

$$L_j = \frac{\prod_{i=1}^{n_j} w(t_{ij})}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}} \frac{\Gamma(n_j + \frac{1}{\delta})}{\left[W(T_j) - W(S_j) + \frac{1}{\delta}\right]^{n_j+\frac{1}{\delta}}}.$$

3.2 Maximum likelihood estimation for the heterogeneous power law NHPP

Specializing the above to the power law (1), we get

$$L_j = \frac{\lambda^{n_j}\beta^{n_j}\left(\prod_{i=1}^{n_j} t_{ij}\right)^{\beta-1}\Gamma(n_j + \frac{1}{\delta})}{\Gamma(\frac{1}{\delta})\delta^{\frac{1}{\delta}}\left[\lambda T_j^\beta - \lambda S_j^\beta + \frac{1}{\delta}\right]^{n_j+\frac{1}{\delta}}}$$

Further, assuming $S_j = 0, T_j = \tau$ for all j, and then taking log and summing over all the m systems, we obtain the full log-likelihood

$$l(\lambda, \beta, \delta) = n\log\lambda + n\log\beta + (\beta-1)\sum_{j=1}^{m}\sum_{i=1}^{n_j}\log t_{ij} + \sum_{j=1}^{m}\log\Gamma(n_j + \frac{1}{\delta})$$
$$- \left[m\log\Gamma(\frac{1}{\delta}) + m\frac{1}{\delta}\log\delta + \left[n + \frac{m}{\delta}\right]\log[\lambda\tau^\beta + \frac{1}{\delta}]\right]$$

In order to find the maximum likelihood estimators for λ, β, δ we first compute

$$\frac{\partial l\,(\lambda, \beta, \delta)}{\partial\lambda} = \frac{n}{\lambda} - \left[\frac{\tau^\beta}{\lambda\tau^\beta + \frac{1}{\delta}}\right]\left[n + \frac{m}{\delta}\right],$$

which when set to 0 implies

$$\hat{\lambda} = \frac{n}{m\tau^{\hat{\beta}}} \tag{7}$$

Next we compute

$$\frac{\partial l\,(\lambda,\beta,\delta)}{\partial \beta} = \frac{n}{\beta} + \sum_{j=1}^{m}\sum_{i=1}^{n_j} \log t_{ij} - \left[\frac{\lambda\tau^{\beta}\log\tau}{\lambda\tau^{\beta}+\frac{1}{\delta}}\right]\left[n+\frac{m}{\delta}\right]$$

which when set to 0, using (7), leads to

$$\hat{\beta} = \frac{n}{n\log\tau - \sum_{j=1}^{m}\sum_{i=1}^{n_j}\log t_{ij}}.$$

Thus, $\hat{\lambda}$ and $\hat{\beta}$ are exactly the same functions of the data as for the power law case without frailities. (Note that this would not be the case if the observation time intervals were not all equal for all the m processes.)

The partial derivative of the log likelihood with respect to δ involves the digamma function ψ defined by

$$\psi(x) = \frac{d}{dx}\log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)},$$

and we get

$$\begin{aligned}
\frac{\partial l\,(\lambda,\beta,\delta)}{\partial \delta} &= -\frac{1}{\delta^2}\sum_{j=1}^{m}\psi(n_j+\frac{1}{\delta}) + \frac{m}{\delta^2}\psi(\frac{1}{\delta}) - m\left[-\frac{1}{\delta^2}\log\delta + \frac{1}{\delta^2}\right] \\
&\quad - \left[-\frac{m}{\delta^2}\log[\lambda\tau^{\beta}+\frac{1}{\delta}] - \frac{1}{\delta^2}\left[\frac{n+\frac{m}{\delta}}{\lambda\tau^{\beta}+\frac{1}{\delta}}\right]\right] \\
&= -\frac{1}{\delta^2}\sum_{j=1}^{m}\psi(n_j+\frac{1}{\delta}) + \frac{m}{\delta^2}\psi(\frac{1}{\delta}) + \frac{m}{\delta^2}\log\delta \\
&\quad - \frac{m}{\delta^2} + \frac{m}{\delta^2}\log[\lambda\tau^{\beta}+\frac{1}{\delta}] + \frac{1}{\delta^2}\left[\frac{n+\frac{m}{\delta}}{\lambda\tau^{\beta}+\frac{1}{\delta}}\right] \\
&= -\frac{1}{\delta^2}\left\{\sum_{j=1}^{m}\psi(n_j+\frac{1}{\delta}) - m\psi(\frac{1}{\delta}) - m\log\delta + m\right\} \\
&\quad + \frac{1}{\delta^2}\left\{m\log\left[\lambda\tau^{\beta}+\frac{1}{\delta}\right] - \frac{n+\frac{m}{\delta}}{\lambda\tau^{\beta}+\frac{1}{\delta}}\right\}
\end{aligned}$$

The likelihood equation given by equating this to 0 is simplified by substituting the estimators for λ and β, which gives an equation of δ alone. No explicit expression for the maximum likelihood estimator $\hat{\delta}$ is available, however, so a numerical method like Newton-Raphson's method needs to be used.

4 Simulations

4.1 Simulate systems from the power law process with heterogeneity

The following probabilistic property of the NHPP can help us to simulate event times of an NHPP from that of an HPP. Namely, if $U_1, U_2, ...$ are the event times of an HPP with intensity 1, then it can be shown that $W^{-1}(U_1), W^{-1}(U_2), ...$ are the event times of an NHPP with CROCOF $W(t)$. Here the inverse function $W^{-1}(u)$ is uniquely determined from $W(t)$ if $w(t) > 0$ for all t.

We now apply this property to the power law NHPP with CROCOF defined by (1). Then it is seen that

$$W^{-1}(u) = (u/\lambda)^{1/\beta}.$$

Thus if $U_1, U_2, \ldots$ are the event times of an HPP with intensity 1, we obtain a simulated realization of the power law NHPP with given parameters λ and β as $(U_1/\lambda)^{1/\beta}, (U_2/\lambda)^{1/\beta}, \ldots$. Note that here that the HPP, $U_1, U_2, \ldots$ can be simulated by first drawing U_1 from an exponential distribution with parameter 1; then letting $U_2 = U_1 + V_2$ where V_2 is an independent draw from the exponential distribution with parameter 1; and so on by adding new independent V_i from the exponential distribution with parameter 1, until the boundary time τ is reached for the transformed variables $(U_i/\lambda)^{1/\beta}$.

In order to simulate from a power law process with gamma distributed frailty, we first draw the value of a for each process, and then for the jth process replace λ by $a_j\lambda$ in the above simulation strategy.

4.2 Simulated single processes

Throughout the simulation study we assume that there are $m = 20$ systems, each observed on the fixed time interval from 0 to $\tau = 10$. The failure processes will be power law NHPPs with basic ROCOF $\lambda\beta t^{\beta-1}$, for varying values of λ and β, but possibly with heterogeneity obtained by multiplying the system intensities by independent random variables a from the gamma distribution with expected value 1 and variance δ.

Figure 2 shows a simulation of a single power law process observed on the time interval [0,10], where parameter values are $\lambda = 2$ and $\beta = 1.5$. Figure 3 shows similarly a single process from a power law process with the same λ and β, but with a frailty parameter $\delta\ (= Var(a)) = 0.2$.

We also illustrate, for the data in Figure 2, the estimated parameters, and the observed Fisher infromation matrix. The ML estimates are $\hat{\lambda}$=1.9926 and $\hat{\beta}$=1.4999, while the observed information matrix is

$$I\left(\hat{\lambda},\hat{\beta}\right)=\begin{bmatrix} 0.0407 & -0.0082 \\ \\ -0.0082 & 0.0018 \end{bmatrix}$$

By inversion of this matrix we obtain on the diagonal the estimated standard errors of $\hat{\lambda}$ and $\hat{\beta}$, respectively, 0.2018 and 0.0423.

4.3 Simulation study

For each setup of parameters we consider $10,000$ simulations, each consisting of $m = 20$ systems. The results are shown in Table 1. For each simulation we estimate parameters and their standard errors by maximum likelihood, and report averages of these numbers based on the $10,000$ simulations. These numbers can hence be viewed as approximations of expected values of the parameters, which enables consideration of possible bias in the estimators. Further, the columns named by "St.D" give empirical standard errors of the corresponding 10,000 computed estimates, obtained as the square roots of empirical variances.

The table also gives the averages (approximation of expected values) of number of failures in the time interval $[0, 10]$ for each parameter combination, as well as the corresponding standard deviations, computed as square roots of the empirical variances. Note that each such number is based on $m \times 10,000$ simulated processes, i.e. 200,000 simulations.

In the discussion below we use the property that the maximum likelihood estimators $\hat{\lambda}$ and $\hat{\beta}$ are given by the same function of the data whatever be the value of δ (see Section 3.2. Thus, the estimates are the same whether we assume the power law model without frailty or the power law model with frailty. (It should be noted that this would not be the case if the observation intervals of the m systems differed). Here it enables us to reach at many interesting conclusions regarding heterogeneity.

The main conclusions to be drawn from the simulation study are given below.

Data	m	True Value			n		Estimates					
10000	20	λ	β	δ	Ave.	St.D	$\hat{\lambda}$		$\hat{\beta}$		$\hat{\delta}$	
							Ave.	St.D	Ave.	St.D	Ave.	St.D
		2	1.5	0	63.1451	8.0541	2.0059	0.2057	1.5008	0.0430	0.0021	0.0036
				0.1	63.0648	21.4421	2.0075	0.2456	1.5005	0.0421	0.0942	0.0352
				0.2	63.2006	29.6606	2.0042	0.2853	1.5013	0.0425	0.1897	0.0651
				0.4	63.0469	40.1756	2.0033	0.3450	1.5009	0.0426	0.3795	0.1208
				0.6	63.3880	49.3983	2.0063	0.4033	1.5014	0.0431	0.5718	0.1750
				0.8	63.0149	56.1749	2.0050	0.4452	1.5013	0.0430	0.7511	0.2223
				1	63.4131	64.1655	2.0033	0.5023	1.5021	0.0432	0.9309	0.2610
			1	0	20.0288	4.4658	2.0051	0.2526	1.0019	0.0509	0.0064	0.0119
				0.2	19.9885	9.9238	1.9987	0.3231	1.0023	0.0509	0.1879	0.0774
				0.4	20.2335	13.4745	2.0052	0.3802	1.0026	0.0508	0.3742	0.1305
				0.6	19.8679	15.8587	1.9968	0.4285	1.0043	0.0514	0.5520	0.1769
				0.8	20.1814	18.2127	1.9923	0.4691	1.0051	0.0518	0.7158	0.2107
				1	20.4585	20.8251	1.9843	0.5149	1.0083	0.0524	0.8679	0.2421
			0.75	0	11.2920	3.3512	1.9992	0.2653	0.7535	0.0507	0.0116	0.0215
				0.2	11.8200	6.8717	2.0535	0.3142	0.7502	0.0493	0.1888	0.0924
				0.4	11.2779	7.8143	1.9923	0.3934	0.7557	0.0519	0.3620	0.1381
				0.6	11.1454	9.2048	1.9846	0.4330	0.7572	0.0522	0.5234	0.1763
				0.8	11.1906	10.4320	1.9702	0.4873	0.7624	0.0539	0.6685	0.2054
				1	11.1918	11.4700	1.9590	0.5285	0.7663	0.0552	0.7917	0.2274

Table 1
Data simulated from the heterogeneity model with varying δ, and estimation done with the heterogeneity model in Section 5.

4.3.1 An ordinary power law is anticipated, while there may be an unrecognized heterogeneity

Consider first the situation where one thinks that the ordinary power law model is the true model, but that in reality there is heterogeneity between the $m = 20$ systems.

For each combination of λ and β is seen that, as δ increases, the average number of failures per system is approximately constant, which is in fact exactly true by a theoretical computation of expected values. However, the standard deviations (St.D) of the number of failures per system increase with δ.

This is caused by the heterogeneity, which in practice means that some system will have a higher failure intensity, while others will have a lower intensity than for the base case. (On average, the intensity will be the same as for the no heterogeneity case, however, since the variables a_i have expected value 1.)

Now let us consider the estimated parameters. It is remarkable that neither the expected value nor the standard error of $\hat{\beta}$ are much influenced by the heterogeneity. As regards $\hat{\lambda}$, this is close to the true value, 2, but is increasingly biased downwards as δ increases. This is most clearly seen when $\beta \neq 1$. Its standard error, on the other hand, clearly increases with δ for all cases. Practical implications of these results are:

(i) For predictions of number of failures of new systems, by an erroneous assumption of *no* heterogeneity, one gets too short predicted intervals for the number of failures in a given time period. In fact, the expected value is correctly estimated, but the variation could be much bigger than expected if heterogeneity is not accounted for. For example, suppose the true values are $\lambda = 2$ and $\beta = 0.75$. In a model not taking heterogeneity into account, the number of failures will be predicted to be in the interval (expected value $\pm$ 2 standard deviations) approximately from 4.5 to 18. If the true heterogeneity variance is, e.g., $\delta = 0.4$, this interval would be much wider, namely from 0 to 27.

(ii) A similar problem is seen for the estimation of λ. Here one would get a too optimistic estimate for the standard error by assuming no heterogeneity, and also a downward bias will be present in the estimate.

4.3.2 The correct model, a power law model with heterogeneity, is used for statistical inference

In this case, the table contains information on all the maximum likelihood estimators of the model, $\hat{\lambda}, \hat{\beta}, \hat{\delta}$.

It is seen that the maximum likelihood estimator $\hat{\delta}$ slightly underestimates the true value of δ. Further, its standared error increases with δ as should be expected. The conclusion is that the estimator $\hat{\delta}$ seems to behave quite satisfactorily. The properties of the estimators $\hat{\lambda}$ and $\hat{\beta}$ are in fact already discussed, since the formulas for their maximum likelihood estimators are the same with and without including δ in the model.

5 The data example from Bhattacharjee et al. (2003)

Recall the closing valve failure example (Bhattacharjee et al. [2]), which was considered in the introduction. There are $m = 104$ systems, each observed on the time interval [0,3286] (hours). The failure times are given in Table 2.

System #	Failure times							
1	610	614	943	2024	2087	2104	2399	2525
2	126	323	943	1132	2087	2399	2426	
3	860	915	1606	3181				
4	10	19	104	2352				
5	293	2567						
6	2434	2676						
7	1963							
8	1262							
9	2501							
10	1963							
11	132							
12	1623							
13	3127							
14	3211							
15	1225							
16	1222							
17-104	–							

Table 2
Failure times for 104 closing valves, with follow-up time of 3286 hours, at two boiling water reactor plants in Finland. Failures type is "External Leakage".

Let the model be as given in Section 3.2. Using the approach in that subsection we get $\hat{\lambda} = 4.594 \times 10^{-4}$, $\hat{\beta} = 0.8215$, $\hat{\delta} = 0.7207$. The estimate of δ thus reveals a considerable heterogeneity between the systems. This heterogeneity is also clearly visible from computation of the standard deviation of the number of failures per system. In fact, the empirical standard deviation of number of observed failures per system is 1.206. On the other hand, using the estimates for λ and β and assuming an ordinary power law model without heterogeneity, the standard deviation would be estimated to 0.3558. The latter number is obtained as the square root of the estimated cumulative intensity of the process

at time 3286, using the fact that expected value equals variance for a Poisson distributed random variable.

6 Conclusion

The motivation for the present paper is the fact that an unobserved heterogeneity between observed systems of the same kind may, if ignored, lead to wrong conclusions or bad predictions of system behaviour.

In the present paper we show, in a partly tutorial manner, how a possible heterogeneity between systems may be included in a statistical investigation of repairable systems. In order to simplify the exposition we considered the fairly standard case of a power law nonhomogeneous Poisson process with heterogeneity of the gamma type. The advantage of the approach is that it leads to relativly simple formulas and procedures, while main ideas, pitfalls and possible remedies are clearly demonstrated.

In a forthcoming paper we will in a similar simulation study explore the relation between models for heterogeneity as considered here, and a certain class of dynamic models, considered for example in a recent paper by Le Gat [4].

References

[1] Ascher H, Feingold H. Repairable systems reliability: modeling, inference, misconceptions and their causes. New York: M. Dekker, 1984.

[2] Bhattacharjee M, Arjas E, Pulkkinen U. Modeling heterogeneity in nuclear power plant valve failure data. In: Mathematical and Statistical Methods in Reliability (Lindqvist BH, Doksum KA, eds.) Singapore: World Scientific Publishing, 2003; pp 341-353.

[3] Crow LH. Reliability analysis for complex, repairable systems. In: Reliability and biometry (Proschan F, Serfling RJ, eds.) Philadelphia, Pennsylvania: SIAM, 1974; pp 379–410.

[4] Le Gat Y. Extending the Yule process to model recurrent pipe failures in water supply networks. Urban Water Journal, 2013, DOI: 10.1080/1573062X.2013.783088

[5] Lindqvist BH. On the statistical modelling and analysis of repairable systems. Statistical Science 2013; 21:532-551.

[6] Meeker WQ, Escobar LA. Statistical methods for reliability data. New York: Wiley, 1998.

[7] Rausand M, Høyland, A. System reliability theory: models, statistical methods, and applications. New York: Wiley, 2004.

[8] Slimacek V, Lindqvist BH. Failure rate of wind turbines modeled by homogeneous Poisson process with covariates and unobserved heterogeneity. In: Safety, Reliability and Risk Analysis: Beyond the Horizon. Proceedings of the European Safety an Reliability Conference, ESREL 2013. CRC Press 2014.

[9] Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 1979;16:439-454.

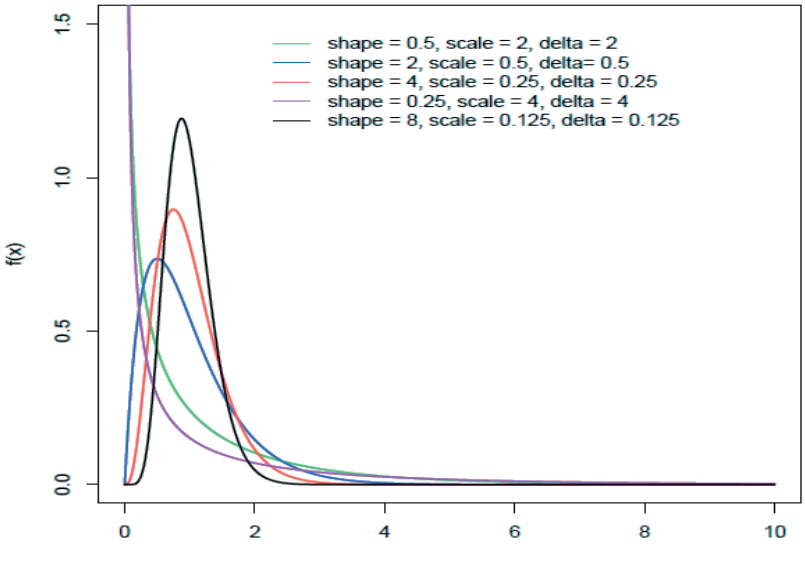

Fig. 1. Graphs of gamma densities with expected value 1

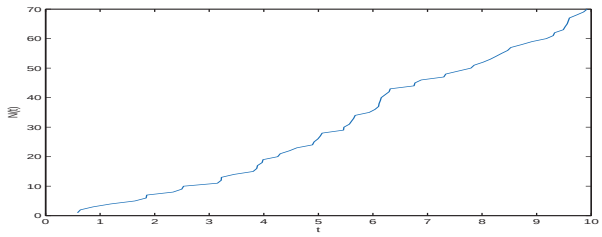

Fig. 2. Cumulative number of failures, $N(t)$, versus time for a power law process with $\lambda = 2$ and $\beta = 1.5$.

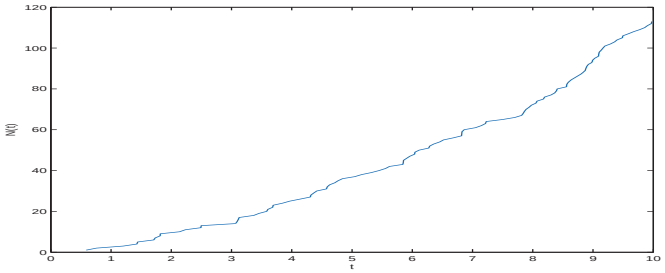

Fig. 3. Cumulative number of failures, $N(t)$, versus time for a single power law process with $\lambda = 2$, $\beta = 1.5$ and δ=0.2.

Paper I.B

Extending minimal repair models for repairable systems: A comparison of dynamic and heterogeneous extensions of a nonhomogeneous Poisson process

Zeytu Gashaw Asfaw [1] and Bo Henry Lindqvist *

Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim, Norway

Abstract

For many applications of repairable systems, the minimal repair assumption, which leads to nonhomogeneous Poisson processes (NHPP), is not adequate. We review and study two extensions of the NHPP, the dynamic NHPP and the heterogeneous NHPP. Both extensions are motivated by specific aspects of potential applications. It has long been known, however, that the two paradigms are essentially indistinguisable in an analysis of failure data. We investigate the connection between the two approaches for extending NHPP models, both theoretically and in data studies including simulated failure processes. In particular we review the computation of the likelihood functions for the two situations and demonstrate a somewhat surprising similarity between them. This similarity is empirically confirmed in the numerical examples.

Key words: Repairable system, LEYP model, frailty, likelihood function, power law NHPP, Monte Carlo simulation.

* Corresponding author.
Email address: bo@math.ntnu.no (Bo Henry Lindqvist).
[1] Email address: zeytu.gashaw@math.ntnu.no (Zeytu Gashaw Asfaw)

1 Introduction

1.1 Minimal repair model (NHPP)

Reliability analyses of repairable systems are commonly made under the assumption of *minimal repairs*, leading to nonhomogeneous Poisson process (NHPP) models. Minimal repair intuitively means that a failed system is restored just back to a functioning state, which is commonly named an "as bad as old" condition. In many cases, such models, although simple, are fully adequate for useful analyses. Indeed, the natural goal is often to return the system to a functioning condition as soon as possible, and it is therefore reasonable that the system after repair is brought to the exact same state it was in just before the failure. The concept of *minimal repair* was apparently introduced by Ascher [2], who considered that the "age" (in the sense of effectiveness) of the system after repair is the same as it was before failure. A classical comprehensive reference to repairable systems, and in particular to minimal repair, is the 1984-book by Ascher and Feingold [3].

In most of the theoretical derivations to follow, we will use a generic intensity function $w(t)$, which as we shall see will play the role of a *baseline* intensity for an NHPP. In applications involving maximum likelihood estimation, the $w(t)$ will typically be a parametric function. In Sections 4 and 5 we shall consider the celebrated *power law* intensity,

$$w(t;\lambda,\beta) = \lambda\beta t^{\beta-1} \text{ for } t > 0, \tag{1}$$

where $\lambda > 0, \beta > 0$ are parameters.

1.2 Heterogeneous extension of NHPP

Asfaw and Lindqvist [4] noted that when conclusions are to be drawn from several similar systems, there may be an unobserved heterogeneity between the systems which, if overlooked, may lead to wrong predictions for future systems of the given kind.

Consider a set of systems, e.g. pumps of the same make, for which one models the failure process by an NHPP with intensity function $w(t)$. The common way of introducing heterogeneity is to modify the intensity for a particular system to the heterogeneous NHPP model by assuming that the intensity is given by

$$w_h(t) = a\, w(t),$$

where $w(t)$ is the *baseline* intensity function, while a is an unobserved positive constant corresponding to the system under study, and which may vary from system to system. Thus, in the case of m systems, each system is assumed to have its own value of the "frailty" a, i.e., $a_1, a_2, ..., a_m$, which in the mathematical model are assumed to be independent realizations from a probability distribution with mean 1 and variance $\delta \geq 0$, say.

In practice, the a_j may be thought of as coming from some unobserved covariate. The randomness of the a_j may well be due to a difference in meterological conditions; or it might be due to different maintenance strategies or even more unquantifiable effects such as differences in maintenance philosophy or attitude. It may of course also be due to variation in the quality of the production of the units themselves.

Asfaw and Lindqvist [4] demonstrated how to derive the likelihood function for data from m power law systems under the assumption of a heterogeneity as described above, when the a_j are independent and gamma-distributed. By maximizing the likelihood function one may then estimate the parameters of the baseline hazard (1), as well as the variance parameter δ of the unobserved a_j.

1.3 Dynamic extension of NHPP

Motivated by applications to urban drinking water infrastructures, Le Gat [8,9] suggested a seemingly different way of modifying the commonly used NHPP methods (as studied, e.g, by Røstum [11]). Water pipe failures are considered to be highly dependent on past events, in the sense that repairs may be harmful, leading to a possible accumulation of failures. Le Gat [8,9] therefore defined the following extension of an NHPP, involving the desired dependency on past failures. For further studies of this model we also refer to Babykina and Couallier [5,6].

The definition of Le Gat's failure process is given in the form of a *conditional intensity function*, here denoted $w_d(\cdot)$, where d stands for *dynamic*,

$$w_d(t) = (1 + \gamma N(t-))w(t). \tag{2}$$

As is common notation, $N(t-)$ denotes the number of failures experienced in the time interval from 0 to t, not including t, while $w(t)$ is the *baseline* intensity function. The interpretation of $w_d(t)$ as a conditional intensity function is as follows,

$$P(\text{failure in time interval } (t, t+h) \mid \text{failure history until time } t)$$

$$\approx (1 + \gamma N(t-))w(t)h$$

for small h. The meaning of this is that, if we imagine to be present at time t (or more precisely, immediately before t), and get to know the exact failure history until that time, then we predict a failure to happen in the interval $(t, t+h)$ with probability $(1 + \gamma N(t-))w(t)h$.

Note the distinction in general between the conditional intensity and the common definition of the ROCOF $r(t)$ (*rate of occurence of failures*, e.g, Ascher and Feingold [3]), which is the corresponding unconditional intensity. The latter can be interpreted as follows. Suppose we are present at time 0 and want to predict whether a failure will occur in the time interval $(t, t+h)$. Then the relevant probability of this is approximately $r(t)h$. For an NHPP, the conditional intensity and the ROCOF are the same, while for the dynamic extension of an NHPP as defined in (2), they are different. (The ROCOF of the dynamic NHPP may in fact be rather difficult to compute).

The resulting "non-memoryless property" of the dynamic NHPP has an important implication as regards predictions of future failures. Consider two disjoint time intervals $[a, b]$ and $[c, d]$ with $a < b < c < d$. While for an NHPP, the number $N(c, d)$ of failures in $[c, d]$ is independent of the number of failures $N(a, b)$ in $[a, b]$, Le Gat [9] showed that for the dynamic NHPP, the failure counts $N(a, b)$ and $N(c, d)$ are *not* independent, with $N(c, d)$ given $N(a, b) = k$ having a negative binomial distribution depending on k.

Following Le Gat [8,9], the failure process defined by (2) may be viewed as an extension of the well known Yule process, which is a birth process with a constant birth rate, and [8,9] therefore denoted the process (2) by the LEYP process (Linear Extension of the Yule Process). We will in the following denote the process by the *dynamic* NHPP model, in accordance with Aalen et al. [1, Chapter 8].

1.4 Heterogeneous versus dynamic extension of NHPP

As indicated above, the motivations for the heterogeneous and the dynamic extensions of the NHPP are quite different. The heterogeneous model assumes that there is a fixed frailty a for each system. As was pointed out in [4], a "footprint" of heterogeneity is a larger variation in number of failures per system than expected from an ordinary NHPP. Such a variation may, however, as well be caused by an underlying dynamic model, where the failure behavior depends on the history of failures in the way that failures may escalate through time.

Despite the apparent difference between the two paradigms, it is a fact that one is not able to distinguish between the two kinds of departures from the ordinary NHPP model. Thus, unexpected variations in number of failures may either be caused by an unobserved covariate, leading to the heterogeneous model; or by a dynamic effect caused by failures in the past. (Or, of course, the truth may be a mixture of the two effects).

Thus, although the heterogeneous and dynamic models may look very different, they also have several similarities. It is the purpose of the present paper to point to such similarities, and to perform a simulation study in order to illustrate certain effects of failure data coming from the dynamic model, while the heterogeneous model is fitted.

1.5 Organization of the paper

The paper is organized as follows. In Section 2 we review the derivation of the likelihood function of the dynamic NHPP, as given, e.g., in Le Gat [9]. We also reconsider the likelihood function of the heterogeneous NHPP as given in [4]. In Section 3 we compare the two likelihood functions, pointing to a somewhat surprising similarity. Then, following Aalen et al. [1], we compare the conditional intensities of the two models, identifying a corresponding similarity. Section 4 reconsiders the example of Bhattacharjee et al. [7], also considered in [4]. Both the dynamic and the heterogeneous models are fitted, both leading to apparently satisfactory fits. In Section 5 we present results of a simulation study, after first having given a recipe for simulation from the dynamic model. The paper is ended by some concluding remarks in Section 6.

2 Likelihood functions of the extended NHPP models

2.1 The likelihood function for the dynamic NHPP

Assume that we have observed a repairable system on the time interval $[0, \tau]$. Let $n(t)$ be the realization of the process $N(t)$ counting failures in $[0, t]$, and let the observed failure times be $t_1, t_2, \ldots, t_{n(\tau)}$. The likelihood function for the data under the dynamic model (2) is given in Le Gat [9], but for completeness we go through the complete derivation below.

Note that even if the process (2) is no longer an NHPP, the formula for the likelihood function in terms of the (conditional) intensity is the same as the

one for an ordinary NHPP (see, e.g., Lindqvist [10]),

$$L_d = \left\{ \prod_{i=1}^{n(\tau)} w_d(t_i) \right\} e^{-W_d(\tau)},$$

where $W_d(\tau) = \int_0^\tau w_d(u)du$.

Consider first

$$\begin{aligned}
W_d(\tau) &= \int_0^\tau w_d(u)du = \int_0^\tau (1+\gamma n(u))w(u)du \\
&= \sum_{k=1}^{n(\tau)} \int_{t_{k-1}}^{t_k} (1+\gamma(k-1)))w(u)du + \int_{t_{n(\tau)}}^{\tau} (1+\gamma n(\tau)))w(u)du \\
&= \sum_{k=1}^{n(\tau)} (1+\gamma(k-1))(W(t_k) - W(_{k-1})) + (1+\gamma n(t))(W(\tau) - W(t_{n(t)})) \\
&= \sum_{k=1}^{n(\tau)} (W(t_k) - W(t_{k-1})) + (W(\tau) - W(t_{n(t)})) \\
&+ \gamma \sum_{k=1}^{n(\tau)} (k-1)(W(t_k) - W(t_{k-1})) + \gamma n(\tau)(W(\tau) - W(t_{n(\tau)})) \\
&= W(\tau) + \gamma \left[n(\tau)W(\tau) - \sum_{k=1}^{n(\tau)} W(t_k) \right]. \qquad (3)
\end{aligned}$$

Next, consider

$$\begin{aligned}
\prod_{i=1}^{n(\tau)} w_d(t_i) &= \prod_{i=1}^{n(\tau)} (1+\gamma n(t_i))w(t_i) \\
&= \left\{ \prod_{i=1}^{n(\tau)} (1+(i-1)\gamma) \right\} \prod_{i=1}^{n(\tau)} w(t_i) \\
&= \frac{\Gamma(\frac{1}{\gamma} + n(\tau))}{\Gamma(\frac{1}{\gamma})} \gamma^{n(\tau)} \prod_{i=1}^{n(\tau)} w(t_i). \qquad (4)
\end{aligned}$$

Here $\Gamma(\cdot)$ is the gamma-function, and the last equality is proven by repeated use of the relation $\Gamma(k+1) = k\Gamma(k)$, valid for all $k > 0$.

By combining (3) and (4), we get the likelihood for a single system,

$$L_d = \left\{ \prod_{i=1}^{n(\tau)} w(t_i) \right\} \cdot \frac{\Gamma(\frac{1}{\gamma} + n(\tau))}{\Gamma(\frac{1}{\gamma})} \cdot \gamma^{n(\tau)} \cdot \frac{e^{\gamma \sum_{i=1}^{n(\tau)} W(t_i)}}{e^{(1+\gamma n(\tau))W(\tau)}}$$
$$= \left\{ \prod_{i=1}^{n(\tau)} w(t_i) e^{\gamma W(t_i)} \right\} \cdot \frac{\Gamma(\frac{1}{\gamma} + n(\tau))}{\Gamma(\frac{1}{\gamma})} \cdot \gamma^{n(\tau)} \cdot \frac{1}{e^{(1+\gamma n(\tau))W(\tau)}}. \quad (5)$$

In practice we will have data from m independent systems of the same kind. Suppose the jth is observed in the time interval $[0, \tau_j]$, with events observed at times $t_{1j}, t_{2j}, ..., t_{n_j(\tau_j)}$, $j = 1, 2, ..., m$. The total likelihood for the m systems is given by a product of the likelihood (5) for each system, thus giving

$$\mathbf{L}_d = \prod_{j=1}^{m} L_{dj},$$

where

$$L_{dj} = \left\{ \prod_{i=1}^{n_j(\tau_j)} w(t_{ij}) e^{\gamma W(t_{ij})} \right\} \cdot \frac{\Gamma(\frac{1}{\gamma} + n_j(\tau_j))}{\Gamma(\frac{1}{\gamma})} \cdot \gamma^{n_j(\tau_j)} \cdot \frac{1}{e^{(1+\gamma n_j(\tau_j))W(\tau_j)}}.$$

If a parametric form of $w(t)$ is given, then $\mathbf{L}_d$ is the basis of maximum likelihood estimation of the parameters of $w(t)$ in addition to γ. If the power law function (1) is chosen for $w(t)$, then the resulting model has been called the Yule-Weibull model by Le Gat [9].

2.2 The likelihood function for the heterogeneous NHPP

We reconsider the heterogeneous NHPP model as described in Section 1.2 and treated more in detail in the paper by Asfaw and Lindqvist [4]. Following [4] (see also [10]), the likelihood function for each single system can be written,

$$L_h = \left\{ \prod_{i=1}^{n(\tau)} w(t_i) \right\} \cdot \frac{\Gamma(\frac{1}{\delta} + n(\tau))}{\Gamma(\frac{1}{\delta})} \cdot \frac{1}{\delta^{\frac{1}{\delta}} \left[\frac{1}{\delta} + W(\tau) \right]^{\frac{1}{\delta} + n(\tau)}}.$$

Rewriting the denominator of the last factor of the above expression, we obtain the alternative expression

$$L_h = \left\{ \prod_{i=1}^{n(\tau)} w(t_i) \right\} \cdot \frac{\Gamma(\frac{1}{\delta} + n(\tau))}{\Gamma(\frac{1}{\delta})} \cdot \delta^{n(\tau)} \cdot \frac{1}{[1 + \delta W(\tau)]^{\frac{1}{\delta} + n(\tau)}}. \quad (6)$$

As for the dynamic NHPP, the likelihood for m systems is given by the product of factors (6), one for each system. (In fact, fitting of heterogeneous NHPPs make sense only if data from several systems ($m > 1$) are available. This also applies to the likelihood (5) for the dynamic case.)

3 Comparison of the dynamic and heterogeneous NHPPs

3.1 Comparison of likelihood functions

If we equate the parameters γ and δ, then the two likelihood functions (5) and (6) differ only in the first and last factors.

In (6), the denominator of the last factor can be approximated as follows,

$$\begin{aligned}[1+\delta W(\tau)]^{\frac{1}{\delta}+n(\tau)} &= [1+\delta W(\tau)]^{\frac{1}{\delta}}\,[1+\delta W(\tau)]^{n(\tau)} \\ &\approx e^{W(\tau)} \cdot e^{\delta n(\tau) W(\tau)} = e^{(1+\delta n(\tau))W(\tau)}, \end{aligned} \tag{7}$$

where the approximation can be justified when $\delta > 0$ is small and $n(\tau)$ is large, with $\delta n(\tau)$ moderate.

If we replace the denominator of the last factor in (6) by (7), then (again equating γ and δ), the expressions (5) and (6) differ only in the first factors, which are, respectively, $\left\{\prod_{i=1}^{n(\tau)} w(t_i) e^{\delta W(t_i)}\right\}$ for the dynamic model and $\left\{\prod_{i=1}^{n(\tau)} w(t_i)\right\}$ for the heterogeneous model.

Now rename the baseline intensity of (5) as $w^*(t)$ and assume that

$$w(t) = w^*(t) e^{\delta W^*(t)}. \tag{8}$$

Then (5) and the approximated version of (6) are the same, except that the $W(\tau)$ in (7) should be replaced by $W^*(t)$. This approximation can be justified by noting that (8) implies

$$W(t) = \frac{e^{\delta W^*(t)} - 1}{\delta} \approx W^*(t)$$

for small δ.

Thus the dynamic model with baseline intensity $w^*(t)$ and the heterogeneous model with intensity $w(t)$ given by (8) are approximately the same in the

sense that the likelihood functions are approximately equal, and the maximum likelihood estimates of, respectively, γ and δ of the two models are hence similar. Note that this similarity requires the connection between the baseline intensities given by (8). We shall see below that this connection also pops up in the comparison of conditional intensity functions of the two models.

3.2 Comparison of conditional intensity functions given the failure history

By using the definition of a conditional intensity function as indicated in Section 1.3, Aalen et al. [1, page 333] computed the conditional intensity of the heterogeneous model with baseline intensity $w(t)$ and gamma-distributed frailties with expected value 1 and variance δ (we will not consider the computations here), to get

$$\lambda(t|\text{history}) = w(t) \cdot \frac{\frac{1}{\delta} + N(t-)}{\frac{1}{\delta} + W(t)} = \frac{w(t)}{1 + \delta W(t)} \cdot (1 + \delta N(t-)). \tag{9}$$

An intuitive interpretation of this result might be useful. Recall that the probability mechanism that produces the heterogeneous NHPP can be described as follows. First, draw a value of a from the given frailty distribution with expected value 1 and variance δ (usually a gamma-distribution). Then run an ordinary NHPP with intensity function $aw(t)$, where $w(t)$ is the given baseline intensity. Since a is unobserved, we do not know at the start of the process whether there will be few or many failures, since this to a large extent will depend on the value of a. Thus if we visit the process at some time t (or rather immediately before t), and want to predict the future based on observation of the process until time t, then the number of failures so far, $N(t-)$, is intuitively very relevant, with a high number indicating an increased probability of a new failure. This effect is precisely what is formulated in (9).

We next note that, after equating the parameters γ and δ, (9) equals the conditional intensity of a *dynamic* model with baseline intensity function

$$w^*(t) = \frac{w(t)}{1 + \delta W(t)}.$$

Solving this equation for $w(t)$ we get in fact precisely (8), thus confirming the informal finding in the previous subsection, and again implying that that heterogeneous model and the dynamic model are in some sense equivalent. We will in the following pursue this idea in a real data example and in a simulation study.

4 Data example (Bhattacharjee et al. [7])

This example was also considered in [4]. It concerns failure data for motor operated closing valves in safety systems at two boiling water reactor plants in Finland. Failures of the type "External Leakage" were considered for 104 valves with a follow-up time of 9 years. The data, given in Table 1, show an apparently abnormal variation in the number of failures per valve, and were therfore analyzed in [7] using a (Bayesian) heterogeneity model, and in [4] using a power law heterogeneous NHPP with gamma distributed frailties. The latter analysis is reproduced here, together with an analysis using the power law dynamic NHPP model described in Section 1.3.

For both the dynamic and the heterogeneous model we use the power law baseline intensity (1).

The maximum likelihood estimates for the two models are given in Table 2. It is remarkable that the estimates of the γ and δ are fairly equal, while the estimated baseline intensities are very different. This is in accordance with the findings in the previous section, although we have not investigated the connection further. It is also remarkable that the estimates of γ and δ are rather high. This is of course caused by the large variation in the number of failures per valve. As a final remark, the computed maximum values of the log likelihood functions of the two models are quite similar, informally suggesting that the models are in some sense equally good (with the dynamic model having a slight advantage).

A way of checking the realism of the two estimated models is to compute, for each of them, the estimated expected number of systems which will have no failures. This obviously equals $104 \cdot P(T_1 > 3286)$.

For the dynamic model, we note that T_1 is simply Weibull-distributed with reliability function $P(T_1 > t) = e^{-\lambda t^\beta}$, so the estimated value for $t = 3286$ is

$$e^{-0.00175 \cdot 3286^{0.563}} = 0.846,$$

giving an estimated expected number of systems with 0 failures given as $104 * 0.846 = 88.0$ (which is exactly the same as the observed one!)

For the heterogeneous model we need to condition upon the value of the frailty variable a which is gamma-distributed. In order to get the estimated value of $P(T_1 > 3286)$ this amounts to computing the expected value

$$E[e^{-\hat{\lambda} \cdot 3286^{\hat{b}} \cdot a}]$$

when a is gamma-distributed with expected value 1 and variance $\hat{\delta}$. This gives 0.848 and hence the estimated number of systems with 0 failures is now 88.2, very close to what we obtained above.

5 Simulation study

5.1 Simulation of failure data from the dynamic NHPP

Consider the process with conditional intensity function given by (2). Suppose the process is observed on the time interval $[0, \tau]$, with N failures occurring at successive times $T_1, T_2, \ldots, T_N$. For convenience, put $T_0 = 0$.

Then for $i = 1, 2, \ldots, N$, for given T_{i-1}, the interfailure time $T_i - T_{i-1}$ has the hazard rate function

$$(1 + (i-1)\gamma)w(t + T_{i-1}) \text{ for } t > 0.$$

Thus, given T_{i-1}, we can simulate a value of T_i by first drawing U_i from the uniform distribution on $[0, 1]$ and solving for T_i the equation

$$U_i = \exp\{-(1 + (i-1)\gamma)(W(T_i) - W(T_{i-1})\}. \tag{10}$$

Here we used the fact that the survival function of a lifetime with hazard rate $z(t)$ equals $\exp\{-\int_0^t z(u)du\}$. (Note that (10) is also clear from results in Le Gat [9]).

It follows from (10) that

$$W(T_i) - W(T_{i-1}) = \frac{-\ln U_i}{1 + (i-1)\gamma}$$

so

$$T_i = W^{-1}\left(W(T_{i-1}) - \frac{\ln U_i}{1 + (i-1)\gamma}\right) \tag{11}$$

for $i = 1, 2, \ldots, N$, where N is the smallest i such that $T_{i+1} > \tau$.

Since it is used in the simulation study below, we display the special case of

(11) when the baseline is given by the power law intensity (1). Then

$$T_i = \left(T_{i-1}^{\beta} - \frac{\ln U_i}{\lambda(1 + (i-1)\gamma)} \right)^{1/\beta} \text{ for } i = 1, 2, \ldots, N. \tag{12}$$

5.2 Simulation scenarios

Throughout the simulation study we assume that there are $m = 20$ systems, each observed on the fixed time interval from 0 to $\tau = 10$. The failure processes will be simulated using (12) with varying values of λ, β and γ.

5.3 Simulation study

For each setup of parameters we consider $10,000$ simulations, each consisting of $m = 20$ systems. The results are shown in Table 3. For each simulation we estimate parameters by maximum likelihood, and report averages of these numbers based on the $10,000$ simulations. These numbers can hence be viewed as approximations of expected values of the parameters, which enables consideration of possible bias in the estimators. Further, the columns named by "St.D" give empirical standard errors of the corresponding 10,000 computed estimates, obtained as the square roots of empirical variances.

Table 3 also gives the averages of number of failures for the single processes in the time interval $[0, 10]$, for each parameter combination. Also, the corresponding standard deviations, computed as square roots of the empirical variances, are given. Note that each such number is based on $m \times 10,000$ simulated processes, i.e. 200,000 simulations.

One aspect that we wanted to study with the simulations is the connection between the parameters γ and δ. It is seen from the table that throughout, there is a remarkably good correspondence between the two. This applies, however, not to the estimate parameters of the power law baseline functions. This should however be expected from the results of Section 3. This effect was also seen in the data example of Section 4. Table 3 shows in particular that the estimates of β increase substantially when γ increases. This is natural since an increasing γ by the definition (2) induces an increasing trend in the failure occurrences. The estimates $\hat{\lambda}$ have a similar effect as compensators for the variation in failures, but less interpretable.

It is clear from the displays of averages and standard deviations of number of failures per system that the simulation study involves systems with high number of failures. Thus the interesting values of γ are small compared to

estimated value in the data example of Section 4. Regarding number of failures, it is seen that the average number increases with γ, which of course is natural. Thus a direct comparison of standard deviations is not so interesting. It is seen, however, that the coefficient of variation (St.D/Ave) are consistently increasing with γ, which also seems reasonable.

6 Conclusion

In this paper we have reviewed and compared two extensions of the NHPP model, the heterogeneous NHPP and the dynamic NHPP. Each of these models are in the literature well motivated by specific applications, shown to imply better fit to data than ordinary NHPPs in certain real cases. For example, Le Gat [8,9] found the dynamic NHPP and its underlying idea to be relevant for applications to failures of water pipes. Bhattacharjee et al. [7] found the heterogeneous extension of the NHPP to give the most plausible explanation of a certain data set for valve failures.

It is therefore interesting to notice that, theoretically, the two paradigms are not distinguishable given ordinary failure data for several systems. As noted by Aalen et al. [1, Section 8.5], this dilemma has in fact been recognized for a long time in many different application areas, ranging from psychiatry to economics.

One aspect of this paper has been to point at the importance of challenging the ordinary NHPP model when appropriate. While the NHPP is usually characterized as a "black box" model, the two extensions considered in this paper are based on exploiting extra insight regarding the particular applications. The essential equivalence of the two studied models, may of course be viewed as a source of further insight and perhaps further investigations regarding the underlying failure mechanisms.

References

[1] Aalen O, Borgan Ø, Gjessing H. Survival and event history analysis: a process point of view. New York: Springer, 2008.

[2] Ascher H. Evaluation of a repairable system reliability using "bad-as-old" concept. IEEE Transactions on Reliability 1968;17,103–110.

[3] Ascher H, Feingold H. Repairable systems reliability: modeling, inference, misconceptions and their causes. New York: M. Dekker, 1984.

[4] Asfaw ZG, Lindqvist BH. Unobserved heterogeneity in the power law nonhomogeneous Poisson process. Submitted manuscript.

[5] Babykina G, Couallier V. Modelling recurrent events for repairable systems under worse than old assumption. In: Advances in Degradation Modeling (Nikulin MS et al., eds.) Statistics for Industry and Technology, DOI 10.1007/978-0-8176-4924-1_22, Birkhauser Boston, 2010.

[6] Babykina G, Couallier V. Modelling pipe failures in water distribution systems: Accounting for harmful repairs and a time-dependent covariate. International Journal of Performability Engineering 2014; 10(1):41-52.

[7] Bhattacharjee M, Arjas E, Pulkkinen U. Modeling heterogeneity in nuclear power plant valve failure data. In: Mathematical and Statistical Methods in Reliability (Lindqvist BH, Doksum KA, eds.) Singapore: World Scientific Publishing, 2003; pp 341-353.

[8] Le Gat Y. Une extension du processus de Yule pour la modélisation stochastique des événements récurrents – application aux défaillances de canalisations d'eau sous pression. Ph.D. thesis, École Nationale du Génie Rural, des Eaux et des Fôrets (ENGREF), Paris, 2009.

[9] Le Gat Y. Extending the Yule process to model recurrent pipe failures in water supply networks. Urban Water Journal, 2013, DOI: 10.1080/1573062X.2013.783088

[10] Lindqvist BH. On the statistical modelling and analysis of repairable systems. Statistical Science 2013; 21:532-551.

[11] Røstum J. Statistical modelling of pipe failures in water networks. Dissertation (Dr.-Engr). Department of Hydraulic and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway, 2000.

System #	Failure times							
1	610	614	943	2024	2087	2104	2399	2525
2	126	323	943	1132	2087	2399	2426	
3	860	915	1606	3181				
4	10	19	104	2352				
5	293	2567						
6	2434	2676						
7	1963							
8	1262							
9	2501							
10	1963							
11	132							
12	1623							
13	3127							
14	3211							
15	1225							
16	1222							
17-104	–							

Table 1
Failure times for 104 closing valves, with follow-up time of 3286 hours, at two boiling water reactor plants in Finland. Failures type is "External Leakage".

Model	$\hat{\lambda}$	$\hat{\beta}$	$\hat{g}$	$\hat{\delta}$	$\log L$
Dynamic	$1.75 \cdot 10^{-3}$	0.563	8.25		-340.32
Heterogeneous	$4.59 \cdot 10^{-4}$	0.822		8.34	-340.79

Table 2
Maximum likelihood estimates of parameters of, respectively, the dynamic model and the heterogeneous model applied to the valve failure data of [7]. The last column shows the maximum of the corresponding log likelihood functions.

Data	m	True Value			n		Estimates					
10000	20	λ	β	γ	Ave.	St.D	$\hat{\lambda}$		$\hat{\beta}$		$\hat{\delta}$	
							Ave.	St.D	Ave.	St.D	Ave.	St.D
		2	1.5	0	63.3176	8.0255	2.0030	0.2019	1.5017	0.0423	0.0020	0.0038
				0.001	65.2978	8.2883	1.9562	0.1972	1.5254	0.0425	0.0026	0.0042
				0.01	88.4093	12.8219	1.5084	0.1668	1.7694	0.0459	0.0111	0.0080
				0.02	127.0301	21.0328	0.9979	0.1159	2.1078	0.0480	0.0229	0.0119
				0.04	287.3850	61.2510	0.2698	0.0349	3.0325	0.0516	0.0387	0.0133
				0.06	719.8430	176.3796	0.0354	0.0044	4.3141	0.0479	0.0569	0.0194
			1	0	19.9410	4.3837	2.0018	0.2592	1.0019	0.0513	0.0062	0.0114
				0.001	20.4110	4.6004	1.9959	0.2521	1.0097	0.0509	0.0069	0.0118
				0.01	21.9510	5.0593	1.9584	0.2518	1.0558	0.0514	0.0135	0.0174
				0.02	24.5330	6.0879	1.9214	0.2596	1.1113	0.0537	0.0238	0.0221
				0.04	30.8490	8.4581	1.8012	0.2496	1.2348	0.0551	0.0474	0.0296
				0.06	37.9590	11.3169	1.6259	0.2481	1.3816	0.0589	0.0682	0.0334
				0.08	50.1180	16.0797	1.4173	0.2131	1.5465	0.0587	0.0763	0.0305
				0.1	64.0240	22.7574	1.1807	0.1888	1.7359	0.0616	0.0943	0.0361
				0.2	269.6050	120.1785	0.2495	0.0397	3.0342	0.0576	0.1902	0.0602
			0.75	0	11.2950	3.3210	2.0022	0.2644	0.7536	0.0490	0.0117	0.0215
				0.001	11.3640	3.5082	2.0046	0.2663	0.7550	0.0513	0.0126	0.0229
				0.01	11.8860	3.5045	2.0044	0.2715	0.7769	0.0520	0.0164	0.0253
				0.02	12.5950	4.1497	2.0310	0.2757	0.7970	0.0511	0.0261	0.0320
				0.04	14.3780	4.8219	2.0294	0.2925	0.8478	0.0541	0.0441	0.0410
				0.06	16.0900	5.8230	2.0615	0.3080	0.8961	0.0544	0.0668	0.0461
				0.08	18.5330	7.1101	2.0409	0.3171	0.9561	0.0574	0.0942	0.0534
				0.2	40.9700	19.5547	1.6950	0.2931	1.4002	0.0611	0.1937	0.0709
				0.4	219.4870	136.0986	0.5884	0.1085	2.5791	0.0532	0.3823	0.1157

Table 3
Data simulated from the dynamic model with varying γ, and estimation done with the heterogeneity model.

Paper II.A

Localized/Shrinkage Kriging Predictors

Zeytu Gashaw Asfaw [1] and Henning Omre *

Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim, Norway

Abstract

The objective of the study is to improve the robustness and flexibility of spatial kriging predictors with respect to deviations from spatial stationarity assumptions. A predictor based on a non-stationary Gaussian random field is defined. The model parameters are inferred in an empirical Bayesian setting, using observations in a local neighborhood and a prior model assessed from the global set of observations. The localized predictor appears with a shrinkage effect and is coined a localized/shrinkage kriging predictor. The predictor is compared to traditional localized kriging predictors in a case study on observations of annual accumulated precipitation. A crossvalidation criterion is used in the comparision. The shrinkage predictor appears as clearly preferable to the traditional kriging predictors. A simulation study on prediction in non-stationary Gaussian random fields is conducted. The results from this study confirms that the shrinkage predictor is favorable to the traditional one. Moreover, the crossvalidation criterion is found to be suitable for selection of predictor. Lastly, the computational demands of localized predictors are very modest, hence the localized/shrinkage predictors are suitable for large scale spatial prediction problems.

Key words: Spatial statistics, Gaussian random fields, Bayesian inference, Conjugate models.

1 Introduction

Spatial prediction is required in many applications, and examples can be found in natural resource mapping, meteorology and image analysis. Consider a re-

* Corresponding author.
Email address: omre@math.ntnu.no (Henning Omre).
[1] Email address: zeytu.gashaw@math.ntnu.no (Zeytu Gashaw Asfaw)

gionalized variable $\{r(x); x \in \mathbf{D} \subset \Re^m\}$ where $r(x) \in \Re^1$ is the variable of interest while x is a reference variable in the domain $\mathbf{D}$. The challenge is to predict the regionalized variable from a set of observations $\mathbf{r}_o = [r(x_1), ..., r(x_n)]$; $x_1, ..., x_n \in \mathbf{D}$. We define the predictors in a probabilistic setting and can also obtain associated predictor uncertainties.

The classical probabilistic approach to spatial prediction is kriging, see Journel and Huijbregts (1978) and Chiles and Delfiner (1999). The traditional ordinary kriging predictor is based on a stationary model for the regionalized variable, with spatially constant expectation and variance, and a shift invariant spatial correlation function. Localized predictors, local neighborhood kriging, see Chiles and Delfiner (1999) can be defined to robustify the predictor with respect to deviations from the stationarity assumptions. The major challenge in using localized predictors is to define the size of the local neighborhood where a bias/variance trade-off must be made. For some spatial correlation structures a screening effect is provided by the observations closest to the prediction location, see Stein (2002), and this effect may robustify the localization. In Anderes and Stein (2011) a weighted localized approach is defined and demonstrated to be useful for non-stationary random fields.

Recent developments in computer and sensor technology have provided enormous spatio-temporal sets of observations, see Johns et al (2003). Computational demands in spatial prediction have been a critical factor and considerable research is devoted to numerical algorithmes for large sparse correlation matrices. Localized predictors may provide an alternative solution to reduce these computational demands.

In the current study we define spatial predictors which are robust with respect to deviations from assumptions about spatial stationary. We define a Gaussian random field with spatially varying expectation and variance. The spatial correlation function is shift invariant and known. We assess the expectation and variance locally by a sliding window approach. In traditional local neighborhood kriging this assessment is done by some maximum likelihood procedure. The new feature of the study is that these local assessments are done by shrinkage estimators in an empirical Bayes setting, inspired by the study of Efron and Morris (1973). The hierarchical, Gaussian random field model discussed in Røislien and Omre (2006) is used locally and the hyperparameters are assessed from the global set of observations. Since the predictor is locally defined it is extremely computationally effecient. We denote the spatial predictor localized/shrinkage kriging.

In the two first sections, various Gaussian random field models for the regionalized variable $\{r(x); x \in \mathbf{D}\}$ are presented with associated model parameter estimators. In the following section, five spatial predictors are defined, one global and four localized. Two of these predictors appear with shrinkage. The

next section defines the evaluation criteria used in the comparison of the predictors. In the next to last section the five predictors are compared on a real set of annual accumulated precipitation data and in a synthetic simulation study. Thereafter, in the last section, the conclusions are forwarded. The paper summarises the major findings in an extended study, reported in Asfaw and Omre (2014).

2 Predictor Models

The spatial predictors will be based on probabilistic models for the regionalized variable $\{r(x); x \in \mathbf{D}\}$.

In traditional kriging prediction $\{r(x); x \in \mathbf{D}\}$ is associated with a stationary Gaussian random field, see Chiles and Delfiner (1999). This assumption entails that:

$$\begin{aligned} E\left[r(x)\right] &= \mu; \quad \forall x \in \mathbf{D} \\ Var\left[r(x)\right] &= \sigma^2; \quad \forall x \in \mathbf{D} \\ Corr\left[r(x'), r(x'')\right] &= \rho(x' - x''); \quad \forall x', x'' \in \mathbf{D} \end{aligned} \tag{1}$$

where expected level $\mu \in \Re^1$, variance level $\sigma^2 \in \Re^1_+$ and spatial correlation function $\rho(x' - x'')$ is positive definite. Note that for these model assumptions the random field is shift invariant, and this property is extensively used to make inference about the model parameters $[\mu, \sigma^2, \rho(.)]$ from the set of observations $\mathbf{r}_o$. This traditional kriging model may be extended to have an expectation surface $\mu(x) = \sum_{l=1}^{L} \alpha_e g_e(x)$ where $\{g_e(x); x \in \mathbf{D}\}$; $l = 1, ..., L$ are known basis surfaces while $\alpha = (\alpha_1, ..., \alpha_L)$ are unknown coefficients. This model corresponds to a spatial regression model, and the shift invariance property is lost, which complicates model parameter inference. Note that for given correlation function $\rho(.)$, maximum likelihood estimates based on $\mathbf{r}_o$ is analytically assessable for the other model parameters, under both these model assumptions.

In the current study it is assumed that $\{r(x); x \in \mathbf{D}\}$ is associated with a general Gaussian random field, which entails that:

$$\begin{aligned} E\left[r(x)\right] &= \mu(x); \quad \forall x \in \mathbf{D} \\ Var\left[r(x)\right] &= \sigma^2(x); \quad \forall x \in \mathbf{D} \\ Corr\left[r(x'), r(x'')\right] &= \rho(x', x''); \quad \forall x', x'' \in \mathbf{D} \end{aligned} \tag{2}$$

with $\mu \in \Re^1$, $\sigma^2 \in \Re^1_+$ and $\rho(x^{'}, x^{''})$ being positive definite. There is obviously lack of shift invariance under these assumptions. We can expect that inference of the spatial model parameters $\{\mu(x); x \in \mathbf{D}\}$, $\{\sigma^2(x); x \in \mathbf{D}\}$ and $\left\{\rho(x^{'}, x^{''}); \quad \forall x^{'}, x^{''} \in \mathbf{D}\right\}$, based on $\mathbf{r}_o$ is complicated. If we fix the correlation function $\rho(.,.)$ we may use localized estimators in a kernel spirit to assess the spatial model parameters $\mu(.)$ and $\sigma^2(.)$. When selecting the size of localization we face a bias/variance trade-off. Large local neighborhood introduce bias in the estimators due to smoothing while small neighborhoods introduce instability in the estimators due to sensoring of observations. In the current study we adress this bias/variance trade-off by defining shrinkage estimators in an empirical Bayes setting.

In order to define the shrinkage estimators we introduce a stationary, hierarchical Gaussian random field model for a local neighborhood $\mathbf{D}_+$ around an arbitrary $x_+ \in \mathbf{D}$, i.e for $\{r(x); x \in \mathbf{D}_+\}$. In this model the expected level m and the variance level s^2 are considered to be random variables. Moreover, $\{[r(x) \mid m, s^2]; x \in \mathbf{D}_+\}$ is defined to be a stationary Gaussian random field, hence:

$$\begin{aligned} E\left[r(x) \mid m, s^2\right] &= m; \quad \forall x \in \mathbf{D}_+ \\ Var\left[r(x) \mid m, s^2\right] &= s^2; \quad \forall x \in \mathbf{D}_+ \\ Corr\left[r(x^{'}), r(x^{''}) \mid m, s^2\right] &= \rho(x^{'} - x^{''}); \quad \forall x^{'}, x^{''} \in \mathbf{D}_+ \end{aligned} \tag{3}$$

By assuming that $\rho(.)$ is known, and that the model parameters $[m \mid s^2]$ and s^2 have prior models which are Gaussian and Inverse Gamma respectively, with:

$$\begin{aligned} E\left[m \mid s^2\right] &= \mu_m \\ Var\left[m \mid s^2\right] &= \tau_m s^2 \\ E\left[s^2\right] &= [\xi_s - 1]^{-1}\gamma_s \\ Var\left[s^2\right] &= [(\xi_s - 1)^2(\xi_s - 2)]^{-1}\gamma_s^2 \end{aligned} \tag{4}$$

where the model parameters are $\mu_m \in \Re^1_+$, $\tau_m \in \Re^1_+$, $\xi_s \in 2 + \Re^1_+$ and $\gamma_s \in \Re^1_+$. The prior model on $[m, s^2]$ is a conjugate model for the stationary Gaussian random field, and the marginal random field $\{r(x); x \in \mathbf{D}_+\}$ will be a T-dist random field, see Røislien and Omre (2006), and be analytically tractable.

The estimators for the model parameters $\mu(x_+)$ and $\sigma^2(x_+)$ in arbitrary location $x_+ \in \mathbf{D}$ are localized to observations in $\mathbf{D}_+$ and with shrinkage according to the localized hierarchical Gaussian model. These estimators will of course

depend on the parameters of the prior model $[\mu_m, \tau_m, \xi_s, \gamma_s]$, and we assess these parameters in an empirical Bayesian spirit from a set of local neighborhood $\mathbf{D}_+$ covering $\mathbf{D}$.

3 Inference of Model Parameters

In order to use the probabilistic models for $\{r(x); x \in \mathbf{D}\}$ to spatial prediction the model parameters must be assessed from the set of observations represented by the $[n \times 1]$-vector $\mathbf{r}_o = [r(x_1), ..., r(x_n)]^T = [r_1, ..., r_n]^T$.

Throughout the study we assume the spatial correlation function $\left\{\rho(x', x'') = \rho(x' - x''); \quad \forall x', x'' \in \mathbf{D}\right\}$ to be known, with associated inter-observation correlation $[n \times n]$-matrix Ω_{oo}. This correlation model must of course also be infered in one way or the other, but in our study we do not account for the uncertainty related to this.

For a stationary Gaussian random field, given the correlation function, the model parameters can be assessed by a maximum likelihood estimator:

$$\hat{\mu} = [\mathbf{i}_n^T \Omega_{oo}^{-1} \mathbf{i}_n]^{-1} [\mathbf{i}_n^T \Omega_{oo}^{-1} \mathbf{r}_o] \tag{5}$$

$$\hat{\sigma}^2 = \tfrac{1}{n} [\mathbf{r}_o - \hat{\mu} \mathbf{i}_n]^T \Omega_{oo}^{-1} [\mathbf{r}_o - \hat{\mu} \mathbf{i}_n]$$

For a general Gaussian random field, given the correlation function, the assessments of the spatial model parameters $\{\mu(x); x \in \mathbf{D}\}$ and $\{\sigma^2(x); x \in \mathbf{D}\}$, are more complicated. We prescribe localized estimators. Consider an arbitrary location $x_+ \in \mathbf{D}$ and parameterize the localization by the k observations in $\mathbf{r}_o$ localized closest to x_+, and denote it as k-localization. Define a binary-selection $[k \times n]$-matrix G_+^k such that $G_+^k \mathbf{r}_o$ is a $[k \times 1]$-vector containing the k observations in the k-localization of x_+. Note that G_+^k easily can be extended to also account for favorable configurations of observations around x_+. The k-localized maximum likelihood estimators for the model parameters are:

$$\hat{\mu}_+^k = \hat{\mu}^k(x_+) = [\mathbf{i}_k^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} G_+^k \mathbf{r}_o] [\mathbf{i}_k^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} \mathbf{i}_k]^{-1} \tag{6}$$

$$\hat{\sigma}_+^{k2} = \hat{\sigma}^{k2}(x_+) = \tfrac{1}{k} [G_+^k \mathbf{r}_o - \hat{\mu}_+^k \mathbf{i}_k]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} (G_+^k \mathbf{r}_o - \hat{\mu}_+^k \mathbf{i}_k)$$

By letting x_+ coinside with the observation locations $[x_1, ..., x_n]$ we obtain the estimators for observation expectation $[n \times 1]$-vector and for diagonal standard deviation $[n \times n]$-matrix respectively.

$$\hat{\boldsymbol{\mu}}_o^k = \left[\hat{\mu}_1^k, ..., \hat{\mu}_n^k\right]^T \tag{7}$$

$$\hat{\Gamma}_o^k = \begin{bmatrix} \hat{\sigma}_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\sigma}_n^k \end{bmatrix}$$

For a stationary, hierarchical Gaussian random field, given the correlation function, we need to assess the model parameters of the prior model $[\mu_m, \tau_m, \xi_s, \gamma_s]$. We choose to make this assessment in an empirical Bayes setting, by considering the localized estimates in the observation locations, $[\hat{\mu}_i^k, \hat{\sigma}_i^{k2}]$; $i = 1, ..., n$, to be a super population of the k-localized estimate in an arbitrary location. Then natural estimators for the parameters of the prior model are dependent on k:

$$\hat{\mu}_m^k = \frac{1}{n} \mathbf{i}_n^T \hat{\boldsymbol{\mu}}_o^k \tag{8}$$

$$\hat{\tau}_m^k = [\hat{\sigma}_\cdot^{k2}]^{-1} \hat{\sigma}_m^{k2}$$

where

$$\hat{\sigma}_\cdot^{k2} = \tfrac{1}{n} Tr[\hat{\Gamma}_o^k]$$

$$\hat{\sigma}_m^{k2} = \tfrac{1}{n} [\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k \mathbf{i}_n]^T [\hat{\boldsymbol{\mu}}_o^k - \hat{\mu}_m^k \mathbf{i}_n]$$

and

$$\hat{\xi}_s^k = [\hat{\sigma}_s^{k2}]^{-1} [\hat{\mu}_s^{k2}] + 2 \tag{9}$$

$$\hat{\gamma}_s^k = \hat{\mu}_s^k \left[[\hat{\sigma}_s^{k2}]^{-1} [\hat{\mu}_s^{k2}] + 1 \right]$$

where

$$\hat{\mu}_s^k = \tfrac{1}{n} \mathbf{i}_n^T \mathbf{s}^2$$

$$\hat{\sigma}_s^{k2} = \tfrac{1}{n} [\mathbf{s}^2 - \hat{\mu}_s^k \mathbf{i}_n]^T [\mathbf{s}^2 - \hat{\mu}_s^k \mathbf{i}_n]$$

$$\mathbf{s}^2 = \left[(r_1 - \hat{\mu}_m^k)^2, ..., (r_n - \hat{\mu}_m^k)^2 \right]^T$$

We have now defined estimators for all model parameters based on the observations $\mathbf{r}_o$. Hence all probabilistic models for the regionalized variable $\{r(x); x \in \mathbf{D}\}$ are fully specified. Focus is on spatial prediction however, and in the following Section we define spatial predictors under the various model assumptions and insert the estimators for the model parameters to obtain operable predictors.

4 Spatial Predictors

Focus of the study is on spatial prediction in a random field $\{r(x); x \in \mathbf{D}\}$ based on a set of observations represented in a $[n \times 1]$-vector $\mathbf{r}_o$. Consider an arbitrary location $x_+ \in \mathbf{D}$ with value $r(x_+) = r_+$. The challenge is to provide a reliable predictor for r_+ based on $\mathbf{r}_o$. We use a squared error loss, hence the predictor is $\hat{r}_+ = \widehat{E}[r_+ \mid r_o]$, with associated estimated predictor variance $\hat{\sigma}^2_+ = \widehat{V}ar[r_+ \mid r_o]$. Note that a predictor for the entire regionalized variable $\{r(x); x \in \mathbf{D}\}$ can be obtained by letting x_+ run over the domain $\mathbf{D}$.

Recall that the correlation function $\left\{\rho(x^{'} - x^{''}); x^{'}, x^{''} \in \mathbf{D}\right\}$ is assumed known and that the inter observation correlation $[n \times n]$-matrix is denoted Ω_{oo}, while the observation to x_+ correlation $[n \times 1]$-vector is denoted ω_{o+}. Recall also that the localization operator G^k_+ is defined such that $G^k_+ \mathbf{r}_o$ is a observation $[k \times 1]$-vector which contain the k observations located closest to x_+.

Glob/Stat/Trad Predictor
This predictor is global and based on a stationary Gaussian random field model with traditional parameter estimates. It corresponds to the frequently used global ordinary kriging predictor, and it is defined by:

$$[r \mid \mathbf{r}_o] \sim Gauss[\mu_{+|o}, \sigma^2_{+|o}]$$

with

$$\mu_{+|o} = \mu + \omega^T_{o+} \Omega^{-1}_{oo} [\mathbf{r}_o - \mu \mathbf{i}_n] \tag{10}$$

$$\sigma^2_{+|o} = \sigma^2 [1 - \omega^T_{o+} \Omega^{-1}_{oo} \omega_{o+}]$$

Note that the predictor is independent of the variance σ^2 while the prediction variance is independent of the observed values $\mathbf{r}_o$ only the location configuration. These are well recognized characteristics of kriging.

The Glob/Stat/Trad predictor with associated predictor variance is defined as:

$$\hat{r}_{GST} = \hat{\mu}_{+|o}$$

$$\hat{\sigma}^2_{GST} = \hat{\sigma}^2_{+|o}$$

which are defined by Expression (10) with the estimates in Expression (5) inserted.

Loc/Stat/Trad Predictor
This predictor is k-localized and based on a stationary Gaussian random field model with traditional parameter estimators. It corresponds to a localized or-

dinary kriging predictor which is frequently used in practice, and it is defined by:

$$[r_+ \mid G_+^k \mathbf{r}_o] \sim Gauss[\mu_{+|o}^k, \sigma_{+|o}^{k2}]$$

with

$$\mu_{+|o}^k = \mu_+^k + [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} \left[G_+^k \mathbf{r}_o - \mu_+^k \mathbf{i}_k \right] \tag{11}$$

$$\sigma_{+|o}^{k2} = \sigma_+^{k2} \left[1 - [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} G_+^k \omega_{o+} \right]$$

Also this predictor is locally independent on σ^2 with predictor variance locally independent on the observed values $\mathbf{r}_o$. Since both μ_+^k and σ_+^{k2} will vary across the field, the predictor and predictor variance will vary across the field.

The Loc/Stat/Trad predictor with associated predictor variance is defined as:

$$\hat{r}_{LST}^k = \hat{\mu}_{+|o}^k$$

$$\hat{\sigma}_{LST}^{k2} = \hat{\sigma}_{+|o}^{k2}$$

which are defined by Expression (11) with the estimates in Expression (6) inserted.

Loc/Stat/Shr Predictor
This predictor is k-localized and based on a stationary Gaussian random field model with shrinkage parameter estimators defined in an empirical Bayes setting. The predictor is denoted the stationary localized/shrinkage kriging predictor and it constitutes one new predictor in the study:

$$[r_+ \mid m_+^k, s_+^{k2}, G_+^k \mathbf{r}_o] \sim Gauss[\mu_{+|o}^k, \sigma_{+|o}^{k2}]$$

with

$$\mu_{+|o}^k = m_+^k + [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} \left[G_+^k \mathbf{r}_o - m_+^k \mathbf{i}_k \right] \tag{12}$$

$$\sigma_{+|o}^{k2} = s_+^{k2} \left[1 - [G_+^k \omega_{o+}]^T [G_+^k \Omega_{oo} [G_+^k]^T]^{-1} G_+^k \omega_{o+} \right]$$

where the posterior expectations for the hyper-parameters are, see Appendix A:

$$\begin{aligned} m_+^k &= E[m \mid s^2, G_+^k \mathbf{r}_o] \\ &= \mu_m^k + \tau_m^k \mathbf{i}_k^T \left[\tau_m^k \mathbf{i}_k \mathbf{i}_k^T + [G_+^k \Omega_{oo} [G_+^k]^T \right]^{-1} \left[G_+^k \mathbf{r}_o - \mu_m^k \mathbf{i}_k \right] \end{aligned} \tag{13}$$

$$s_+^{k2} = E[s_+^{k2} \mid G_+^k \mathbf{r}_o] = \left[\xi_s^k + \tfrac{k}{2} - 1 \right]^{-1}$$

$$\times \left[\gamma_s^k + \tfrac{1}{2}\left[[G_+^k\mathbf{r}_o - \mu_m^k\mathbf{i}_k]^T\left[[G_+^k\Omega_{oo}[G_+^k]^T] + \tau_m^k\mathbf{i}_k\mathbf{i}_k^T\right]^{-1}[G_+^k\mathbf{r}_o - \mu_m^k\mathbf{i}_k]\right]\right]$$

Expression (12) follows from the definition of a stationary hierarchical Gaussian random field conditioned on the model parameters and on $G_+^k\mathbf{r}_o$. Expression (13) follows from the prior model of $[m, s^2]$ being conjugate, hence the posterior models are analytically assessible, and so are their expectations. Note in particular that $m_+^k = E[m \mid G_+^k\mathbf{r}_o]$ hence independent of s_+^{k2}. Note also that the predictor is independent of the variance while the predictor variance is dependent on the actual observed values.

The predictor appears with shrinkage through the estimators for shift and scaling parameters m_+^k and s_+^{k2}. The actual observations weights are independent of m_+^k and s_+^{k2}.

The Loc/Stat/Shr predictor with associate predictor variance is defined as:

$$\hat{r}_{LSS}^k = \hat{\mu}_{+|o}^k$$

$$\hat{\sigma}_{LSS}^{k2} = \hat{\sigma}_{+|o}^{k2}$$

which are defined by Expressions (12) and (13) with the estimators in Expressions (8) and (9) inserted.

Loc/Non-stat/Trad Predictor
This predictor is k-localized and based on a general Gaussian random field model with traditional parameter estimators. This predictor is surprisingly seldom used, and it is defined as:

$$[r_+ \mid G_+^k\mathbf{r}_o] \sim Gauss[\mu_{+|o}^k, \sigma_{+|o}^{k2}]$$

with

$$\mu_{+|o}^k = \mu_+^k + \sigma_+^k[G_+^k\Gamma_o^k\omega_{o+}]^T[G_+^k\Gamma_o^k\Omega_{oo}\Gamma_o^k[G_+^k]^T]^{-1}\left[G_+^k[\mathbf{r}_o - \boldsymbol{\mu}_o^k]\right] \tag{14}$$

$$\sigma_{+|o}^{k2} = \sigma_+^{k2}\left[1 - [G_+^k\Gamma_o^k\omega_{o+}]^T[G_+^k\Gamma_o^k\Omega_{oo}\Gamma_o^k[G_+^k]^T]^{-1}G_+^k\Gamma_o^k\omega_{o+}\right]$$

where

$$\boldsymbol{\mu}_o^k = \left[\mu_1^k, ..., \mu_n^k\right]^T$$

$$\Gamma_o^k = \begin{bmatrix} \sigma_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^k \end{bmatrix}$$

Note that this predictor is dependent of the variance variability across the field while the predictor variance is independent of the observation values.

The Loc/Non-stat/Trad predictor with associated predictor variance are defined as:

$$\hat{r}^k_{LNT} = \hat{\mu}^k_{+|o}$$

$$\hat{\sigma}^{k2}_{LNT} = \hat{\sigma}^{k2}_{+|o}$$

which are defined by Expression (14) with the estimates in Expressions (6) and (7) inserted.

Loc/Non-stat/Shr Predictor
This predictor is k-localized and based on a non-stationary Gaussian random field model with shrinkage parameter estimators defined in an empirical Bayesian setting. The predictor is denoted the non-stationary localized/shrinkage kriging predictor and it constitutes another new predictor in the study:

$$[r_+ \mid m^k_+, \mathbf{m}^k_o, s^{k2}_+, S^k_o, G^k_+\mathbf{r}_o] \sim Gauss[\mu^k_{+|o}, \sigma^{k2}_{+|o}]$$

with

$$\mu^k_{+|o} = m^k_+ + s^k_+[G^k_+S^k_o\omega_{o+}]^T[G^k_+S^k_o\Omega_{oo}S^k_o[G^k_+]^T]^{-1}\left[G^k_+[\mathbf{r}_o - \mathbf{m}^k_o]\right] \qquad (15)$$

$$\sigma^{k2}_{+|o} = s^{k2}_+\left[1 - [G^k_+S^k_o\omega_{o+}]^T[G^k_+S^k_o\Omega_{oo}S^k_o[G^k_+]^T]^{-1}G^k_+S^k_o\omega_{o+}\right]$$

where m^k_+ and s^{k2}_+ are defined as in Expression (13) while

$$\mathbf{m}^k_o = \left[m^k_1, ..., m^k_n\right]^T$$

$$S^k_o = \begin{bmatrix} s^k_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s^k_n \end{bmatrix}$$

are defined from the x_+-centred m_+ and s^{k2}_+ shifted to the observation locations $[x_1, ..., x_n]$.

The predictor appears with shrinkage through the estimators of m^k_+ , s^{k2}_+ and m^k_i , s^{k2}_i; $\quad i = 1, ..., n$. Hence both shift and scaling, as well as observation weights are influenced by the shrinkage effect. This is a full shrinkage predictor.

The Loc/Non-stat/Shr predictor with associated predictor variance are defined as:

$$\hat{r}^k_{LNS} = \hat{\mu}^k_{+|o}$$

$$\hat{\sigma}^{k2}_{LNS} = \hat{\sigma}^{k2}_{+|o}$$

which are defined by Expression (15) with the estimates in Expressions (8) and (9) inserted.

Crossvalidation Calibrated (CVC) Predictors
We have defined five predictors, one global and four localized. One challenge with localized predictors is lack of global anchoring. The variance estimates are localized and coupled with the expectation estimates which may cause wrong scaling of the prediction variances. We introduce crossvalidation calibrated (CVC) predictors to provide a global calibration.

Consider an arbitrary predictor in an arbitrary location $x_+ \in \mathbf{D}, \hat{r}_+ = \widehat{E}[r_+ \mid \mathbf{r}_o]$. The predictor may be global or localized, and it is based on the observations $\mathbf{r}_o = [r_1, ..., r_n]$ in locations $[x_1, ..., x_n]$. The associated prediction variance is $\hat{\sigma}^2_+ = \widehat{V}ar[r_+ \mid \mathbf{r}_o]$.

Define the crossvalidation predictors with associated predictor variances in the observation locations $[x_1, ..., x_n]$:

$$\hat{r}_i = \widehat{E}[r_i \mid \mathbf{r}_{o(-i)}] \quad ; \quad i = 1, ..., n$$

$$\hat{\sigma}^2_i = \widehat{V}ar[r_i \mid \mathbf{r}_{o(-i)}] \quad ; \quad i = 1, ..., n$$

where $\mathbf{r}_{o(-i)}$ represents observations $\mathbf{r}_o$ with observations r_i removed.

The normalized crossvalidation errors are defined as:

$$e_i = [\hat{\sigma}_i]^{-1}[r_i - \hat{r}_i] \quad ; \quad i = 1, ..., n$$

Under a fully specified model, these errors will be centred at zero and be scaled to unity. Consider the estimators:

$$\hat{\mu}_e = \frac{1}{n}\sum_{i=1}^n e_i$$

$$\hat{\sigma}^2_e = \frac{1}{n}\sum_{i=1}^n [e_i - \hat{\mu}_e]^2$$

and the $\hat{\mu}_e$ and $\hat{\sigma}^2_e$ should be close to zero and unity, respectively.

We define the CVC predictor and CVC prediction variance in an arbitrary location $x_+ \in \mathbf{D}$ by:

$$\tilde{r}_+ = \hat{r}_+ + \hat{\sigma}_+ \hat{\mu}_e \tag{16}$$

$$\tilde{\sigma}^2_+ = \hat{\sigma}^2_e \hat{\sigma}^2_+$$

Note that the corresponding normalized crossvalidation errors will have $\hat{\mu}_e$ and $\hat{\sigma}_e^2$ which are identical to zero and unity, respectively. These CVC predictors with associated CVC prediction variances will be used in the study.

5 Evaluation Criteria

We have defined several spatial CVC predictors with associated CVC prediction variances. We need to select one based on the set of observations $\mathbf{r}_o$ only. Recall that all predictors have normalized crossvalidation errors centred at zero and scaled to unity.

The precision of the CVC predictors $\tilde{r}_+$ measured by mean square crossvalidation error:

$$PMSE = \frac{1}{n}\sum_{i=1}^{n}[r_i - \tilde{r}_i]^2 \tag{17}$$

may vary. In the CVC predictor the scale of the normalized crossvalidation error is identical to unity but large deviations may be reduced by large prediction variances in this measure. We use PMSE as measure for precision in the CVC predictor $\tilde{r}_+$ and we prefer small values of PMSE, of course.

The precision of the CVC prediction variances $\tilde{\sigma}_+^2$ is indicated by the dependence between crossvalidation squared errors $[r_i - \tilde{r}_i]^2$ and corresponding prediction variances $\tilde{\sigma}_i^2$. Since these are variance estimates we define the mean square measure:

$$VMSE = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{[r_i - \tilde{r}_i]^2}{\tilde{\sigma}_i^2} - 1\right]^2 \tag{18}$$

Recall that normalized crossvalidation squared error in this expression is centred exactly at unity. We use VMSE as measure for precision in the CVC prediction variance $\tilde{\sigma}_+^2$ and we prefer small values of VMSE, of course.

By comparing values of PMSE and VMSE for different CVC predictors we are able to evaluate the relative quality of the predictors. Normally, the criterion PMSE is considered more important than criterion VMSE.

5.1 Empirical Studies

It is difficult to compare the various predictors analytically, hence we conduct an empirical comparision. We present two case: set of observations of yearly accumulated precipitation and a simulation study on a general Gaussian random field.

US Precipitation Study
We consider observations of 1997 accumulated precipitation in 1001 locations in an area of the US, see Figure 1. The observations is a subset of much larger data set, see Data (2014) and Johns et al (2003).

By inspecting the observations in Figure 1 we note relatively dense, uniform coverage of observations with a slight south-eastern trend in the values. The spatial correlation function is inferred under a stationary model, see Figure 2, and the following model is fitted:

$$\rho(\tau) = exp\left\{-\left[\frac{\tau}{3.5}\right]^{1.4}\right\}; \quad \tau \geq 0$$

with $\tau = \left|x^{'} - x^{''}\right|$. This spatial correlation model is used throughout the study. The localized predictors require that the number of observations in the neighborhood k is specified. We have, after a small preliminary study, chosen k=10.

We compare the five alternative CVC predictors defined Section 4 by using the evaluation criteria defined in Section 5. The results from the evaluation are displayed in Figure 3 thru 9 and Table 1.

The results from the Glob/Stat/Trad CVC predictor are displayed in Figure 3 and 4 and in Table 1. Figure 3 display the actual crossvalidation predictions and crossvalidation prediction standard deviations in the observations locations. The predictions can be compared to the actual observations in Figure 1, and no dramatic deviations are seen since the observation coverage is dense. Figure 4 display the normalized crossvalidation errors both spatially and as a histogram. Note that the histogram is centred to zero and scaled to unity since the CVC predictor is used. The locations of large errors seem to fall in the south-eastern corner where the trend effect is largest. The values of the evaluation criteria are listed in Table 1, first column ($k = 1000$). The two first lines MNE and MSNE contain the values of $\hat{\mu}_e$ and $\hat{\sigma}_e^2$ respectively, hence the empirical moments of the normalized crossvalidation errors of the non-calibrated predictor. The predictor appears as well centred but with downward biased prediction variances. The two next lines contain the values of the evaluation criteria PMSE and VMSE, for the CVC predictor. The former criteria is related to prediction precision while the latter is related to precision in the prediction variance. These criteria provide the base for comparision of the various CVC predictors.

The results from the Loc/Stat/Trad k=10 CVC predictor are displayed in Figure 5 and 6 and in Table 1. The formats are identical to the ones discussed in the previous paragraph. The crossvalidation predictions and prediction standard deviations in Figure 5 are very similar to the results for the global predic-

tor in Figure 3. The normalized crossvalidation errors in Figure 6 do deviate from the results for the global predictor in Figure 4. Note that the large errors tend to be more uniformly located in the area and that the histogram have somewhat lighter tails. From Table 1 we see that the non-calibrated predictor is well centred but with downward biased prediction variances. The evaluation criteria PMSE and VMSE of the Loc/Stat/Trad k=10 CVC predictor have values that are favorable compared to the global CVC predictor. Not much favorable for prediction but clearly favorable for prediction variance.

The results from the Loc/Stat/Shr k=10 CVC predictor are displayed in Figure 7 thru 9, and in Table 1, on similar formats as above. The predictor relies on a set of hyper-parameters that defines the prior model for localized expectation and variance. These prior models are assessed in an empirical Bayesian spirit from the complete set of observations. For the current predictor with k=10 we obtain the prior model displayed in Figure 9. The crossvalidation predictions and crossvalidation standard deviations in Figure 7 appear as very similar to the results for the other CVC predictors in Figure 3 and 5. The normalized crossvalidation errors in Figure 8 appear as similar to the ones for the traditional localized predictor in Figure 6. Note, however, that the histograms are different in the sense that the histogram of the shrinkage predictor have lighter tails than for the traditional one. This is very much in the shrinkage spirit, since extreme predictions, often caused by unstable model parameter estimates, are dampened towards the centre of the model. From Table 1, we observe that the non-calibrated predictor is well centred but with downward biased prediction variances. The evaluation criteria PMSE and VMSE for the Loc/Stat/Shr k=10 CVC predictor are both favorable to both the traditional global and localized predictors. Minor improvement on prediction precision is obtained, while the precision in the prediction variance is clearly improved.

The results from the Loc/Non-stat/Trad k=10 CVC predictor and the Loc/Non-stat/Shr k=10 CVC predictor are only summarized in Table 1. By comparing the values of the evaluation criteria PMSE and VMSE to the other predictors, we conclude that the precision in prediction is clearly poorer. Note, however, the improvement in precision for the prediction variance. These results may indicate that there is a trade-off in the precision of prediction and prediction variance.

To summarize the current study, we conclude that localized predictors can be clearly favorable to the global one. Favorable both in prediction precision and particularly in precision of prediction variance. The localized models are robust with respect to deviations from assumptions of global stationary, and this robustness causes improvements of the localized predictors. Localized stationary predictors are favorable to the localized non-stationary ones, since the precisions in prediction are clearly better. The precision in the prediction variance can, however, be improved by non-stationary predictors. In

the non-stationary models we must estimate expectation and variance in each observation location, which introduces additional uncertainity in the model. This uncertaininty is dominating the advantage of using a more general model. Among localized, stationary predictors, the shrinkage predictor is clearly favorable to the traditional one. The prediction precision is slightly better, while the precision in prediction variance is clearly favorable. For localized models we need to make bias/variance trade-offs when selecting the localization. Using a regularizer representing the global variability in the parameter estimators provide more stable estimates. It is not surprising that the effect is largest for prediction variance, since traditional variance estimators are notoriously unstable.

In the current study we have used a localization with k=10. In the extended study, see Asfaw and Omre (2014), we have evaluated the sensitivity to choise of k-value. If we reduce k considerably, to for example k=4, the localized predictor are poorer than the global one. This is probably caused by overfitting to the observations. Results for k in the range 8 to 16 are consistent with the results in the current study with localized shrinkage predictors clearly favorable. By increasing k we will of course in the limit have the localized and global predictors to coinside.

Simulation Study
We conduct a simulation study on a general Gaussian random field model $\{r(x); x \in \mathbf{D} \subset \Re^1\}$ with $\mathbf{D} = [1, ..., 200]$ discretized to grid $\mathcal{L}_{\mathbf{D}} = \{1, ..., 200\}$. The expectation and variance fields are displayed in Figure 10.a, and note the non-stationarity with inter-dependence. The known correlation function is $\rho(\tau) = exp\{-0.2\tau^{1.5}\}$ with $\tau = \left|x'' - x'\right|$. We generate one realization from this random field model and use realizations at locations $\mathcal{L}_{\mathbf{o}} = \{1, 10, ..., 190, 200\}$ as observations, hence $n = 21$.

With known model parameters, the correct predictions and prediction variances in locations $\mathcal{L}_{\mathbf{D}}$ are analytically assessable, see Figure 10.b. The predictor reproduces the observations and the prediction variances reflects the non-stationarity. We use the various CVC predictors Glob/Stat/Trad, Loc/Stat/Trad and Loc/Stat/Shr with localization k=$\pm$4 when relevant, and obtain results as in Figure 10.c-e. Note that the global predictor have regression towards the observation average and prediction variance accounting only for the observation configuration. The localized predictors are fairly similar, capturing the local variability in the observations. The shrinkage-version appears as somewhat dampened relative to the traditional one. This dampening is caused by the prior models on local expectation and variance which are assessed in an empirical Bayesian spirit. The prior model for this realization are displayed in Figure 11. We also perform prediction based on the non-stationary models but the results are not displayed in Figure 10.

To summarize, the global predictor, which corresponds to classical ordinary Kriging, appears as highly unreliable due to deviations from the stationarity assumptions, and will not be further evaluated. The localized predictors are difficult to distinguish by visual inspection.

The evaluation criteria discussed in Section 5 can be revised to characterize the deviations between the CVC predictions and prediction variances and the correct ones in Figure 10.a. We repeat 1000 realizations and average the criteria to obtain APMSC and AVMSC as evaluation criteria for prediction and prediction variance respectively, see Table 2 with $k=\pm 4$. We repeat this procedure for varying localizations k. Note that for localized, stationary predictors the shrinkage-versions are favorable to the traditional one for both criteria and all k, except one extreme case with large k. For the localized, non-stationary predictors the shrinkage-versions are uniformly favorable to the traditional one.

One typical feature is observed for localized/stationary predictors for criterion APMSC, characterizing precision in prediction, for varying localizations k. The traditional predictor makes bias/variance trade-offs, with poor performances for large k due to bias and for small k due to instability in parameter estimates. Localization $k=\pm 4$ appears as favorable. The shrinkage predictor stabilizes the variance estimates by shrinkage and tolerates smaller localizations with $k=\pm 2$. We consider the Loc/Stat/Shr $k=\pm 4$ CVC predictor to be the favorable one, since the APMSC criterion is seen as more important than the AVMSC one.

The evaluation criterion discussed above require the correct predictors to be avaliable, which is not the case in real studies. We have also computed the crossvalidation criteria discussed in Section 5 which always are available. We average over 1000 realizations to obtain APMSE and AVMSE, and the values are listed in Table 2. For these criteria the shrinkage-versions are uniformly favorable to the traditional ones for all cases. We consider the Loc/Stat/Shr $k=\pm 2$ CVC predictor to be best based on these criteria, of course.

To summarize, the crossvalidation based criteria are representative for the exact criteria based on the correct predictions. Note that the former can be computed in real studies with only one set of observations avaliable. The Loc/Stat/Shr CVC predictors are favorable to other predictors, although the localization k tends to be somewhat underestimated by crossvalidation.

In the current study we have only used one expectation and variance function. In the extended study, see Asfaw and Omre (2014), we have considered many other cases. The results are consistent with the ones presented here, and it is demonstrated that the traditional predictors are particularly sensitive to

deviations from stationarity in expectation which also influences the variance estimates of course. Lastly, recognize that the synthetic simulation case is Gaussian, no outliers and no heavy-tailed distributions are involved. In spite of this, we find that the shrinkage predictors are favorable to the traditional ones. In presence of outliers and heavy-tailed models, we expect the shrinkage predictors to perform even more favorably.

6 Conclusion

We specify two versions of localized, shrinkage CVC predictors, one based on local stationarity and one with local non-stationarity. The shrinkage is defined in an empirical Bayes setting while the crossvalidation calibration (CVC) ensures correct global scaling of the predictor variance. The introduction of spatial shrinkage predictors constitutes the new feature of the study, and they are termed localized/shrinkage kriging predictors.

The localized/shrinkage kriging predictors are compared to the traditional kriging predictors, both global and localized, in a study on real precipitation data and in a synthetic simulation study. We use two crossvalidation based criteria in the comparison. The localized/shrinkage kriging predictors are found to be clearly favorable to traditional kriging predictors on the real data set of yearly accumulated precipitation. The synthetic study is based on a Gaussian random field with spatially varying expectation and variance which make local predictors suitable. The localized/shrinkage kriging predictors appear as clearly favorable to traditional localized kriging predictors. The shrinkage predictors based on local stationarity seems to be the superior ones.

Our recommendation is to use localized, shrinkage kriging predictors, based on a local stationarity model, for spatial prediction whenever deviation from stationarity in the observations is suspected. Even for a stationary Gaussian model localized, shrinkage kriging predictors can be preferable to global ordinary kriging, if focus is on computational demands.

References

[1] Anderes, E. B. and Stein M. L.; 2011: "Local likelihood estimation for non-stationary random fields", Journal of Multivariate Analysis, Vol. 102, pp. 506-520.

[2] Asfaw, Z. G. and Omre, H.; 2014: "Localized/Shrinkage Kriging Predictors". Preprint Series in Statistics no.1, 2014, Department of Mathematical Sciences, NTNU, Trondheim, Norway.

[3] Chiles, J.P and Delfiner,P; 1999: *Geostatistics: Modeling spatial uncertainty*, Wiley, New York, 695p.

[4] Data; 2014: http://www.image.ucar.edu/GSP/Data/US.monthly.met/

[5] Efron, B. and Morris, C.; 1973: "Stein's estimation rule and its competitors-An empirical bayes approach", Journal of the American Statistical Association, Vol. 68, No. 341, pp. 117-130.

[6] Johns, C. , Nychka, D., Kittel, T. , and Daly, C.; 2003: "Infilling sparse records of spatial fields", Journal of the American Statistical Association, Vol.98, No. 464, pp. 796-806.

[7] Journel, A.G and Huijbregts, C.J; 1978: *Mining Geostatistics*, Academic Press Inc, London.

[8] Stein, M. L. ; 2002: "The Screening Effect in Kriging", The Annals of Statistics, Vol. 30, No. 1, pp. 298-323

[9] Røislien, J. and Omre, H.; 2006: "T-distributed random fields: A parametric model for heavy-tailed well-log data," Mathematical Geosciences, Vol. 38, No.7, pp. 821-849.

Appendix A

Hierarchical, Gaussian model

Consider a stationary hierarchical, Gaussian random field $\{r(x); x \in \mathbf{D} \subset \Re^m\}$ and a set of observations $\mathbf{r}_o = [r(x_1), ..., r(x_n)]$, $x_1, ..., x_n \in \mathbf{D}$.

The pdf are:

$$[\mathbf{r}_o \mid m, s^2] \sim f[\mathbf{r}_o \mid m, s^2] = Gauss_n[m\mathbf{i}_n, s^2\Omega_{oo}]$$

$$= [2\pi]^{\frac{-n}{2}} \left|s^2\Omega_{oo}\right|^{-\frac{1}{2}} exp\left\{\frac{-1}{2}\left[\mathbf{r}_o - m\mathbf{i}_n\right]^T \left[s^2\Omega_{oo}\right]^{-1}\left[\mathbf{r}_o - m\mathbf{i}_n\right]\right\}$$

$$[m \mid s^2] \sim f[m \mid s^2] = Gauss_1[\mu_m, \tau_m s^2]$$

$$= [2\pi]^{\frac{-1}{2}} [\tau_m s^2]^{-\frac{1}{2}} exp\left\{\frac{-1}{2}\left[\tau_m s^2\right]^{-1}\left[m - \mu_m\right]^2\right\}$$

$$[s^2] \sim f[s^2] = InvGam_1[\xi_s, \gamma_s]$$

$$= [\Gamma(\xi_s)]^{-1}\gamma_s^{\xi_s}[s^2]^{-[\xi_s+1]} exp\left\{-\gamma_s[s^2]^{-1}\right\}$$

Note that:

$$E[s^2] = [\xi - 1]^{-1}\gamma_s \quad ; \quad \xi_s > 1,\, \gamma_s > 0$$

$$Var[s^2] = \left[(\xi - 1)^2(\xi - 2)\right]^{-1}\gamma_s^2 \quad ; \quad \xi_s > 2,\, \gamma_s > 0$$

From the pdfs above

$$[m \mid s^2, \mathbf{r}_o] \sim f[m \mid s^2, \mathbf{r}_o] = \frac{f[\mathbf{r}_o \mid m, s^2]f[m \mid s^2]}{\int f[\mathbf{r}_o \mid m, s^2]f[m \mid s^2]dm}$$

$$= Gauss_1(\mu_{m|o}, \sigma^2_{m|o})$$

$$\mu_{m|o} = \mu_m + \tau_m \mathbf{i}_n^T \left[\tau_m \mathbf{i}_n \mathbf{i}_n^T + \Omega_{oo}\right]^{-1} [\mathbf{r}_o - \mu_m \mathbf{i}_n]$$

$$\sigma^2_{m|o} = \tau_m s^2 - \tau_m s^2 \mathbf{i}_n^T \left[\tau_m s^2 \mathbf{i}_n \mathbf{i}_n^T + s^2\Omega_{oo}\right]^{-1} [\tau_m s^2 \mathbf{i}_n]$$

$$[s^2 \mid \mathbf{r}_o] \sim f[s^2 \mid \mathbf{r}_o] = \frac{\int f[\mathbf{r}_o \mid m, s^2]f[m \mid s^2]f[s^2]dm}{\int\int f[\mathbf{r}_o \mid m, s^2]f[m \mid s^2]f[s^2]dmds^2}$$

$$= InvGam_1(\xi_{s|o}, \gamma_{s|o})$$

$$\xi_{s|o} = \xi_s + \tfrac{n}{2}$$

$$\gamma_{s|o} = \gamma_s + \tfrac{1}{2}\left[[\mathbf{r}_o - \mu_m \mathbf{i}_n]^T \left[\tau_m \mathbf{i}_n \mathbf{i}_n^T + \Omega_{oo}\right]^{-1} [\mathbf{r}_o - \mu_m \mathbf{i}_n]\right]$$

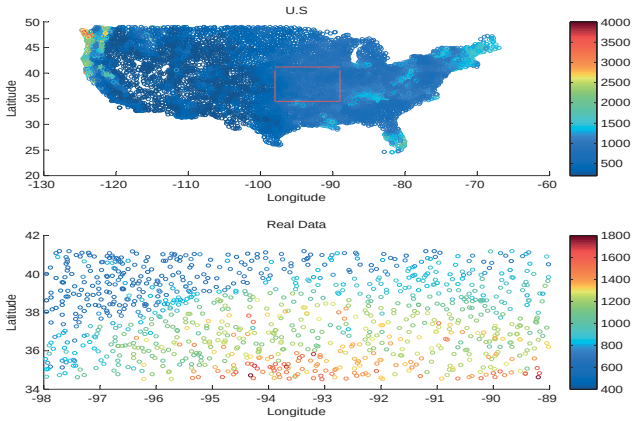

Fig. 1. The 1997 accumulated precipitation observations in the US with sub-area to be studied (top). The 1001 observations used in the study (bottom).

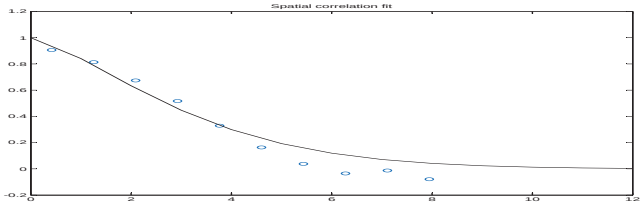

Fig. 2. Spatial correlation function used with estimated correlation values.

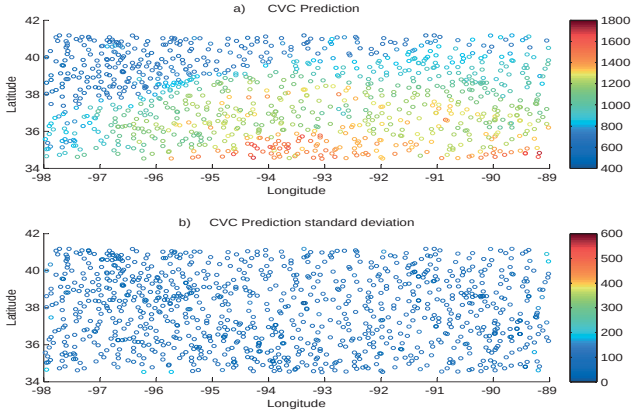

Fig. 3. Glob/Stat/Trad CVC Predictor-ordinary kriging.

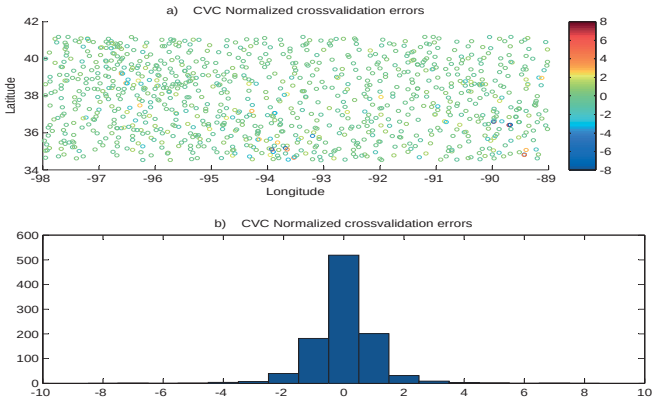

Fig. 4. Glob/Stat/Trad crossvalidation errors.

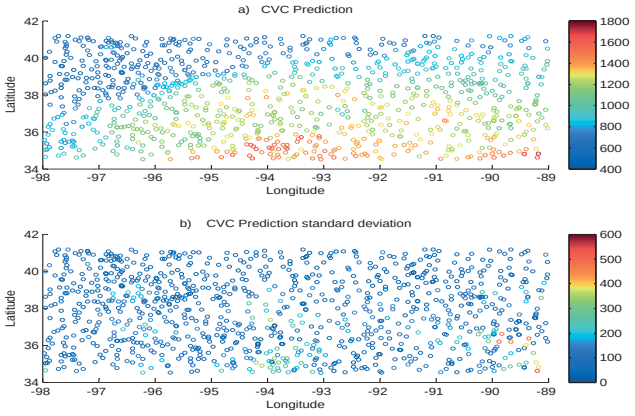

Fig. 5. Loc/Stat/Trad k=10 CVC Predictor.

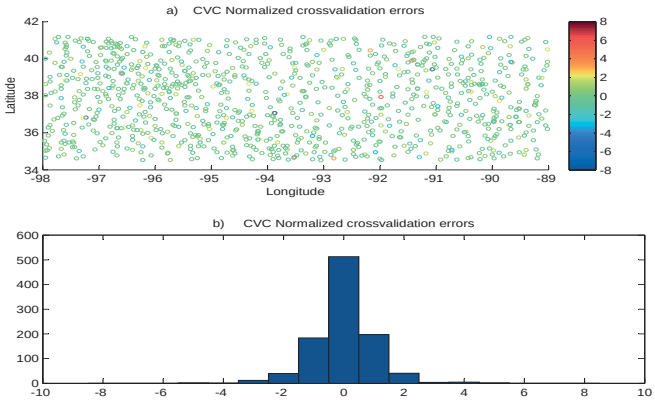

Fig. 6. Loc/Stat/Trad k=10 crossvalidation errors.

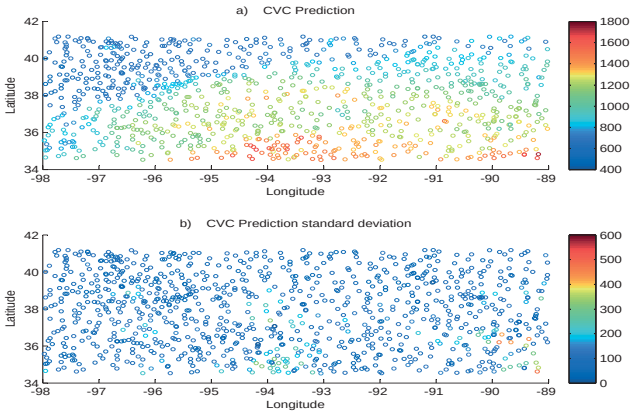

Fig. 7. Loc/Stat/Shr k=10 CVC Predictor.

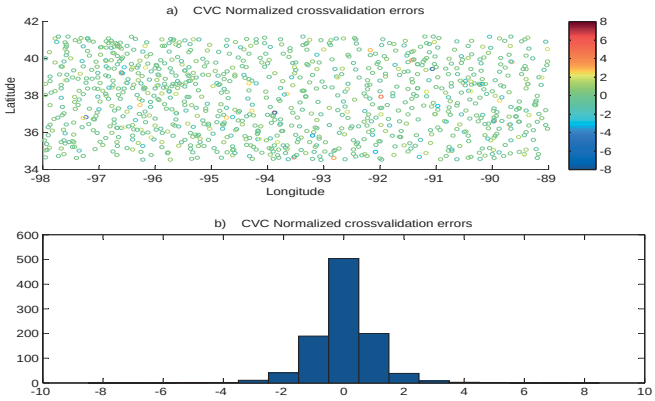

Fig. 8. Loc/Stat/Shr k=10 crossvalidation errors.

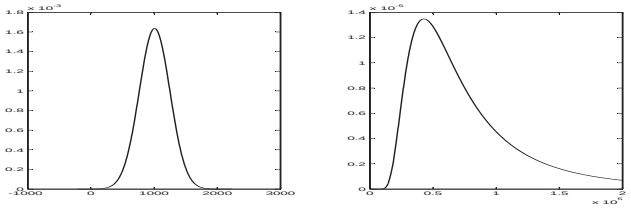

Fig. 9. US Precipitation study. Priors model for expectation and variance.

Model	Localized/Stationary			Localized/Non-stationary	
	Traditional		Shrinkage	Traditional	Shrinkage
Test D.	$k = 1000$	$k = 10$	$k = 10$	$k = 10$	$k = 10$
MNE	1.0148e-17	-5.3792e-18	5.9892e-17	1.5084e-16	-1.1579e-16
MSNE	1.5399	2.9740	3.3928	9.2487	5.1702
PMSE	$6.8758e+03$	$6.8745e+03$	$6.8654e+03$	$2.2181e+04$	$9.3555e+03$
VMSE	9.8749	5.9746	5.2027	4.0787	4.3475

Table 1
US Precipitation study. Evaluation criteria: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE)and Variance mean squared error (VMSE).

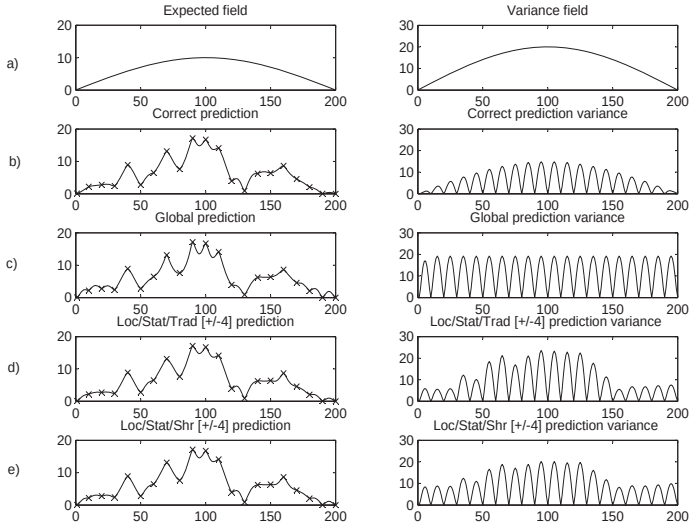

Fig. 10. General Gaussian random field. Expectation and variance field (a), Predictions and prediction variances for one realization (b-e).

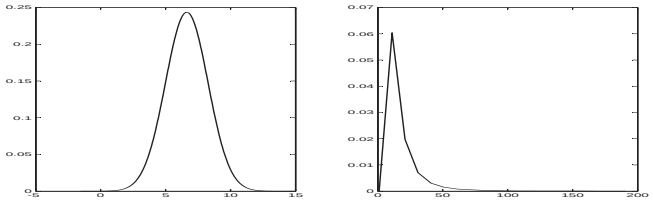

Fig. 11. General Gaussian random field. Prior model for expectation and variance for one realization.

Model	Localized/Stationary						Localized/Non-stationary					
	Traditional			Shrinkage			Traditional			Shrinkage		
Test D.	±6	±4	±2	±6	±4	±2	±6	±4	±2	±6	±4	±2
APMSC	0.3665	0.3357	1.0100	0.3968	0.2838	0.2563	0.4815	0.6669	2.1658	0.4259	0.4305	0.6264
AVMSC	0.5287	1.0939	159.3760	0.3927	0.5944	1.2491	0.5108	1.0369	138.6639	0.3860	0.5723	1.2327
APMSE	18.2536	17.8311	19.8851	16.9502	15.6486	13.8070	18.6699	18.4643	23.7843	17.0732	15.8705	14.1365
AVMSE	1.8936	2.0277	2.9888	1.7271	1.7306	1.4011	1.8908	2.0086	2.4491	1.7277	1.7187	1.3483

Table 2
General Gaussian random field. Evaluation criteria: Average prediction mean squared correct (APMSC), Average variance mean squared correct (AVMSC), Average prediction mean squared error (APMSE) and Average variance mean squared error (AVMSE).