Titles and abstracts

Smögen workshop 2010

Magne Aldrin (with Ragnar Bang Huseby): Disease dispersal in marine fish aquaqulture - A statistical model with unobserved infection

Abstract: Infectious diseases constitute a constant threat to the Norwegian fish farming industry with major economic implications, in addition to being a problem for fish welfare and the environment. We have developed a statistical model to describe how an infectious disease is transmitted between fish farms. Such models may be used to investigate the effect of various measures to reduce the risk of infection. Of special interest would be simulations of various scenarios with regard to geographic location of fish farms, as well as effects of vaccination. The work is performed in close cooperation with the National Veterinary Institute.

For each disease outbreak, we know the time when the outbreak was recorded, but the actual time of infection is unknown. In addition, a farm may be infected and then become infectious even if no outbreak has been recorded at that farm. We model this as two separate processes, a latent infection process and an observable outbreak process conditioned on occurrence of infection. The infection rate is modeled as a composition of contributions of transmission from four possible sources: adjacent infected farms, infected farms within shared local contact networks, previous infected fish cohort resident at the same farm, and other sources. The sea distance between pairs of farms, which takes into account the coastal geography, is an important explanatory variable in the model. Another important factor is the time sequence of the observed outbreaks.

In order to estimate the model by accounting for the unobserved infection we use a Bayesian data augmentation approach utilizing MCMC techniques. We have investigated transmission of three diseases by using data from Norway from 2003 to 2009, with number of observed disease outbreaks varying from about 60 for one disease to several hundreds for the two other diseases.

David Bolin: Spatial models generated by nested SPDEs

Abstract: A new class of stochastic field models is constructed using nested stochastic partial differential equations (SPDEs). The model class is computationally efficient, applicable to data on general smooth manifolds, and includes both the Gaussian Matérn fields and a wide family of fields with oscillating covariance functions. Non-stationary extensions are obtained by spatially varying the parameters in the SPDEs, and non-Gaussian extensions are obtained by driving the SPDE with non-Gaussian noise. Finally, the model class is used to estimate daily ozone maps from a large data set of spatially irregular global total column ozone measurements.

Ottmar Cronie: Estimation in a spatio-temporal marked point process

Abstract: The Renshaw-Särkkä spatio-temporal growth-interaction model (RS-model) has recently been studied in (Särkkä and Renshaw, 2006). The basis of this spatio-temporal marked point process is an immigration-death process which governs the number of arrivals of individuals (occurring according to a Poisson process) whose locations are uniformly distributed in the region of interest. This process also governs the individuals' deaths (individuals are assigned iid exponential life-times). Upon arrival an individual is assigned a mark which changes with time according to a deterministic growth equation which consists of two parts; an individual growth term and a spatial interaction term. Since spatial interaction takes place, naturally, individuals may also suffer death through competition.

In this talk we will discuss the improvements made in the maximum likelihood (ML) estimation of the immigration-death process parameters (Cronie and Yu, 2010). Thus far the growth and interaction parameters have been estimated using the method of least squares. We here will discuss the RS-model and its (possible) ML estimation when it is altered in such a way that the deterministic growth equation which governs the marks' growth is replaced by its stochastic differential equation analogue. Possibly, if there is time, we will briefly mention some edge correction methods developed for the RS-model (Cronie, 2010).

Keywords: Immigration-death process, Maximum likelihood estimation, Stochastic differential equation, Spatio-temporal marked point process.

References

Cronie, O. (2010). Some edge correction methods for marked spatio-temporal point process models. *Preprint series - Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg*, 2010:9.

Cronie, O., Yu, J. (2010). Maximum likelihood estimation in a discretely observed immigration-death process. *Research report: Swedish University of Agricultural Sciences - Centre of Biostochastics*, 2010:01.

Särkkä, A., Renshaw, E. (2006). The analysis of marked point patterns evolving through space and time. *Computational Statistics & Data Analysis 51*, 1698–1718.

Arnoldo Frigessi: A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims

Abstract: Climate change will affect the insurance industry. We develop a Bayesian hierarchical statistical approach to explain and predict insurance losses due to weather events at a local geographical scale. The number of weather-related insurance claims is modelled combining generalized linear models with spatially smoothed variable selection. Using Gibbs sampling and reversible jump MCMC, the model is fitted on daily weather and insurance data from each of the 319 municipalities of southern and central Norway for the period 1997-2006. Out-of-sample predictions from the model are very good. Our results show interesting regional patterns in the impact of different weather covariates. In addition to being useful for insurance pricing, our model can be used for short-term predictions based on weather forecasts and long-term predictions based on downscaled climate models.

Philip Gerlee: The effect of space on the war of attrition

Abstract: In the game-theoretic model "war of attrition" the strategy of a player consists of a single real number which corresponds to how long the agent is willing to wait. In the two-contestant game, the winner is the player with the longest waiting time (WT). The winner collects a unit award, while the loser receives a consolation prize of $k < 1$. It has previously been shown that a population of agents playing this game (with an implicit time-cost, i.e. players with long waiting times play fewer games), can exhibit both stationary and oscillatory distribution of WT depending on the value of k. We consider the effect of space on the dynamics, and analyse a coarse-grained version of the model, which is formally equivalent to a stochastic cellular automata. The results show that the system exhibits a phase-transition in the WT-distribution, where for small k the long WTs dominate, while for large k short WTs are most abundant. However, when the spatial component is removed and the model is well-stirred this effect disappears and co-existence is prominent.

Peter Guttorp: Space-time trends in temperature extremes in south central Sweden

Abstract: We look at data from some of the synoptic sites from SMHI. The aim is to look for time trends in surface temperature minima, which from climate model simulations are expected to be stronger than those in average temperatures. We fit GEV models to the different sites using maximum likelihood methods, study the effect of altitude, estimate time trends, and investigate what common parameter models are possible. Doing spatial interpolation of the model parameters, we estimate the distribution at a site not used in the model fitting, and compare the distribution of observed and fitted data. Finally we demonstrate how wavelet tools can be used to estimate nonlinear time trends with attendant simultaneous confidence intervals.

Gudmund Hermansen: Model robust inference and model selection issues for stationary Gaussian time series.

Abstract: We consider stationary Gaussian time series where the dependency structure is potential misclassified, meaning that the true dependency is determined by some unknown spectral density g, but where we are working within a parametric class $\mathcal{F}_\theta$ that does not necessarily include the true spectrum. This leads to a least false parameter estimation perspective, where we show that the maximum likelihood estimator converges in true P_g-probability to a minimizer θ_0 of a distance function $d(g, f_\theta)$. This distance function is naturally motivated and also leads to model selection criteria.

We also obtain the limit distribution for the normalized estimator and show that it converges to a Gaussian distribution with a so called "sandwich" covariance matrix. The results are derived under the full likelihood and also using the Whittle approximation, where both estimates are shown to share the same limit properties. The results are illustrated through analytical and numerical examples.

Jenny Jonasson: Statistical modeling of scanning transmission electron microscopy images of crystal interfaces with atomic resolution

Abstract: In this talk some ideas will be presented for modeling High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM) images of crystals. In a crystal the atomic structure is a lattice which can be observed by the electron microscope. The main aim is to study the interface between two crystalline materials with similar structure. We want to describe the position of the interface and its width, i.e. the degree of intermixing between the materials. Another question is if the structure deviates from a regular lattice in the interface. In the end we hope to build a model for the entire structure and estimate parameters from the images. This is work in progress and some preliminary results will be presented.

Wengang Mao: A comparitive study of extreme value estimation and application on ship structures

Abstract: In this paper two methods to predict the extreme responses are compared. The first one uses quantiles of the long term distribution of response maximums while the second estimates upcrossings of high level by the response during the design life. It is shown that the methods give the same predictions if the short term distribution of local maximums is estimated by means of the narrow band approximation. The focus is on the zero mean non-Gaussian responses which are characterized by the standard deviation, skewness, kurtosis and zero up-crossing frequency which varies with changing sea conditions. A relation between the response parameters and the significant wave heights is studied. Crossings of high levels are computed by means of the transformed Gaussian process. The Winterstein's transformation, which is defined by the stress parameters, is employed. The full-scale measurements of a 2800 TEU container ship during the first six months of 2008 are used to validate the methods. Keywords: Extreme response, up-crossing, non-Gaussian response, significant wave heights, skewness, kurtosis

Jacques de Maré (joint with Sara Lorén): Maintenance for reliability

Abstract: The optimal replacement problem for components with stochastic lives has an appealing solution based on the TTT-transform. The issue is revisited for components which are regularly inspected and where statistical uncertainties are taken into account by means of the method of predicted profile likelihood. The ideas are applied on crack growth data on a low pressure nozzle in a jet engine. It turns out that the standard method is not directly applicable and that the effect of uncertainties on the replacement times is not easy to predict.
Keywords: Condition based maintenance, Periodic inspection, Stochastic component, Predictive profile likelihood, Replacement

Mari Myllymäki: Second-order spatial analysis of nerve fibers

Abstract: The aim of our project is to introduce new statistical methods for analysing and modelling the spatial distribution of epidermal nerve fiber patterns in the epidermis. Preliminary neurologic studies indicate qualitative differences in the spatial patterns of nerve fibers based on pathophysiologic conditions in the subjects. Of particular interest is the progress of diabetic neuropathy. It appears that the spatial distribution of nerve fibers becomes more "clustered" as neuropathy advances. Our question of interest is whether or not the nerve fibers are significantly more "clustered" in diabetic subjects than in normal subjects. We analyse some samples from thighs of non-diseased subjects and subjects with mild, moderate or severe diabetes. The nerve patterns of the different groups are compared by using (functional) second-order summary statistics for spatial point and fiber processes. The spatial point pattern under consideration is the set of locations where the nerves enter the epidermis from the dermis. Further, we discuss analysis of the larger data set consisting of replicates of healthy subjects.

Alessandro Ottavi: Bayesian spatial smoothing over complex domains

Abstract: Spatial smoothing methods often assume the region of interest is $\Re^2$ but in practice we might have a river, lake or a mountain, which, at least "physically", prevents smoothing over the river, lake or mountain. Such geometrical "constraints" is not easy to deal with properly.

In this work, we try to not see these physical constraints as "annoying constraints" but rather reinterpret physical constraints into geometrical properties and then include the geometry into the model. To do this, we have to leave $\Re^2$ and define smoothing methods on a manifold in $\Re^3$. The essential idea, is to make use of the ideas by Lindgren and Rue (2007), which showed how Matérn fields on a manifold could be constructed using the corresponding stochastic partial differential equation (SPDE). Moreover, they also achieved a representation with Markov properties which ensure fast computations. We show how their approach could be extended to deal with spatial smoothing over complex domains which automatically respect and adjust for physical constraints.

This is joint work the H.Rue.

Krzysztof Podgorski: Spectral representation and ergodic theorem for stationary process driven by asymmetric Laplace distributions

Abstract: Laplace motion is defined as the Levy motion arising from an asymmetric Laplace distribution. Stationary processes can be obtained by integration of time dependent functions with respect to such a motion. Although these are second order processes having also higher moments, their properties differ significantly from the ones for Gaussian processes. For example, the harmonizable processes are not ergodic while the spectral representation of ergodic ones involves stochastic measures with dependent increments. We investigate both the (random) time average limit in the non-ergodic case and spectral representation in the ergodic case. Potential consequences for applications are discussed as well.

Jonas Ringsberg: Survivability analysis of a struck ship with damage opening - influence from model and material properties uncertainties

Abstract: The conditions for damage stability and survivability of a ship struck by collision in arbitrary sea-state are, from a structural point of view, determined by the size and shape of the damage opening in its side shell. Today, there exist no modelling guidelines for engineers how such calculations should be carried out to ensure that ships are safely designed with respect to crashworthiness. The current presentation demonstrates the importance of incorporating systematic parameter sensitivity analyses/uncertainty analyses in ship design. The work presented is part of a European network which will propose modelling guidelines to classification societies who have the main responsibility for ships' safety during their design phase.

Explicit finite element analyses (FEA) are presented of a ship-to-ship collision scenario where the damage opening of a struck ship is calculated for a selection of damage models and material properties determined by testing. In the analyses, model uncertainty is introduced as a possible (user-related) insecurity in the selection of the most appropriate damage model for the structure analysis. The uncertainty in material properties is introduced in the in the constitutive material model description and the material parameters used in the damage models. The size and shape of the damage openings predicted by the FEA are used in following dynamic damage stability analyses in which the struck ship is subjected to wave motions in arbitrary sea-state and flooding into the damage opening.

The time to capsize, $T_{capsize}$, is used as a parameter to define "survivability". The probability of survival is calculated for a number of sea-state conditions. The influence on $T_{capsize}$ from model and material properties uncertainties, and the natural variation in sea-state conditions, is presented in "capsize band" diagrams. These diagrams show if a ship will survive the collision scenarios defined by IMO (International Maritime Organisation), or if its structural design must be changed to meet the requirements prescribed by classification societies.

Mats Rudemo: Maximum likelihood estimation of diffusion coefficient distributions from particle tracking

Abstract: For analysis of data from experiments with diffusing particles from heterogeneous populations log-normal distributions are typically used for the distribution of diffusion coefficients. We suggest using inverse gamma distributions instead. This give closed form analytical expressions for the likelihood functions, which greatly simplifies the analysis. Models with one inverse gamma distributions and models with mixtures of such distributions are studied. In particle tracking we often have observations only in a limited focal depth which causes a sampling bias as fast particles are sampled more often. This bias is studied in detail in a separate paper and here we only use a simplified sampling bias correction. However, this simplified bias correction is particularly suitable together with inverse gamma distributions as it results in new inverse gamma distributions with changed parameters. Several empirical single particle tracking datasets with nominal particle size of 100 nm and 200 nm are studied.

Igor Rychlik:

Abstract:

Magnus Röding: Statistical methods for analysis of single particle tracking experiments

Abstract: Tracking single particles using different microscopic techniques is becoming increasingly important in a wide field of research within the nano- and biosciences. We present some initial results on e.g. estimation of particle concentration from 2D microscopic video sequences.

Aila Särkkä: Models for the structure of concrete

Abstract: Granular, sintered and multiphase materials are commonly used for example in building industry. We focus on materials which consist of spherical objects (pores or grains) embedded in a matrix material. Macroscopic properties of such materials, e.g. mechanical strength, fracture behaviour or percolation, are heavily influenced by their microstructure. Three-dimensional computer tomographic images combined with methods from image processing and spatial statistics yield a powerful source of information on microstructural characteristics such as the size distribution of the spheres or their spatial arrangement. Based on this information, models from stochastic geometry can be fitted to the microstructure of the observed material. Changing the model parameters, "virtual" materials with a modified microstructure could then be generated. The simulation of physical properties in these virtual microstructures allows one to investigate relationships between certain geometrical characteristics and macroscopic properties of the material.

We will discuss modelling of materials consisting of spherical objects by studying a sample of refractory concrete. The material consists of corundum grains and air pores, both of nearly spherical shape, embedded in a solid matrix. We will start by including only the corundum grains, not the air pores, in the analysis. Several model alternatives for concrete are considered in the literature: a marked Gibbs point process model, a modified random sequential adsorption model, a hard sphere packing generated by the force-biased algorithm, and a random-shift model. The last model is simulated by randomly and independently shifting the spheres in a hard sphere system avoiding overlaps.

Here, we introduce another model alternative and investigate its capability to generate patterns comparable to those observed in concrete structures. The models are based on immigration growth interaction processes introduced by Renshaw and Särkkä, where individuals arrive on the study area according to a Poisson process, and then, either die naturally with some rate or grow according to a deterministic growth model. By allowing movement of the particles, we are able to generate patterns with much higher packing density than the ones generated by the original growth interaction process. We will discuss the choice of the growth and interaction functions needed for the immigration growth interaction process as well as estimation of the parameters of the model.

Geir Storvik: Plague in Kazakhstan - A case study

Abstract: Plague, caused by the bacterium Yersina pestis, is known for its impact on human history through a series of pandemics. The primary reservoir of the bacterium, and the origin of the pandemics, is thought to be populations of wild rodents in Central Asia. Transmission between individuals occur through flea bites. The ecology and epidemiology of plague is complex and not fully understood.

One of the very few datasets suitable for detailed analysis of the dynamics of plague is the dataset for the Pre-Balkhash plague system in Kazakhstan, where the main host and vectors of plague are great gerbils and their fleas. This area, which has been monitored since 1949, has a unique combination of a high spatial resolution and a long study period.

In this talk I will discuss the use of spatio-temporal modeling for analysis of the Kazakhstan data. The main goal is to estimate the effects of population density of the host as well as climatic factors on plague. Our modeling strategy will be through dynamic models describing temporal changes in a spatial complex.

Prevalence of a disease is usually modelled on the log or logistic scale. A challenge here is that plague can disappear for a period of time before re-emerging. We include this possibility in the model by introducing a binary variable indicating whether plague is present or not. Presence of plague is further categorised into three different types:

- /Persistence/ corresponding to presence of plague in the same site for the previous period,
- /Spread/, corresponding to presence of plague in neighbor sites in the previous period and
- /Invasion/ in which case the presence of plague can not be related to previous presence in the neighborhood of the site.

As typical for spatio-temporal modelling, computational challenges are huge. We perform inference through Markov Chain Monte Carlo utilising sparsity in precision matrices involved. By modelling spatial correlations in the binary variables indirectly through spatial random effects, also the binary variables can be efficiently sampled.

This work is part of a project running at the Centre of Ecological and Evolutionary Synthesis (CEES) at the Department of Biology, University of Oslo and involves Nils Chr. Stenseth, Lise Heier, Hildegunn Viljugrein, Bård Øyvind Kvaal as well as many international collaborators.

Thomas Svensson: Using prior knowledge without Bayes

Abstract: In engineering there is a strong need for managing experiences and experimental results by means of parameterized properties. Values from experience, company standard or literature should be continuously improved based on new experiences and tests, but rational methods to do it are usually not available. The immediate solution that comes into a statisticians mind is Bayesian updating, but this technique may be too difficult to apply without experiences. I will here present an alternative for linear regression cases. It was presented in the MMR conference in Glasgow 2007 by Vasily Krivtsov from the Ford company and can be summarized as follows: Regard prior knowledge about a parameter in linear regression model as an extra observation. Estimate the uncertainty of this knowledge as a fraction of the random error variance in the other observations and use weighted regression to estimate the parameters. The technique will here be compared to the Bayesian method in an example from fatigue, where we assume prior knowledge about the damage exponent in Wöhler curve estimation.

Sofia Tapani: Do neighbouring cells have a common mother-cell ancestor in early embryos?

Abstract: Before differentiation cells can have many different fates in the future embryo, as can be displayed in a so called fate map. How much mixing takes place before gastrulation? A new statistical method is presented aiming at the problem of "cell mixing" in the early embryo. This method can aid in answering fundamental questions of cell polarization, omnipotence, and migration at development.

Jonas Wallin: Chaos expansions of non-linear response to an input at its level crossing

Abstract: The Slepian model describes the behaviour of a stochastic process at its level crossings. We study what happens if the Slepian model enters as an input to a non-linear system. While for a linear system the response is explicitly known, this more general situation has no explicit and exact solution. The following three-step approach of handling a non-linear system are investigated for the van der Pool and Duffie oscilators. First, we expand and approximate the Slepian model input using orthogonal Karhunnen-Loeve expansion with independent random Gaussian coefficients. Then we follow Galerkin's method and require that the error of an approximate solution to be orthogonal to the low order polynomials in the basis of random polynomials. This leads to a set of deterministic equations for the coefficients of the approximation corresponding to low order polynomials of Gaussian variables. Finally, we solve for the non-random coefficients and by using them along with the corresponding polynomials in Gaussian random variables we study the approximated stochastic response. We discuss several issues that need to be addressed to facilitate such an approach. 1) The choice of initial conditions, the statistical distribution for initial values and its effect on the response; 2) How long prior to the occurrence of a level crossing the non-linear equations have to be run so that initial values and the assumed level crossing are not affecting each other. 3) How the system behaves depending on the height of the level. All these issues are supported by simulation studies.

keyword: Karhunen-Loéve, Galerkin method, nonlinear pde, Slepian model, polynomial approximation, random intial condition, Duffie oscillator, Van der Pool oscillator.

Jorg Wegener: Stochastic fields with embedded shallow water dynamics

Abstract: The deterministic shallow water equations (SWEs) are frequently used as a toy model to verify new techniques supposed to enter the milieu of numerical weather prediction. To combine stochastic and de- terministic modeling, we propose a random field whose dynamics are governed by the shallow water formalism. The dynamics are intro- duced by embedding deterministic velocities into a stochastic tempo- spatial Gaussian model. In this way, a dynamically inactive stochastic field with given spatial and temporal covariance structure gains dy- namics that in general follow a deterministic pattern. The observed motion can be interpreted as movement of a stochastically distorted shallow water flow.

The seminar will be a journey through the areas of fluid dynamics, weather prediction and stochastic modeling.

Jun Yu: Wavelet based noise reduction and parameter estimation in magnetic resonance signals

Abstract: The project deals with development of biostochastic models for analysis of spatial-temporal signals with applications in magnetic resonance. We developed a noise reduction algorithm for 4D MRI signals, based the wavelet transform and Gaussian scale mixtures. We also derived the maximum likelihood estimators for the parameters of a modified Rice distribution, which arises from the MR signals. Asymptotic properties of the estimators were obtained and an ECM-algorithm for simultaneous estimation was developed.