# Titles and abstracts

Smögen workshop 2004

**Clive Anderson**: The influence of shape on the estimation of the sizes of large inclusions in metals

**Jenny Andersson**: Inclusions in cast iron modelled by non-overlapping spheres

**Stefano Beretta**: Application of a weakest link model for fatigue strength

**Anders Ekberg**: Inspection of railway infrastructure by image analysis
**Abstract**: Railways are expensive structures that are designed to be functional for a long period of time. To this end, the infrastructure needs to be inspected so that defects and potential hazards that occur over time will be detected and may be measured. This is especially important since the traffic operating a railway is relying on the infrastructure to be fully functional (at emergency braking, the X2000 high speed train will need 1.75 km to go from 200 km/h to full stop). However, the inspection of the infrastructure is complicated by the length of the tracks. There are about 11 500 km of railway tracks in Sweden (more than 1 million km world-wide). The inspection today is mainly made by foot and each section of the track is inspected up to six times per year. The inspection is facilitated by checklists and a computerised report system. Another option would be to mount a video-system on a train and identify maintenance needs and potential hazards from the video-image. This would permit more frequent inspections (in particular if the identification is automated) and could aid the in-field inspections. This introduction will focus on presenting the practical problem as a basis for a discussion of possibilities using image analysis. The presentation will give examples of what can (and is today) detected by an image analysis system and problems that need to be overcome. It is hoped that the presentation could initiate a discussion on which levels of accuracy that may be expected, accuracy vs cost, possibilities of connecting an image analysis system to other systems (e.g. for measuring track forces).

**Claus Ekstrøm**: Improved spot shape modeling

**Pavel Grabarnik**: Improving the maximum pseudo-likelihood estimator for spatial point processes
**Abstract**: Parameter estimation for spatial processes based on the likelihood is difficult because the normalising constant in the likelihood is intractable except of a few simple models.

Among alternatives proposed to overcome the difficulties the maximum pseudo-likelihood (MPL) method is competitive approach due to its computational ease. However, the MPL estimator looses much relative efficiency and has a substantial bias in case of strong interactions. Hence there is a need to look for estimators that still retain the computational simplicity but all that perform better. Baddeley (2000) proposed a general method of constructing estimators for parametric inference in a case when a probability model can be defined as the invariant distribution of an auxiliary Markov process. These estimators were called time-invariance estimators.

The performance and optimality of the time-invariance estimators is an open question. We investigate the performance empirically by simulation experiments and compare the quality of some time-invariance estimators for various models. We propose a modification of the pseudo-likelihood function and demonstrate that the quality of the estimator based on the new function is better than the standard maximum pseudolikelihood estimator, whereas computing does not take much extra calculations. We applied the new estimator to an analysis of the real data for which the usual MPL method gives unsatisfactory results.

**Gilles Guillot**: A spatial statistical model for landscape genetics
**Abstract**: Landscape genetics is a new discipline which aims to provide information on how landscape and environmental features influence population genetic structure. The first key step of landscape genetics is the spatial detection and location of genetic discontinuities between populations. However, efficient methods to achieve this task are lacking. In this talk, we first clarify what is conceptually involved in the spatial modeling of genetic data. Then we describe a Bayesian model implemented in a Markov Chain Monte Carlo scheme which allows inference of the location of such genetic discontinuities from individual geo-referenced multi-locus genotypes, without a priori knowledge on populational units and limits. In this method, the global set of sampled individuals is modeled as a spatial mixture of panmictic populations, and the spatial organization of populations is modeled through the colored Voronoi tessellation. Besides spatially locating genetic discontinuities, the method quantifies the amount of spatial dependence in the dataset, estimates the number of populations in the studied area, assigns individuals to their population of origin, detects individual migrants between populations, while taking into account uncertainty on the location of sampled individuals. The performance of the method is evaluated through the analysis of simulated datasets. Results show good performances for standard datasets (e.g. 100 individual genotyped at ten loci with ten alleles per locus), with high but also low levels of population differentiation (e.g. $F_{\mathrm{ST}} < 0.05$). The method is then applied to a set of eighty-eight individuals of wolverines (*Gulo gulo*) sampled in north-western

United States and genotyped at ten microsatellites.

**Hugo Hammer**: Some new estimators for binary Markov random field
**Abstract**: Parameterestimation in Markov random fields are challenging. It is possible to calculate the maximum likelihood (ML) estimator. Unfortunately the ML estimator preserve only local properties in the data. It will be suggested another way to do parameter estimation where we have more control of the properties preserved by the estimated parameters.

**Pär Johannesson**: Fatigue life prediction based on variable amplitude tests

**Elena Kabo**: Image analysis of defects of railway wheels: a challenge for mathematicians
**Abstract**: Material defects in railway wheels is a problem that relates to the reliability and safety of railway operation. Defects in the wheel material naturally develop in the production process. These defects are normally not detectable under railway operation until they have grown to a size that endanger safe operations. Instead these defects have to be found by ultrasonic inspections and similar methods. Surface defects on the other hand arise during operations. These defects may be divided into different categories, such as wear, rolling contact fatigue cracks, thermal cracks, martensite formation, wheel flats, indentations, etc. It is often hard to distinguish between these different categories by visual inspection. This is even more the case since the wheels are inspected (and photographed) in workshops where lightning conditions are poor. Further, the wheel surfaces are often worn/grimy and there are often a mix of different defects. Consequently, it is hard to categorise the defects and maintain a strict classification of damages. Thus, wheels are often unnecessarily reprofiled, which naturally costs money and may introduce operational disturbancies. This is the reason why image analysis could be a useful tool in aiding the workshop staff. The complications described above make it a challenge for statisticians to identify and classify material surface defects such as cracks, martensite formations, and microstructural changes. The question on how image quality increases analysis accuracy can also be of interest. This introduction will focus on presenting the practical problem as a basis for a discussion of possibilities. Different types of defects will be shown and their features will be discussed.

**Magnus Karlsson**: Modelling of Lateral Loads for Fatigue Life Calculations
**Abstract**: A method for predicting fatigue life for vehicle components sensitive to lateral loads is presented. Most methods used at present are directly trying to describe the loads from service measurements. The many

different aspects of the loads make this a complicated task. Alternatively it is possible to describe the different aspects causing the loads separated from each other and thereby gain information and understanding about the underlying effects that sum up to the loads. The main load sources considered in this work are influences of the road, the driving style, and the vehicle dynamics.

The lateral loads that a component will experience will primarily be determined by the lateral acceleration. The lateral acceleration can be split into two parts, one part decided by the curvature of the road and the speed, corresponding to the centripetal acceleration and another part which is a combination of the effects of the specific driver's behaviour when driving through a curve as well as of the road vibration input and the surrounding traffic.

From road construction requirements it is reasonable to see the curvature of each curve as a trapezoid. From a measurement of the curvature, trapezoids can be estimated corresponding to each curve. These curvature estimates are fitted to a statistical distribution. A regression model for the relationship between the curvature and the speed at which the driver chooses to drive through the curve is then worked out. The curvature and the speed determine the centripetal acceleration. The part of the lateral acceleration that is due to the specific driver's behaviour, the road vibration input and the surrounding traffic can be modelled as a random process. The lateral acceleration can therefore be seen as the centripetal acceleration with a random noise.

The cumulative fatigue damage caused by the lateral acceleration corresponds to the two parts mentioned above. The damage is then approximated with the damage for the centripetal acceleration with a correction term added for the maximum value of the random process over each curve, and the damage caused by the random process. By a discretisation of the centripetal acceleration it is shown that it is possible to approximate the load as a Markov chain, for which the exact Rain flow count matrix can be calculated. Thus it is also possible to calculate the expected damage for that part. The random noise is shown to be a transformed gaussian process, for which an upper limit of the cumulative fatigue damage can be found. Measurements made on a Volvo Truck in Brazil are used to study the model.

**Mats Kvarnström**:

**Claudia Lautensack**: Random tessellations for modelling of sintered structures

**Abstract**: For various reasons material scientists are interested in models

for the geometry of sintered structures. In the development of a model allowing the treatment of arbitrary stages of the sinter process you are confronted with several questions. The first task is the need for a description of the locations of the sinter particles which can be used to create a model for the initial state. Random dense packings as well as marked (Gibbs) point processes might leed to an appropriate solution of this problem. Laguerre (or power) tessellations are considered an adequate model for the final state of the sinter process. So the second part of the talk concentrates on several questions arising in the mathematical analysis of random Laguerre tessellations.

**Traci Leong and Lance Waller**: Estimating the pair-correlation function from images of epidermal nerve fibers.
**Abstract**: The previous speakers present results exploring second-order summaries of spatial patterns exhibited in images of epidermal nerve fibers in the skin. We contrast cumulative measures of spatial pattern (K and L functions) with a scale-specific measure of dependence, the pair-correlation function (g function), and apply different approaches to newly collected data. The new data are in a different spatial format, and are relatively sparse requiring additional thought in estimation of the g function. We present some exploratory results and outline additional inferential needs.

**Finn Lindgren**: Classification of fluorescence spectra from tumour images acquired in vivo

**Sara Lorén**: Estimating inclusion size distribution
**Abstract**:An important design property of steel is the fatigue limit, i.e. the load level where a specimen has infinite life. For the material it can also be defined an endurance limit, which is the stress level at which the specimen has a certain life. It will not break before $n_o$ cycles. Obviously it is of interest to estimate these limits. The maximum inclusion size in clean steels influences the fatigue limit behaviour. Therefore it is also interesting to estimate the distribution of the inclusion size which causes the failure. Here the results from staircase tests are used to estimate the endurance limit and the inclusion size distribution.

The staircase test is combined with the $\sqrt{area_{max}}$-model to estimate the distribution of the inclusion size where the failure occurs. This can be done by using stress levels with or without inclusion measurements.

By combining the $\sqrt{area_{max}}$-model and the assumption that the inclusion size above a fixed threshold follows a generalized Pareto distribution, the stresses from the staircase test can be used to estimate the probability that a component will fail. Under these assumptions the probability that the

component will fail under a rotating bending test can be estimated using data from a uniaxial test on a different component and vice versa.

**Marianne Månsson**: Dead leaves, lilyponds and other models for non-overlapping grains
**Abstract**: In a random model of non-overlapping grains the volume fraction is the proportion of space covered by the grains. We consider the volume fraction of some examples of such models, for instance the intact grains of Matheron's "dead leaves" model and the lilypond model. The grains are supposed to have a fixed, convex shape, while the sizes and orientations may be random. The focus is on how the shape of the grains affects the volume fraction.

**Jacques de Maré**: The Wicksell problem for spheres whose sections have Gumbel distributed radii

**Tommy Norberg**: Restoration and simulation of partially observed discrete geological structures

**Viktor Olsbo and Aila Särkkä**: How is the spatial pattern of epidermal nerve fibers affected by diabetic neuropathy?
**Abstract**: Recent breakthroughs in imaging of skin tissue reveal new details on the distribution of nerve fibers 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. We consider the set of nerve trunks to be a realization of a marked spatial point process and use second-order properties of the observed data to describe the degree and scale of clustering between nerve trunks. We compare patterns observed in images taken from the thighs of one normal (disease-free) individual, and two images each taken from the thighs of subjects with mild, moderate, and severe diabetes and report measurable differences in the spatial patterns of nerve trunks associated with disease status.

**Antti Penttinen**: External marking of point patterns
**Abstract**: External or geostatistical marking of a point pattern means that each object of the point pattern is provided with a mark, derived from a random set or from a continuous-parameter random field. The point pattern and the random field or random set are assumed to be independent.

External marking is considered in literature for many purposes. Lotwick (1984) gives an example of an unmarked point process with dependence between the points but having the same second order structure as the Poisson process. The construction relies on externality in marking. Mase (1996) models observed gauge measurements for rainfalls by an independent random field. Wälder and Stoyan (1997, 2000) deal with the problem of mark-variogram and its relation to the geostatistical variogram of a random field.

Schlather *et al.* (2004) is an important study on the question whether a marked point process is externally marked. They suggest two devices for detecting evidence against the hypothesis of external marking. These devices are special cases of the cross-mark correlation of Stoyan (1987) and $f$-correlation by Penttinen and Stoyan (1989). Hence, edge corrected estimators are available. Straightforward statistical tests for external marking are not known.

In this presentation we consider external marking as a means of generating simple marked point process models. Two explicit constructions are suggested. The first one is for bivariate point processes with external marking of the "species". The second point process model is a marked point process couterpart of the mixed model, commonly applied in statistical modelling of dependent observations. Further we consider devices on studying external marking for vector-valued marked point processes. Two examples on the study of external marking are given.

References:

H.W. Lotwick, H.W. (1984). Some models for multitype spatial point processes, with remarks on analysing multitype patterns. *Journal of Applied Probability*, 21: 575–582.

Mase, S. (1996). The treshold method for estimating total rainfall. *Annals of the Institute of Statistical Mathematics*, 42: 201–213.

Penttinen, A.K. and Stoyan, D. (1989). Statistical analysis of a class of line segment processes. *Scandinavian Journal of Statistics*, 16: 153–161.

Schlather, M., Ribeiro Jr, P.J. and Diggle, P.J. (2004). Detecting dependence between marks and locations of marked point processes. *Journal of the Royal Statistical Society B*, 66: 79–93.

Stoyan, D. (1987). Statistical analysis of spatial point processes: A soft-core model and cross-correlations of marks. *Biometrical Journal*, 29(8): 971–980.

Wälder, O. and D. Stoyan, D. (1997). Models of markings and thinnigs of Poisson processes. *Statistics*, 29: 179–202.

Wälder, O. and Stoyan, D. (2000). On variograms in point process statistics, II: Models of markings and ecological interpretation. *Biometrical Journal*, 42(2): 171–187.

**Mats Rudemo**: Modelling from images, some examples in biology and material science

**Abstract**: Imaging is an increasingly important source of data for statistical analysis and modelling. In this talk a number of examples are discussed including two-dimensional gel images for measuring protein expression, microarrays for measuring gene expression, finding individual trees from aerial photographs, identifying 3D agarose gel structure from 2.5D transmission electro-microscopy images and finding inclusions in steel from ultrasonic images.

**Katja Schladitz**: Characterisation of the microstructure of filter materials

**Abstract**: Traditionally, in quality control for materials, microscopic images of 2d sections are analysed. For filter materials like open foams or nonwovens this approach is not feasible as the preparation of 2d sections is costly or even impossible and most information about the 3d geometry can not be recovered from the 2d sections. Tomographic images of microstructures provide the full 3d information. In this talk, a general method for measuring geometric characteristics based on integral geometric formulae is presented as well as model based methods for open foams and nonwovens.

**Anders Sjögren** (with Erik Kristiansson, Mats Rudemo and Olle Nerman): Weighting of microarrays to improve quality of inference - an empirical Bayes approach

**Abstract**: Studying gene expression levels using the DNA microarray technology is a process involving several consecutive steps, both of biological and analytical nature, each with varying quality. Established techniques for inference on microarray data do not take the variations in quality into account in a quantitative manner. They are therefore left with the choice of entirely removing or keeping arrays, basing the decision on ad-hoc methods. We propose a model where array-specific variance components are introduced in an empirical Bayes framework. Based on that model we derive a statistic for effects in paired situations, such as treatment experiments. In the statistic, the arrays are effectively weighted by the reciprocal of their variances, giving arrays of better quality higher weights. The method is of general nature in the sense that it is only utilising the actual data at hand and it is therefore applicable to both two-channel cDNA and oligonucleotide (Affymetrix) microarrays. The performance of the method is evaluated on simulated data and to some extent on real-world data. The talk will contain a brief introduction to the DNA microarray technology.

References:

G. Smyth. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology, 3(1), 2004.

**Sara Sjöstedt-de Luna**: Resampling methods for non-stationary spatial data

**Abstract**: Subsampling and block resampling methods have been suggested in the literature to non-parametrically estimate the variance and distribution of statistics computed from spatial data. Usually stationary data are required. However, in empirical applications, the assumption of stationarity often must be rejected. This talk presents nonparametric methods to estimate the variance and distributions of (functions of) sample means based on non-stationary spatial data using subsampling and resampling. It is assumed that data is observed on a lattice in some region of R. In the data considered, the information in the different picture elements (pixels) of the lattice are allowed to come from different distributions, with smoothly varying expected values, or with expected values decomposed additively into directional components. Furthermore, pixels are assumed to be locally dependent, and the dependence structure is allowed to differ over the lattice. Consistent variance and distribution estimators for (functions of) sample means are provided under these assumptions. An example with applications to forestry, using satellite data, is discussed. This talk is largely based on collaborative work with Magnus Ekstrm.

**Johan Svensson**: Discrete Approximations of Life Distributions in Optimal Replacement

**Abstract**: We use discretization of continuous random variables in a stochastic optimization problem. A measure connected to an optimization model that suggests what parts to replace in an aircraft engine is presented. The optimization model requires a discrete random variable with restrictions on the points of support. Four methods to make discretizations are discussed and adapted to the constraints of the model. The methods are compared and the choice concerning the number of points of support is discussed. Finally the consequence of using a so called narrow scenario tree is commented upon.

**Thomas Svensson**: Complexity versus scatter in fatigue modelling

**Lance Waller** (with Traci Leong and Andrew Barclay): One dimensional point processes and endangered species: Comparing intensity functions in a study of sea turtle nesting sites on Juno Beach, Florida

**Abstract**: We explore variations in the observed spatial pattern of sea turtle nesting behavior at Juno Beach, Palm Beach County, Florida for the 1998-2000 nesting seasons. Of particular interest is an assessment of possible effects due to a 990-foot fishing pier constructed in 1998-1999. The data include approximately 8,000-10,000 emergence locations per nesting season over 6 miles of beach with locations identified by global positioning system (GPS) units with sub-meter accuracy. Typical statistical analyses

(Chi-square and Kruskal-Wallis tests) suggest some significant changes between years in counts of emergences for certain marked zones but do not readily identify where significant local differences occur along the beach.

We conduct a spatial analysis by estimating the density of emergences (number of emergences per unit length of beach) as a function of beach location, and compare densities of emergences between nesting seasons, densities of nesting and non-nesting emergences within each season, and densities between species (green and loggerhead) within each season. The approach reveals significant decreases in emergence density (nesting, non-nesting, and total) near the pier in the first post construction year (1999) in contrast to 1998 and 2000. The approach also reveals a possible distributional shift in nesting locations in the second year post construction even though total emergence counts are similar. Finally, the approach suggests an impact of the pier on nesting behavior, i.e. a reduced probability of nesting per emergence in the immediate vicinity of the pier.