Thesis for the degree of Doctor of Philosophy

Percolation: Inference and Applications in Hydrology

Oscar Hammar

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
Gothenburg, Sweden, 2011

Percolation: Inference and Applications in Hydrology
Oscar Hammar
ISBN 978-91-628-8395-9

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Sweden
Telephone +46 (0)31 772 1000

Printed at the Department of Mathematical Sciences
Gothenburg, Sweden, 2011

Percolation: Inference and Applications in Hydrology
Oscar Hammar
Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg

Abstract

Percolation theory is a branch of probability theory describing connectedness in a stochastic network. The connectedness of a percolation process is governed by a few, typically one or two, parameters. A central theme in this thesis is to draw inference about the parameters of a percolation process based on information whether particular points are connected or not. Special attention is paid to issues of consistency as the number of points whose connectedness is revealed tends to infinity. A positive result concerns Bayesian consistency for a bond percolation process on the square lattice $\mathbb{L}^2$ - a process obtained by independently removing each edge of $\mathbb{L}^2$ with probability $1-p$. Another result on Bayesian consistency relates to a continuum percolation model which is obtained by placing discs of fixed radii at each point of a Poisson process in the plane, $\mathbb{R}^2$. Another type of results concerns the computation of relevant quantities for the inference related to percolation processes. Convergence of MCMC algorithms for the computation of the posterior, for bond percolation on a subset of $\mathbb{L}^2$, and the continuum percolation, on a subset of $\mathbb{R}^2$, is proved. The issue of convergence of a stochastic version of the EM algorithm for the computation of the maximum likelihood estimate for a bond percolation problem is also considered.

Finally, the theory is applied to hydrology. A model of a heterogeneous fracture amenable for a percolation theory analysis is suggested and the fracture's ability to transmit water is related to the fractures median aperture.

Keywords: percolation, inference, consistency, Markov chain Monte Carlo, hydrology

Acknowledgments

I would like to thank my supervisor Olle Häggström for his guidance and valuable comments on my manuscripts and my co-supervisor Anastassia Baxevani for reading my manuscripts and giving me useful feedback. I would also like to thank my co-authors Lisa Hernqvist, Gunnar Gustafson and Åsa Franson for good cooperation and Erik Järpe for comments on Paper 1. Moreover, I would like to thank Erik Kristiansson for all encouragement and Emilio Bergroth and other co-workers for good friendship. I would also like to take the opportunity to thank Evert Eggelind and Jakob Eidenskog for their encouragement during my undergraduate studies

Finally I would like to thank Kristina, mum, dad, my sisters with families, my parents-in-law and the rest of my relatives and friends for their inspiration and support.

Oscar Hammar
Gothenburg
November 2011

Table of contents

List of included papers

This thesis contains the following papers:

1. Oscar Hammar *Inference in a Partially Observed Percolation Process* Submitted to Latin American Journal of Probability and Mathematical Statistics

2. Oscar Hammar *Bayesian Consistency in a Partially Observed Percolation Process on the Infinite Square Lattice*

3. Oscar Hammar *Bayesian Consistency in a Partially Observed Continuum Percolation Process*

4. Oscar Hammar, Lisa Hernqvist, Gunnar Gustafson, Åsa Fransson *Relating the Hydraulic Aperture and the Median Physical Aperture for Rock Fracture with Large Aperture Variance using Percolation Theory* Submitted to International Journal of Rock Mechanics and Mining Sciences

Introduction

This thesis concerns applied percolation theory. The first three papers relate to statistical inference for percolation processes, focusing on consistency of such inferential procedures. In the last paper, percolation theory is used to model and analyse a real world phenomenon in hydrology.

In this introduction we provide with the necessary background. We initiate this presentation by providing with definitions of the different percolation models considered in this thesis: bond, site and continuum percolation. We discuss some central results in percolation theory in relation with the material in this thesis.

The two main approaches to statistical inference, the Bayesian and the frequentist, are considered in this thesis. We state the aim of inference in the two approaches and give some historical view of the development of the central concept of consistency. The few results on inference for discrete percolation are reviewed and the more comprehensive literature on inference in a different set-up than ours for continuum percolation processes is commented on.

The hydrological application in Paper 4 concerns flow of water in a fractured rock. We therefore give some background on earlier use of percolation theory to describe flow of water in fractured or porous media.

Bond percolation

The origin of percolation theory can be found in Broadbent and Hammarsley (1957), where what is today known as bond percolation, is introduced. Broadbent and Hammersley considered bond percolation on the lattice $\mathbb{L}^n$, with vertices $\mathbb{Z}^n$ and edges connecting vertices at a unit Euclidean distance. This process is given by independently declaring each edge of $\mathbb{L}^n$ open with probability p and closed with probability $1-p$.

The motivation for the model was given for $n = 3$. The authors described a porous stone by a large subset of $\mathbb{L}^3$ imagining the edges of this lattice being channels which with probability p, are wide enough to transmit water. The purpose of the model was to answer whether or not the centre of the stone would get wet if the stone would be put into water. This question is the prototype for the type of questions percolation theory is concerned with. Rephrased in terms of the mathematical model the question is: Does there exist an open path from the origin reaching the boundary of a large proportion of $\mathbb{L}^3$? In the limit, as the size of the subset of $\mathbb{L}^3$ tends to infinity in a suitable way, this question concerns the existence of an infinite open cluster of connected vertices.

Although initially considered on $\mathbb{L}^n$, percolation processes living on other graphs have been studied intensively and classes of graphs where interesting new phenomena occur have been found, see e.g. Häggström and Jonasson (2006). However, our primary interest in this work remains on bond percolation processes living on $\mathbb{L}^n$.

The critical probability p_c

The central question in percolation theory concerns the forming of clusters. With a path defined as an alternating sequence of vertices and edges connecting consecutive vertices, a path is said to be open respectively closed if all its edges are open respectively closed. Open and closed clusters are defined as sets of vertices connected by open respectively closed paths. To study the open clusters formed by a bond percolation process on $\mathbb{L}^n$ it is natural to define C to be the open cluster containing the origin of $\mathbb{L}^n$. Of particular interest in percolation theory is the possibility of an *infinite* open cluster. To study this, the percolation probability is defined as

$$\theta(p) = P_p(|C| = \infty),$$

where P_p denotes the probability measure on the set of all configurations of open and closed edges of $\mathbb{L}^n$ when retention parameter p is used. It is intuitively clear, and easily proved using a standard coupling argument, see e.g. Lindvall (1992), that $\theta(p)$ is increasing in p. It is therefore natural to define the critical probability as

$$p_c = \sup\{p : \theta(p) = 0\}.$$

Much of the interest in percolation theory steams from the fact, proved by Broadbent and Hammarsley (1957), that p_c is non-trivial, i.e. that $0 < p_c < 1$, for bond percolation on $\mathbb{L}^n$ for $n \geq 2$. It is intuitively clear (and follows from the translation invariance of the lattice $\mathbb{L}^n$ and the probability measure P_p) that if $\theta(p) = 0$ then *any* vertex belongs almost surely to a finite open cluster and thus all open clusters are almost surely finite. It is also possible to show that if $\theta(p) > 0$, then an infinite open cluster almost surely exists. This *phase transition property*, i.e. the drastic change in the macroscopic behaviour of the process at a critical value, is central to percolation theory.

Bond percolation on $\mathbb{L}^2$

We now turn to bond percolation on $\mathbb{L}^2$ which is the process considered in Paper 2. The value of p_c for this process attracted a lot of attention during the first decades following the introduction of percolation theory in 1957.

Hammersley (1959) proved that $0.35 < p_c < 0.65$ and the lower bound was soon enhanced to $0.5 \leq p_c$ when Harris (1960) proved $\theta(0.5) = 0$. Although generally believed, it turned out to be hard to prove that $p_c = 0.5$. A great moment in the history of percolation theory was when Kesten (1980) finally proved this long standing conjecture. Except for the trivial case $n = 1$, $n = 2$ is the only dimension of $\mathbb{L}^n$ for which the exact value of p_c is known.

Of particular interest for our application of percolation in Paper 2 is the question of possible co-existence of infinite open and closed clusters for bond percolation on $\mathbb{L}^2$. Combined with the result of Harris that $\theta(0.5) = 0$ the Kesten result ($p_c = 0.5$) implies that there is almost surely no infinite open cluster at p_c. By symmetry of the labellings 'open' and 'closed' this rules out the possibility of co-existence of infinite open and closed clusters for all $p \in [0, 1]$ almost surely.

Site percolation

There is an alternative way of introducing randomness in a graph. A site percolation model on a graph is given by independently declaring each vertex open with probability p and closed with probability $1 - p$. The physical interpretation is that fluid can flow through the open vertices and the edges but not through the closed vertices. Also site percolation processes posses phase transitions and the exact value of the critical probability p_c for site percolation is known for a few lattices. In particular, for site percolation on the triangular lattice, which is considered in Paper 4, it was proved by Wierman (1981) that $p_c = 1/2$.

Continuum percolation

Bond and site percolation processes are discrete in the sense that the positions of the sites are fixed. In contrast, the positions of the sites of a continuum percolation model are randomly spread out in a continuum such as $\mathbb{R}^2$. The first continuum percolation model was the Poisson blob model introduced by Gilbert (1961). The basic version of this model is given by placing discs of fixed radius ρ at each point of a Poisson process in the plane, $\mathbb{R}^2$.

Gilbert introduced the Poisson blob model to describe the transmission of information between telecommunication stations. The stations are distributed randomly (in the plane) and each station can communicate with other stations within distance 2ρ. The percolation theoretic question asked by Gilbert was whether or not there exists an infinite cluster of communicating stations.

While there are other ways of defining continuum percolation models, the Poisson blob model is the most well-studied. There are also numerous possible generalisations of the basic Poisson blob model. One generalisation

considered by Gilbert is to let the discs have random radii. Other natural generalisations are to consider the Poisson blob model in higher dimensions and allowing more general sets than discs or spheres. However, our focus is on the basic Poisson blob model in $\mathbb{R}^2$ with discs of fixed radii, which is the model considered in Paper 3.

The critical density λ_c

A realisation of the basic Poisson blob model partitions $\mathbb{R}^2$ in an occupied and a vacant component, where a point belongs to the occupied component if and only if it is covered by at least one disc. A connected subset of the open component is referred to as an occupied cluster and we denote by W the occupied cluster containing the origin. With P_λ denoting the probability distribution corresponding to the basic Poisson blob model on $\mathbb{R}^2$ with intensity λ and radii 1 and $d(D) = \sup_{x,y\in D} |x - y|$ denoting the diameter of a set $D \subset \mathbb{R}^2$, a percolation probability can be defined as

$$\theta(\lambda) = P_\lambda(d(W) = \infty). \tag{1}$$

Gilbert considered a related quantity, $\theta'(\lambda)$, defined as the P_λ-probability that the number of points in W of the underlying Poisson process is infinite. It is easy to prove that $\theta'(\lambda)$ coincides with $\theta(\lambda)$ and we, from now on write $\theta(\lambda)$ for both. Gilbert proved that $\theta(\lambda) = 0$ for sufficiently small λ and that $\theta(\lambda) > 0$ for sufficiently large λ, thereby proving the existence of a non-trivial, i.e. positive and finite, critical density $\lambda_c = \inf\{\lambda : \theta(\lambda) > 0\}$. Straightforward arguments were used to derive rough bounds on λ_c.

The upper bound on λ_c was derived by comparing the Poisson blob model on $\mathbb{R}^2$ with a bond percolation process on $\mathbb{L}^2$ and using the critical value p_c for the latter model to derive a bound on the critical density of the former model. With the best bound $p_c < 0.65$ established at the time, the analysis of Gilbert yielded $\lambda_c < 2.10$ and with $p_c = 0.5$ it yielded $p_c < 1.38$.

To derive a lower bound on λ_c, the number of points in W of the underlying Poisson process was compared to the number of points of a simple branching process. With $S(r, x)$ denoting the closed disc with radius r centred in x and with $\mathbf{0}$ denoting the origin of $\mathbb{R}^2$, the first generation of points of the branching process is given by a Poisson process on $S(1, \mathbf{0})$ with intensity λ. If $q_1, \ldots, q_K$ denotes the n^{th} generation of points of the branching process, then the $(n + 1)^{\text{th}}$ generation of points is given by independent Poisson processes with intensity λ on $S(1, q_k)$ for $k = 1, \ldots, K$. The probability that this branching process generates at least n points is clearly larger than the probability that W contains at least n points of the Poisson process underlying the Poisson blob model. Since a disc with radius 1 has area π it is intuitively clear, and a basic result in branching processes theory (Asmussen and Hering,

1983), that if $\lambda < \pi^{-1}$, then there will almost surely be only finitely many points generated by the branching process. Refining this argument slightly, Gilbert arrived at $0.140 < \lambda_c$. The bounds derived by Gilbert have been enhanced but the interval of possible values of λ_c is still relatively wide. The best rigorous bounds, $0.174 < \lambda_c < 0.843$, are derived in Hall (1985) using techniques similar to those of Gilbert.

The critical density $\lambda_c^\star$

There is another natural notion of clusters in the Poisson blob model. As well as asking whether there exists an infinite *occupied* cluster one might ask whether there exists an infinite *vacant* cluster, i.e. a connected subset of the vacant region. With V denoting the vacant cluster containing the origin, a percolation probability for the vacant cluster can be defined as

$$\theta^\star(\lambda) = P_\lambda(d(V) = \infty).$$

Results concerning the critical density $\lambda_c^\star = \sup\{\lambda : \theta^\star(\lambda) > 0\}$ for a vacant cluster have turned out to be much harder to prove than those concerning λ_c. For the special case where the radii of discs in a Poisson blob model are given by an almost surely bounded random variable, it was proven by Roy (1990) that $\lambda_c = \lambda_c^\star$.

A crucial issue in Paper 3 is the possibility of co-existence of an occupied and a vacant infinite cluster in the basic Poisson blob model on $\mathbb{R}^2$ with discs of fixed radii. For this model, we have $\lambda_c = \lambda_c^\star$ and it was proven by Alexander (1996) that $\theta(\lambda_c) = \theta^\star(\lambda_c) = 0$, which rules out co-existence of an occupied and a vacant infinite cluster almost surely in this case.

Power laws

The phase transition property of percolation processes has made these models attractive to physicists in their quest to understand this empirically known phenomenon. A lot of the early research in percolation theory (e.g. Kirkpatrick (1973), Sur et al. (1976)) focused on the function $\theta(p)$ for values just above the critical value where these phase transitions take place. Simulations indicated that for p just above p_c, $\theta(p)$ behaves roughly as a power of $p - p_c$, i.e. $\theta(p) \approx (p - p_c)^\beta$ for some critical exponent β. Power laws were also suggested for quantities other than $\theta(p)$.

The only rigorously proved power laws are given by Smirnov and Werner (2001) whose results concern the special case of site percolation on the triangular lattice. They proved the existence of the conjectured critical exponent $\beta = 5/36$ relating θ and $p - p_c$ for this special case. In addition, they proved a power law that is crucial to our analysis of flow in a fracture in Paper 4.

This result concerns the correlation length ξ of a percolation process, which in terms of the retention parameter p is defined as

$$\xi(p) = \big(\lim_{n\to\infty} -\frac{1}{n} \log P_p(0 \leftrightarrow n) \big)^{-1},$$

where $P_p(0 \leftrightarrow n)$ denotes the probability of an open path from the origin to a point n steps away along the x-axis when each site is open with probability p (Grimmett, 1999). Smirnov and Werner (2001) proved that for the special case of site percolation on the triangular lattice,

$$\lim_{p\to p_c} \xi(p) = (p - p_c)^{\nu + o(1)} \quad \text{where } \nu = -3/4. \tag{2}$$

Statistical inference

A problem of statistical inference starts with a question concerning some real world phenomenon. In order to answer this question a model of the phenomenon is assumed and data are collected. The process of choosing a model that accurately describes reality and is amenable for analysis is an important step. However, the main part of this thesis concerns the step of the inference process after a statistical model has been accepted.

A statistical model for data, $\mathbf{X}$, is a collection of probability distributions $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ on $\mathcal{X}$, where $\mathcal{X}$ is the sample space of $\mathbf{X}$, i.e., the set of all possible outcomes of $\mathbf{X}$. In this thesis we consider only parametrized models which means that the parameter θ is finite dimensional. The aim of statistical inference is to use data $\mathbf{X}$ to draw inference about $\theta \in \Theta$.

Types of statistical inference

There are two main frameworks for statistical inference: the Bayesian and the frequentistic. The Bayesian approach to inference is the older of the two, having its roots in a paper by Thomas Bayes in 1763, and was the dominant approach until the frequentist school emerged during the first half of the twentieth century. The main figure in the development of the frequentist school was Ronald Fisher who, in a series of publications leading up to a famous paper (Fisher, 1922), laid down a new framework for statistical inference. In this famous paper, the concept of consistency was introduced which, with its Bayesian counterpart is a central theme in Papers 1-3. Other important contributors to the frequentist school were Jerzy Neyman and Egon Pearson who formulated the method of hypothesis testing (Neyman and Pearson, 1933) and Kolmogorov who formulated the axioms of probability on which the frequentist school is based (Kolmogorov, 1933).

The new frequentistic ideas about inference stimulated a formalization of the old Bayesian ideas about inference. The most comprehensive framework for the Bayesian approach to inference was formulated by Savage (1954), where a non-frequentist alternative to Kolmogorov's axioms of probability was stated. During the second half of the twentieth century the statistical community was divided into frequentists and Bayesians with, sometimes heated debates between proponents of the two schools. Nowadays, while the debate is still ongoing, more statisticians conform to a pragmatic view where the two approaches are seen as complementary (Bayarri and Berger, 2004). In the subsequent sections we demonstrate the main differences between the two approaches.

Different interpretations of probability

Statistical methods are based on probability theory and the discrepancy between the Bayesian and the frequentistic schools emerges already at the interpretation of the fundamental concept of probability. Simply stated, in the Bayesian context probability can be assigned to any event, which allows probabilities to be used to express personal degree of belief concerning events. Such a use of the probability concept is not accepted in the freqentist context. Here probabilities are only assigned to events resulting from some 'experiment'. The probability of an event is then interpreted as the limiting frequency of times the event occurs, as the number of identical experiments conducted tends to infinity.

Inference in the Bayesian context

The different interpretations of probability leads inevitably to different ideas of how to use data to draw inference about $\theta \in \Theta$. With the Bayesian interpretation it is natural to express knowledge about θ by a probability distribution over Θ. The knowledge about θ before data have been observed is expressed by a prior distribution and the aim of Bayesian inference is to use data to update the prior distribution to a posterior distribution which reflects an enhanced knowledge about θ.

A prior distribution Π defines, together with the statistical model $\{P_\theta : \theta \in \Theta\}$, a probability distribution Δ_Π over $\mathcal{X} \times \Theta$. The posterior distribution $\Pi(\cdot|\mathbf{X})$ is the Δ_Π-probability of θ given data $\mathbf{X}$, which is computed using Bayes' Theorem, i.e. for a suitable subset A of Θ,

$$\Pi(A|\mathbf{X}) = \Delta_\Pi(A|\mathbf{X}) = \frac{\Delta_\Pi(A \cap \mathbf{X})}{\Delta_\Pi(\mathbf{X})}. \tag{3}$$

An accumulation of the probability mass of the posterior in a certain region of Θ signifies an increased belief of the value of θ being in that region.

Inference in the frequentistic context

In the frequentist context there is an unknown $\theta_0 \in \Theta$ and data are assumed to be generated under P_{θ_0}. The aim of inference is to use data to conclude the most plausible value or values for θ_0. For this, an estimator, which is a function of the data $\mathbf{X}$, is used. An important class of estimators were introduced by Fisher in terms of the likelihood function, which for given data $\mathbf{X}$, is $P_\theta(\mathbf{X})$ viewed as a function of θ. The maximum likelihood estimator (MLE) $\hat{h}(\mathbf{X})$ maximizes this function, i.e.,

$$\hat{h}(\mathbf{X}) = \arg\max_{\theta \in \Theta} P_\theta(\mathbf{X}). \tag{4}$$

Usually more than a point estimate is reported, e.g. a confidence interval or a decision to reject or accept a hypothesis. In this case the notion of probability is used very differently from the way it is used in the Bayesian context. Whereas in the Bayesian context probability statements are made about θ the probability statements in the frequentist context are made about the procedures by which inference is drawn, i.e. 0.95 is the limiting frequency of times a 95% confidence interval covers θ_0, as the number of intervals constructed tends to infinity.

Consistency

Consistency is an asymptotic property of an inference procedure which guarantees that the correct inference is drawn in the limit as the amount of data tends to infinity. In the following discussion it is convenient to assume $\mathbf{X} = (X_i)_{i=1}^{\infty}$ and let $\mathbf{X}_n = (X_i)_{i=1}^{n}$.

Consistency in the Bayesian approach

An example of consistency for independent and identically distributed (i.i.d.) Bernoulli random variables using a prior with positive and continuous density on $(0, 1)$ were given already by Laplace in the nineteen century. However, he did not consider this as an example of a general concept and an early definition of Bayesian consistency is instead given by Doob (1949).

A sequence $\Pi(\,\cdot\,|\mathbf{X}_n)$ of posterior distributions is said to be (strongly) consistent at θ if for each neighbourhood U of θ,

$$\lim_{n\to\infty} \Pi(\,\cdot\,|\mathbf{X}_n) = 1 \quad P_\theta\text{-a.s.}$$

This property clearly reflects an increasingly strong correct belief about θ. Doob (1949) showed a general result on Bayesian consistency under weak conditions for all parameters in a set with prior measure 1. An alternative is

to require consistency for *all* parameters in the parameter space. A central result on Bayesian consistency in this direction is given by Schwartz (1965) who proved that, if the prior puts positive probability on every Kullback-Leibler neighbourhood of the true parameter, then the posterior accumulates in all weak neighbourhoods of the true parameter. This result is central for the proofs of Bayesian consistency in Papers 1-3.

Consistency in the frequentist approach

A desirable property of an estimator $h_n = h(X_1, \ldots, X_n)$ of θ is that it is unbiased, i.e. that $E_\theta[h] = \theta$, where E_θ denotes expectation under P_θ. This, together with a small variance for the estimator, guarantees that the estimated value is close to the true value with high probability. A milder requirement is to demand that a sequence of estimators $(h_n)_{n=1}^\infty$, where $h_n = h(X_1, \ldots, X_n)$, is *unbiased in the limit* as n tends to infinity, i.e. that $\lim_{n\to\infty} E_\theta[h_n] = \theta$. If additionally the variance of h_n tends to zero as n tends to infinity, then the sequence of estimators is consistent. In this case an estimated value of θ, based on a large sample, is close to the true value with high probability. Formally, $(h_n)_{n=1}^\infty$ is said to be (strongly) consistent if $\lim_{n\to\infty} h_n = \theta$ almost surely.

Early results for consistency of MLE's were given by Fisher. These results were elaborated during many years and the conditions under which consistency holds were not always clear. Rigorous proofs of consistency of MLE's were given by Wald (1949) and Cramer (1946b,a). These results concern the i.i.d. case. In Paper 1 we prove consistency for the MLE for data that are independent but not identically distributed.

Inference for percolation processes

Results in the literature on inference for discrete percolation processes are limited. We present some of the few exceptions. Meester and Steif (1998) considered estimation of various quantities such as $\theta(p)$ and $I_{\{\theta>0\}}$ from a realisation of a bond percolation process on $\mathbb{L}^d$ observed in a $B(n) = [-n, n]^d \subset \mathbb{L}^d$ and showed that their frequentistic estimation procedure is consistent as the size n of the box $B(n)$ tends to infinity.

Larson (2010) considered an inference problem for a first passage percolation process. This is a time dependent version of ordinary percolation, introduced by Hammersley and Welsh (1965). A flow, e.g. water, is spreading on a graph between neighbouring vertices via the edges. The passage times, i.e. the times it takes for the flow to spread between neighbouring vertices, are independent and identically distributed random variables. The process of how the vertices are wetted is observed but the process of the edges trans-

mitting the water is unobserved. Consequently, when a vertex with several wet neighbours is wetted it is not known where the flow came from. The objective is to estimate the distribution of passage times based on this limited information. The inference problem and the type of data studied by Larson are similar to what we consider in Papers 1-3.

Inference for the Poisson blob model, especially in two dimensions, has been studied a lot. Often Poisson blob models with general random sets, i.e. not necessarily discs, centred at the points of a Poisson process are considered. However, the main challenge is the same. In the general set-up, a Poisson blob model on $\mathbb{R}^2$ is observed in a finite box $B(n) = [-n, n]^2$ and the intensity of the underlying Poisson process, as well as parameters governing the random sets, are sought to be estimated. The main difficulty arises as some sets might be completely covered by other sets. The literature on suggestions of inference procedures in presence of this difficulty is quite large and consistent estimators in the frequentist context of the intensity of the underlying Poisson process are available (Molchanov, 1995).

Percolation and hydrology

Although Broadbent and Hammersley used a motivating example of hydraulic flow when introducing their model in 1957, the adoption of percolation theory in hydrology was initially slow. The first serious attempts to use percolation theory to describe flow in heterogeneous media were made in the eighties. Two examples are Wilke et al. (1985) who evaluated the permeability of a fractured rock by a percolation model using simulations and Halperin et al. (1985) who used heuristic arguments to estimate critical exponents relating fluid permeability of a disordered media to percolation parameters. In addition, a method called critical path analysis, originally developed for analysis of a current flow, was adopted for the analysis of hydraulic flow.

Critical path analysis

Ambegaokar et al. (1971) introduced critical path analysis (CPA) in the context of electronic transport. They considered a large volume consisting of randomly distributed sites. Each site has a given energy and sites i and j are connected by a conductance $G_{i,j}$ depending on their relative positions and energies. Electrons are hopping between sites i and j with an intensity depending on $G_{i,j}$.

The authors argued that when the conductances vary over many orders of magnitude the network of conductances can be considered to be composed of three parts. The first part consists of isolated regions of high conductivity, the second part is a set of relatively few resistors of moderate conductivity

connecting the isolated regions of high conductivity and the third part consists of resistors with low conductivity which has limited contribution to the systems conductivity. The first two parts make up what is called the critical subnetwork. The resistors in the part with moderate conductances define a critical conductance G_c which is the lowest conductance of the critical subnetwork. Ambegaokar and co-authors argued that this critical conductance governs the conductance of the whole system.

Developments of critical path analysis

Critical path analysis was refined by Shante (1977) who argued that the conductance of the system is governed by an optimal conductance somewhat smaller than the critical conductance given by Ambegaokar and co-authors. Starting from the critical subnetwork given by Ambegaokar and co-authors, he added all resistors with conductances somewhat smaller than the critical conductance. The created subnetwork contains worse conductors than the critical subnetwork considered in Ambegaokar et al. (1971), but the number of paths through the system is increased. Shante argued that this is the relevant subnetwork to consider. An optimization procedure is carried out to find the optimal conductance, which balances between large conductances along the paths and a large number of paths.

Katz and Thompson (1986) translated the ideas of CPA for transport of electrons to transport of fluid flow in porous media. This work as well as other attempts, by e.g. Charlaix et al. (1987) and Hunt (2005), to use CPA to describe fluid flow relies on a conjectured power law relating the transport capacity of a percolation process to the difference $p - p_c$. In Paper 4 a critical path analysis is carried out for the transport of water in a single heterogeneous fracture. This critical path analysis is not based on conjectured critical exponents. We model a single fracture with a site percolation process on the triangular lattice and use one of the rigorously proven power laws given by Smirnov and Werner (2001).

Summary of papers

Summary of Paper 1 - Inference in a Partially Observed Percolation Process

A type of inference problem relating to percolation processes is introduced and a first special case is analysed. A distinguishing feature from earlier work on inference for percolation process is the type of data that is considered. The data are defined in terms of a set of observation points, which for a bond percolation process is a subset of the vertices of the graph on which the process lives. Each data point is specified to carry the information on whether or not a particular pair of observation points are connected by an open path. This situation is referred to as the percolation process being partially observed as opposed to the situation when a full realisation of a percolation process on a subgraph is observed.

Both Bayesian and frequentist consistency is considered for inference in a partially observed bond percolation process. In order to prove consistency results, the bond percolation process is restricted to live on a particular class of graphs consisting of identical copies of isolated finite subgraphs. With this imposed restriction the data from the partially observed process has a structure similar to i.i.d. observations, a fact used to prove consistency.

In addition to the focus on consistency, another issue is the computation of relevant quantities for the inference procedures. Theses quantities, i.e. the maximum likelihood estimate (MLE) in the frequentist approach and the posterior distribution in the Bayesian approach, does not allow direct computations. Instead, algorithms approximating these quantities are needed. We implement a Markov chain Monte Carlo (MCMC) algorithm for the computation of the posterior distribution and a stochastic version of the EM algorithm for the computation of the MLE and present proofs that these algorithms converge in a suitable sense. The paper also contains an extensive simulation study which evaluates the Bayesian and frequentist approach with respect to accuracy and computation load. It is found that the stochastic EM algorithm introduces some bias due to problems with finding initial values in presence of a phase transition of the model.

Summary of Paper 2 - Bayesian Consistency in a Partially Observed Percolation Process on the Infinite Square Lattice

In this paper the issue of Bayesian consistency for inference in a partially observed bond percolation process is explored further. Here we consider the more challenging and physically more interesting case of bond percolation on

the infinite square lattice $\mathbb{L}^2$. In this setting the connectedness of pairs of observation points are not independent which leads to dependent data. The basic idea for handling this problem is to, for each of a number of pairs of observation points, place a box around the pair and to consider connectedness of the pair of points by a path within the box. With non-overlapping boxes these events are clearly independent. By considering pairs of observation points within larger and larger boxes and applying results from percolation theory we can control the dependence. It is shown that if each element of $\mathbb{Z}^2$ is included independently in the set of observation point with some probability $r > 0$ and if the prior has full support on the parameter space, then there is Bayesian consistency for all parameters in a set of prior measure 1.

Summary of Paper 3 - Bayesian Consistency in a Partially Observed Continuum Percolation Process

In this paper inference for a partially observed percolation process is elaborated futher by considering a two-parameter percolation process. The inference problem considered in Papers 1 and 2 is translated to a continuum by instead of bond percolation study the basic Poisson blob model with discs of fixed radii. In this context the set of observation points is a subset of $\mathbb{R}^2$ and the data consist of information on connectedness of pairs of observation points by curves entirely contained in the occupied region.

Inference is drawn about the two dimensional parameter (λ, ρ), where λ is the intensity of the underlying Poisson process and ρ is the radii of the discs centred at the points of the Poisson process. As in Paper 2 the data from the model are non-independent. However, the main challenge is the dimension of the parameter (λ, ρ). It is shown that if the observation points are given by a homogeneous Poisson process independent of the Poisson blob model and if the prior has full support on the parameter space then there is Bayesian consistency for all parameters in a set of prior measure 1.

An algorithm for the computation of the posterior distribution is presented and proved to converge in a suitable sense. Moreover, we present some inference on simulated data which indicates that it is relatively easy to infer the covered area proportion, but it is harder to decide whether there are many small discs or few larger ones.

Summary of Paper 4 - Relating the hydraulic aperture and the median physical aperture for rock fracture with large aperture variance using percolation theory

Paper 4 is quite different from Papers 1-3 and does not relate to inference for percolation processes. Instead, we model a fracture with highly varying

physical aperture and use percolation theory to evaluate its ability to transmit water. The aim is to relate the median physical aperture, a_c, to the fracture's transmissivity T, i.e. the rate of water flow through the fracture. This relation is used to relate a_c to a measurable quantity b (the hydraulic aperture) with a known relation to T. The varying physical aperture is captured by a model where the fracture is partitioned into hexagonal cells and each cell is assigned independently a log normally distributed aperture.

If a fracture has constant aperture, then a physical law (the cubic law) is applicable for relating its transmissivity to its aperture. This law is applied to each cell of the model to transform its assigned aperture to a local transmissivity. For fractures with the level of variation of physical aperture considered in the paper the ratio between the 90^{th} and the 50^{th} percentile of the distribution of local transmissivities is of order 100. With this large difference between well and really well transmitting cells the flow through the fracture naturally takes place in a few dominant paths. The transmissivity of one of these dominant paths is determined by the worse cell along the path. With the highly varying local transmissivities it is important to carefully determine the level of transmissivity of the dominant paths, which is done using critical path analysis. This analysis includes strong simplifications and the result is a lower bound, rather than an exact value, on the fractures transmissivity under the model.

The hydraulic aperture, b, of a fracture is related to the transmissivity T of the fracture by the cubic law, i.e. for for some constant C,

$$T = Cb^3. \tag{5}$$

The critical path analysis yields a relation between the median physical aperture, a_c, and the transmissivity, T, of the fracture:

$$T > C(da_c)^3. \tag{6}$$

Combining (5) and (6), the reciprocal of the factor d gives a bound of the factor of overestimation when using b as an estimate of a_c. The value of d depends on the level of variance in physical aperture and for the fractures we consider, the analysis gives $d^{-1} \in (3.34, 4.03)$. We use simulations to get a rough estimate on the level of underestimation of the transmissivity introduced by the CPA. Compensating by this estimated factor we arrive at $d^{-1} \in (1.97, 2.36)$ for the fractures considered.

References

Alexander, K. S. (1996). The RSW Theorem for Continuum Percolation and the CLT for Euclidean Minimal Spanning Trees. *The Annals of Applied Probability*, 6:466–494.

Ambegaokar, V., Halperin, B. I., and Langer, J. S. (1971). Hopping Conductivity in Disordered Systems. *Physical Rewiew B*.

Asmussen, S. and Hering, H. (1983). *Branching Processes*. Birkhauser, Boston.

Bayarri, M. J. and Berger, J. O. (2004). The Interplay of Bayesian and Frequentist Analysis. *Statist. Sci.*, 19:58–80.

Broadbent, S. R. and Hammarsley, J. M. (1957). Percolation processes I. Crystals and mazes. *Proceedings of the Cambridge Philosophical Society*, 53:629–641.

Charlaix, E., Guyon, E., and Roux, S. (1987). Permeability of a Random Array of Fractures of Widly Varying Apartures. *Transport in porous media*, 2:31–43.

Cramer, H. (1946a). A contribution to the theory of statistical estimation. *Skand. Akt. Tidskr*, 29:85–94.

Cramer, H. (1946b). *Mathematical Methods of Statistics*. Princeton Univeristy Press.

Doob (1949). Applications of the theory of martingales. *Le Calcul des Probabilités et ses Applications*, pages 23–27.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society, A*, 222:309–368.

Gilbert, E. N. (1961). Random Plane Networks. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):533–543.

Grimmett, G. (1999). *Percolation*. Springer, 2 edition.

Häggström, O. and Jonasson, J. (2006). Uniqueness and non-uniqueness in percolation theory. *Probability Survey*, 3:289–344.

Hall (1985). On continuum percolation. *Ann. Prob*, 13:1250–1266.

Halperin, B. I., Feng, S., and Sen, P. N. (1985). Differences between Lattice and Continuum Percolation Transport Exponents. *Phys. Rev. Lett*, 54:2391–2394.

Hammersley, J. M. (1959). Bornes supérieures de la probabilité critique dans un processus de filtration . *Le Calcul de Probabilitité et ses Applications*, pages 17–37.

Hammersley, J. M. and Welsh, D. J. A. (1965). *Bernoulli, Bayes, Laplace Anniversary Volume*. Springer.

Harris, T. E. (1960). A lower bound for the critical probability in a certain pecolation process. *Proceedings to the Cambridge Philosophical Society*, 56:13–20.

Hunt, A. (2005). *Percolation Theory for Flow in Porous Media*. Springer.

Katz, A. J. and Thompson, A. H. (1986). Quantitative prediction of permability in porous rock. *Physical review B*, 34:8179–8181.

Kesten, H. (1980). The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. *Communications in Mathematical Physics*, 74:41–59.

Kirkpatrick, S. (1973). The nature of percolation 'channels'. *Solid State Communications*, 12:1279–1283.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin.

Larson, K. (2010). Estimation of the passage time distribution on a graph via the EM algorithm. Technical report, Umeå University.

Lindvall, T. (1992). *Lectures on the Coupling Method*. Wiley Series in Probability and Mathematical Statistics.

Meester, R. and Steif, J. (1998). Consistent Estimation of Percolation Quantities. *Statistica Neerlandica*, 52:226–238.

Molchanov, I. S. (1995). Statistics of the Boolean Model: From the Estimation of Means to the Estimation of Distributions. *Advances in Applied Probability*, 27:63–86.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. Ser. A*, 231:289–337.

Roy, R. (1990). The RSW theorem and the equality of critical densities and the 'dual' critical densities for continuum percolation on $\mathbb{R}^2$. *Ann. Prob.*, 18:1563–1575.

Savage, L. J. (1954). *The foundations of statistics*. New York: John Wiley and Sons.

Schwartz, L. (1965). On Bayes procedures. *Probability Theory and Related Fields*, 4(1):10–26.

Shante, V. K. S. (1977). Hopping conduction in quasi-one-dimensional disordered compounds. *Physical review B*, 16(6):2597–2612.

Smirnov, S. and Werner, W. (2001). Critical exponents for two-dimentional percolation. *Mathematical Research Letters*, 8:729–744.

Sur, A., Lebowitz, J. L., Marro, J., Kalos, M. H., and Kirkpatrick, S. (1976). Monte Carlo studies of percolation phenomena for a simple cubic lattice. *Journal of Statistical Physisc*, 15:345–353.

Wald, A. (1949). Note on the Consistency of the Maximum Likelihood Estimate. *Ann. Math. Statist.*, 20:595–601.

Wierman, J. C. (1981). Bond percolation on honeycomb and triangular lattices. *Adv. in Appl. Probab.*, 13:298–313.

Wilke, S., Guyon, E., and de Marsily, G. (1985). Water penetration through fractured rocks: Test of a tridimensional percolation description. *Mathematical Geology*, 17:17–27.