World Applied Sciences Journal 2 (5): 548-551, 2007
ISSN 1818-4952

Predicting the Excursion Set of Gaussian Random Field

[1]Q.D. Katatbeh, [2]M.T. Alodat and [3]J.W. Kristof

[1]Department of Mathematics and Statistics, Jordan University of Science and Technology, Irbid, Jordan
[2]Department of Statistics, Yarmouk University, Irbid, Jordan
[3]Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Canada

Abstract: The statistics of the excursion set of random fields are common measures of reliability for many engineering systems. Predicting the excursion set of a random field and its statistics are the topics of this paper. A method for predicting the excursion set of a smooth and stationary Gaussian random field is discussed.

Key words: Gaussian field . excursion set . soil contamination

INTRODUCTION

Understanding various engineering properties of the soil is the goal of many geotechnical problems [2]. This includes the intrinsic soil properties, the shear strength, the soil type and the level of contamination in the soil. Crude oil contamination in the soil is one of the factors that affects the shear strength and we are interested in determining the statistical measures for this soil property in a given region. If we observe the soil property in a given region C, then we may want to predict the soil property in another region D [3]. These properties are unknown and may be modeled by a set of random variables. Since these quantities vary spatially, they may be modeled by a random function or random field. A random field is simply a collection of random variables indexed by a spatial set. The main interest of the researchers is to determine the probability distribution of the soil properties in a region of interest. For example, we may be interested in determining the probability that a soil property exceeds some given threshold u in a region of interest, or the proportion of the space where the soil property exceeds u [5, 7]. These statistical properties are used as measures of reliability for the soil used in the structures. The aim of this paper is to predict the excursion set and some of its characteristics of a smooth and stationary Gaussian random field in a given region of interest based on a realization of the field on a region.

To setup the notation assume that the region of interest is $C \subset R^d$, the d-dimensional Euclidean space. A family of random variables$\{X(t),\ t \in C \subset R^d\}$, $d>1$, is called a d-dimensional random field. If $d = 1$, the family is called a random process. For every random field $X(t)$, two functions can be defined, the mean function $\mu(t) = E\{X(t)\}$ and the covariance function $K(t,\ s) = \mathrm{cov}\{X(t),\ X(s)\}t,\ s \in C$. A d-dimensional random field is called a Gaussian random field if $(X(t_1),\ldots, X(t_n))$ is a multivariate normal distribution for every choice$\{t_1,\ldots,t_n\} \in C$.

The mean and the covariance matrix of $(X(t_1),\ldots, X(t_n))$ are given by $\mu = (\mu(t_1),\ldots, \mu(t_n))$ and

$$M = \left(K(t_i, t_j)\right)_{i,j=1}^{n}$$

The random field $X(t)$ is said to be stationary if $(X(t_1),\ldots, X(t_n))$ and $(X(t_1+h),\ldots, X(t_n+h))$ have the same distribution for any $h \in R^d$ and is said to be isotropic random field if $(X(t_1),\ldots, X(t_n))$ and $(X(qt_1),\ldots, X(qt_n))$ have the same distribution for any rotation q in R^d. For a stationary random field the mean function is constant, i.e., $\mu(t) = \mu$ for every $t \in R^d$.

In this paper, we assume that $X(t)$ is smooth and stationary Gaussian random field with mean $\mu(t) = \mu$ and variance $\mathrm{Var}\{X(t)\} = \sigma^2$.

Let $X_i(t)$ be the first derivative of $X(t)$ with respect to the i^{th} coordinate of t and $X_{ij}(t)$ be the second partial derivative of $X(t)$ with respect to i^{th} and j^{th} coordinates. We also assume that the following condition is fulfilled

$$\max_{i,j} E\left\{\left|X_{ij}(t) - X_{ij}(0)\right|^2\right\} \le c\|t\|^2$$

for $c>0$ and t in some neighborhood of 0. Here $\|.\|$ denotes the Euclidean norm in R^D.

The excursion set of a random field $X(t)$ in C above a level u is defined as the set of points $t \in C$ for

Corresponding Author: Dr. Q.D. Katatbeh, Department of Mathematics and Statistics, Jordan University of Science and Technology, Irbid, Jordan

which $X(t) \geq u$. Let us denote the excursion set of $X(t)$ in C above u by $A(X, u, C)$. The excursion set is very important and has been studied extensively in [1]. With probability tending to one as $u \to \infty$, the excursion set of smooth Gaussian random field $X(t)$ has simpler topology, i.e., it is a union of disjoint convex components where each convex component contains one local maximum of $X(t)$. Moreover, N, the number of convex components of $A(X, u, C)$, follows approximately the Poisson distribution [1]. The mean of this Poisson distribution is given by

$$E\{N\} = \mathrm{vol}(C)\det(\Lambda)^{\frac{1}{2}}\sigma^{-(2d-1)}u^{d-1}(2\pi)^{-(d+1)/2}\exp\left(-\frac{u^2}{2\sigma^2}\right) \quad (1)$$

Where, vol(C) is the volume of C and Λ is the covariance matrix of $(X_1(t),\ldots, X_d(t))$. Then $E\{N\}$ can be used to find the following accurate approximation for $P\{\sup_{t\in C} X(t) \geq u\}$:

$$P\left\{\sup_{t\in C} X(t) \geq u\right\} \approx E\{N\}$$

So the problem of approximating $P\left\{\sup_{t\in C} X(t) \geq u\right\}$ is reduced to the problem of approximating $E\{N\}$.

THE PROBLEM

Let t be a location in a region of interest C and $X(t)$ be the soil property at t. Let $X(t)$ be a smooth and stationary Gaussian random field. Let $X_1 = (X(t_1),\ldots,X(t_n))$ be the observed values of $X(t)$ at the locations $t_1,\ldots t_n \in C \backslash D = \{t \in C: t \notin D\}$, where $D \subset C$. Our aim is to predict the excursion set of $X(t)$ and its characteristics in the domain D, i.e., to predict $X_2 = (X(s_1),\ldots,X(S_m))$ where $s_1,\ldots,s_m \in D$.

If we denote an n-dimensional vector of ones by l_n, then from the multivariate normal theory, the stacked vector $X^T = (X_1, X_2)$ has (n+m)-dimensional multivariate normal distribution with mean $\mu = \mu l_{n+m}$, where l_{n+m} is (n+m)-dimensional vector of ones and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}$$

Where,

$$\Sigma_{11} = \left(K(t_i,t_j)\right)_{i,j=1}^{n}, \Sigma_{22} = \left(K(s_i,s_j)\right)_{i,j=1}^{m}$$

and

$$\Sigma_{12} = \left(K(t_i,s_j)\right)_{i=1,j=1}^{n,m}$$

The conditional distribution of X_2 given $X_1 = x_1$ is also m-dimensional multivariate normal with mean

$$\mu_{2,1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)$$

and covariance matrix

$$\Sigma_{2.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}^T$$

Where,

$\mu_1 = \mu l_n$ and $\mu_2 = \mu l_m$. The mean $\mu_{2.1}$ is a function in x_1 which can be used to predict X_2. Various covariance functions for $X(t)$ are available in the literature [4]. A common choice is the following one

$$K(t,s) = \exp\left(-\frac{1}{2\tau^2}\|t-s\|^r\right), \ r \in (0,2], \tau > 0 \qquad (2)$$

PREDICTION

Let us denote the predictive distribution of X_2 given $X_1 = x_1$ by $f(x_2|x_1)$. The predictive distribution of X_2 given $X_1 = x_1$ depends on the parameters σ^2, μ and τ^2. So we estimate them using the data X_2 and then we plug the estimates in the density $f(x_2|x_1)$. To predict the characteristic of the excursion set $A(X, u, D)$ in D, we simulate a large sample from the distribution of X_2 given $X_1 = x_1$. These realizations can then be used to predict the size of the excursion set, the cluster size, the number of components above u and $\sup_{t\in D} x(t)$.

The general form of the predictor is $E\{H(X_2)|X_1 = x_1\}$, where $H(X_2)$ denotes a characteristic of the excursion set of $X(t)$ in D. Since it is not possible to simulate a random field on a compact set D, we use instead $\tilde{D}$, a grid of D. Let X_{2j}, $j = 1,\ldots, M$ be M realizations from $f(x_2|x_1)$. Then the following characteristics of the excursion set can be predicted

Size of $A(X, u, D)$: The size of $A(X, u, D)$ can be predicted as:

$$E\left\{\left|A(x,u,\tilde{D})\right| \mid X_1 = x_1\right\} = \frac{1}{M}\sum_{j=1}^{M}\left|A(X_{2j},u,\tilde{D})\right|.$$

Cluster size of $A(X, u, D)$: For large u, let $S(X_{2j},\tilde{D})$ be the cluster size of X_{2j} on $\tilde{D}$, $j = 1,\ldots, M$.

Then the cluster size can be predicted as follows

$$E\left\{S(x,\tilde{D}) \mid X_1 = x_1\right\} = \frac{1}{M}\sum_{j=1}^{M} S(X_{2j},\tilde{D})$$

Number of clusters N: Let $\tilde{N}_j$ be the number of components of X_{2j} (t) on $\tilde{D}$. Then N can be predicted by:

$$E\{N \mid X_1 = x_1\} = \frac{1}{M}\sum_{j=1}^{M} \tilde{N}_j.$$

$\sup_{t\in\tilde{D}} X(t)$: The supremum of X(t) on D can be predicted by

$$E\left\{\sup_{t\in\tilde{D}} X(t) \mid X_1 = x_1\right\} = \frac{1}{M}\sum_{j=1}^{M} \max\{X_{2j}\}$$

PREDICTION INTERVALS

Based on a large number of realizations from $f(x_2|x_1)$, we can find a 95% prediction intervals for A(X, u, D), N, S(X,u) and $\sup_{t\in D}X(t)$.

The following algorithm is designed to find these prediction intervals.

Simulate $X_{2j}(t)$, j = 1,..., M, realizations from $f(x_2|x_1)$.

For each j, find $A(X_{2j},u,\tilde{D})$, the excursion set of X_{2j} on $\tilde{D}$.

For each excursion set in 2, find the clusters size, the number of clusters and the max $\{X_{2j}\}$.

For each characteristic you find in 3, the prediction interval is [L, U], where L and U are the 2.5% and the 97.5 percentiles of the empirical distribution.

ESTIMATION OF μ, σ^2 AND τ^2

Since the parameters μ, σ^2 and τ^2 are unknown and the predictive density $f(x_2|x_1)$ depends on these parameters, we plug their estimates in $f(x_2|x_1)$. The Maximum Likelihood Estimates (MLE's) of μ, σ^2 and τ^2 are the values $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\tau}^2$ which maximize the likelihood function

$$L(\mu,\sigma^2,\tau^2) = (2\pi)^{-\frac{n}{2}} \det(\Sigma_{11})^{-\frac{1}{2}} \exp\begin{pmatrix} -\frac{1}{2\sigma^2}\left(X_1 - \mu 1_n\right)^T \\ \Sigma_{11}^{-1}\left(X_1 - \mu 1_n\right)\end{pmatrix}$$

SIMULATION

We restrict our simulation to the case d = 1. Simulation of a Gaussian process in $D = [0, A]$ is equivalent to simulation a Gaussian vector on a grid of D. So, to simulate a stationary Gaussian process X(t), $t\in D = [0, A]$ with covariance function K(t,s), we follow the following steps:

Consider the grid

$$\tilde{D} = \{0 = t_1,...,t_B = A\}$$

Find the covariance matrix $\Sigma = \left(K(t_i,t_j)\right)_{i,j=1}^{B}$ and the mean vector $\mu = \mu 1_B$

Simulate a random vector of length B from $N_B(\mu, \Sigma)$

We consider the covariance function (2) and the value μ = 0, σ = 1 and τ = 1 to simulate a sample path of X(t) on the interval C = [0, 256]. Then the data is divided into two vectors X = (X_1, X_2) where X_1 represents the first 128 entries of X and X_2 the remaining 128 entries. So X_1 is considered as the observed data and X_2 as the reference data for our prediction. We use the theory developed in this paper to predict the characteristics of the excursion of X(t) in D = [129, 256]. The excursion set of the reference data X_2 has the observed characteristics: $A(X_{2j},2,\tilde{D}) = 4$, $S(X_{2j},\tilde{D}) = 1$, N= 4 and max X_2 = 2.5544. A large sample of M = 5000 realizations from $f(x_2|x_1)$ is simulated and 95% prediction intervals for these excursion set characteristics are obtained. We summarize the results in Table 1.

Table 1: Prediction interval for excursion set characteristics

Characteristics	$A(X,2,\tilde{D})$	$S(X,\tilde{D})$	N	$\sup_{t\in\tilde{D}} X(t)$
Prediction interval	[0, 8]	[0, 3]	[0,6]	[1.735, 3.462]

CONCLUSION

In this paper, we considered the problem of predicting the excursion set and its characteristics for a smooth and stationary Gaussian random field. We did not get closed forms for the predictors, but we obtained them based on a large sample from the predictive density. Simulation shows that the prediction intervals contain the observed characteristics.

REFERENCES

1. Adler, R., 1981. The geometry of random fields. John Wiley and Sons, New York.
2. Baecher, G.B. and J.T. Christian, 2003. Reliability and statistics in geotechnical engineering. Chichester, West Sussex, England.
3. Benkoski, S.J., M.G. Monticino and J.R. Weisinger, 1991. A survey of the search theory literature. Naval Research Logistics, pp: 38469-38494.

4. Degroot, D.J. and G.B. Baecher, 1993. Estimating auto-covariance of in-situ soil properties. J. Geotech. Eng., 119: 147-166.
5. Faber, M.H., D. Straub and M.A. Maes, 2006. A computational framework for risk assessment of RC structures using indicators. Computer-Aided Civil and Infrastracture Engineering, 21: 216-230.
6. Leadbetter, M.R. and G.V. Spaniolo, 2002. On statistics at level crossings by a stationary process. Statistica Neerlandica, 56: 152-164.
7. Straub, D. and M.H. Faber, 2005. Risk based inspection planning for structural systems. Structural Safety, 27: 335-355.