



Doctoral Thesis

## Estimating temporal change of soil properties

**Author(s):**

Papritz, Andreas Jürg

**Publication Date:**

1993

**Permanent Link:**

<https://doi.org/10.3929/ethz-a-000922284> →

**Rights / License:**

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH No. 10233

# Estimating Temporal Change of Soil Properties

A dissertation submitted to the  
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH  
for the degree of  
DOCTOR OF NATURAL SCIENCES

presented by

ANDREAS JÜRG PAPRITZ  
dipl. Natw. ETH  
born February 5, 1960  
citizen of Albligen BE

accepted on the recommendation of  
Prof. Dr. Hannes Flühler, examiner  
Prof. Dr. Hans Rudolf Künsch, co-examiner  
Prof. Dr. Richard Webster, co-examiner

1993

# Kurzfassung

Diese Arbeit vergleicht statistische Verfahren zur Schätzung zeitlicher Veränderungen von Bodeneigenschaften. In der Schweiz bestehen sowohl ein nationales wie auch kantonale Netze von Beobachtungsflächen, auf welchen die Schadstoffbelastung des Bodens periodisch erfasst wird. Aus diesen Daten müssen Schlüsse über die Veränderung der Belastung gezogen werden. Diese Fragestellung verlangt nach generalisierten Least-Squares-Schätzverfahren zur Bestimmung der Parameter eines statistischen Modells, weil die Zufallsterme im Modell als statistisch abhängig vorausgesetzt werden müssen. Ungenügende Kenntnisse über typische Autokorrelationsmuster von Bodeneigenschaften unter einer bestimmten Landnutzungsart verhindern aber vorderhand eine Anwendung dieser Methode, und Schlüsse über generelle Veränderungen müssen aus einem einfachen Modell abgeleitet werden, welches voraussetzt, dass die Schätzwerte von einzelnen Beobachtungsflächen statistisch unabhängig sind. Die Fragestellung reduziert sich somit darauf, verschiedene Verfahren zur Schätzung von mittleren Differenzen über Beobachtungsflächen zu vergleichen.

Sowohl modellbezogene, geostatistische Methoden wie auch klassische Stichprobenpläne können dazu verwendet werden. Im modellbezogenen Ansatz wird ein Datensatz, der aus wiederholten Messungen besteht, als eine Realisierung mehrerer kreuzkorrelierter Zufallsprozesse betrachtet. Mit diesem Modell wurden Cokriging Gleichungen zur Schätzung der zeitlichen Differenzen hergeleitet. Die Methode kann zur Schätzung von Punktwerten und zur Bestimmung von lokalen Mittelwerten verwendet werden, und sie weist unter den linearen Verfahren den kleinsten, mittleren quadrierten Fehler auf.

Voraussetzung für die Anwendung dieser Methode ist die Kenntnis

der Variogramme oder der Kovarianzfunktionen der Zufallsprozesse. Das Kreuzvariogramm kann aber bei wiederholter Beprobung des Bodens nicht mit der Momentenmethode geschätzt werden, weil für jeden Punkt nur eine Messung zur Verfügung steht. Sind die Variogramme beschränkt, so können an ihrer Stelle die Kovarianzfunktionen verwendet werden, deren Schätzung immer möglich ist. Sind sie aber unbeschränkt, so muss entweder das Pseudokreuzvariogramm benutzt werden, oder das normale Kreuzvariogramm wird mit der vorgeschlagenen Least-Squares Methode bestimmt. Eine Analyse der Stationaritätsbedingungen zeigte, dass das Pseudokreuzvariogramm nur für Zufallsprozesse existiert, deren Differenzen in zweiter Ordnung stationär sind. Diese Bedingung ist strenger als diejenige, welche für die Existenz des normalen Kreuzvariogrammes notwendig ist. In der Praxis kann das Pseudokreuzvariogramm nur dann verwendet werden, wenn die Autovariogramme beschränkt sind oder wenn sie für grosse Lagdistanzen mit der gleichen Rate zunehmen.

Die geostatistische Methode kann auch zur Schätzung einer globalen mittleren Veränderung, d.h. der gemittelten Differenz über eine ganze Untersuchungsfläche eingesetzt werden. Dazu wird das Untersuchungsgebiet in kleine Parzellen unterteilt, und deren Mittelwerte werden mittels Co-kriging geschätzt. Der globale Mittelwert wird dann als gewichtetes Mittel der lokalen Schätzwerte berechnet. Eine theoretische Analyse zeigte, dass die geostatistische Methode bei Mittelung über viele Realisierungen eines Zufallsprozesses für jede beliebige räumliche Anordnung der Messpunkte mindestens so präzise wie die klassischen Stichprobenpläne sein sollte. In der Praxis ist dieses Optimalitätskriterium aber nicht in jedem Fall von Bedeutung, weil üblicherweise nur eine Realisierung vorliegt. In diesem Fall ist es wichtiger zu wissen, welche Schätzmethode zum kleinsten Schätzfehler führt, wenn die gleiche Realisierung wiederholt beprobt wird.

Zur Untersuchung dieses Problems wurden eine Realisierung zweier grosser, auto- und kreuzkorrelierter Zufallfelder generiert. Der Simulation lag ein statistisches Modell zugrunde, das sich aus einer Untersuchung über die zeitliche Veränderung des pH-Werts in einem Waldboden ableiten liess. Die geostatistische Methode zur Schätzung der globalen mittleren Differenz wurde mit geschichteten und systematischen Zufallsstichproben verglichen. Bei den klassischen Stichprobenplänen wurde zudem der Effekt untersucht,

der sich aus der Paarung der Probenahmestellen zu den beiden Zeitpunkten ergibt. Jedes Schätzverfahren wurde mit zufälliger Wahl der Stichprobenpunkte 300 mal wiederholt. Das geostatistische Verfahren war weniger präzise als gepaarte systematische und gepaarte stratifizierte Stichproben. Die kleinere Effizienz war vor allem durch den zusätzlichen Aufwand bedingt, der zur Schätzung der Variogramme notwendig ist. Die Paarung der Probenahmestellen vergrößerte die Präzision von geschichteten Stichproben deutlich: gepaarte stratifizierte waren praktisch ebenso präzise wie gepaarte systematische Zufallsstichproben. Bei fehlender Paarung war die letztere Methode aber deutlich präziser. Für kleine Stichprobenumfänge sind gepaarte geschichtete Zufallsstichproben systematischen vorzuziehen. Letztere sind zwar etwas effizienter, sie erlauben aber nicht, die Varianz der mittleren Differenz aus den Werten einer Stichprobe erwartungstreu zu schätzen.

Die statistische Analyse der zeitlichen Veränderung der pH-Werte bestätigte im wesentlichen diese Befunde. Eine Abschätzung der Erwartungswerte und der Varianzen der Schätzfehler unter dem Modell der Zufallsprozesse legte zudem den Schluss nahe, dass sich die gleichen Resultate auch für weitere Realisierungen der Zufallsfelder finden liessen.

# Abstract

The thesis compares statistical methods for estimating temporal change of soil properties. In Switzerland networks of sites for monitoring the contamination of soil have been established both by federal and cantonal authorities. Soil is repeatedly surveyed on these plots, and the data on temporal change are used to assess overall trends of soil contamination in the country. Determining the parameters of a statistical model asks for generalized least-squares estimation because one has to assume that the random terms of the model are spatially autocorrelated. However, the information on typical patterns of spatial correlation found for soil under different land use is still too limited to follow that approach. The general development must therefore be assessed with a simple model that assumes statistical independence of the estimates measured on different plots. Inference can then focus on estimating mean temporal differences on monitoring plots.

Both the model-based geostatistical method and the classical random sampling schemes can be used for this purpose. In the model-based approach the observations measured on different occasions are regarded as a realization of a set of auto- and cross correlated random processes. By using this model the cokriging equations for estimating temporal differences at unsampled points were derived. The extension to block cokriging enables us to estimate local mean differences on small pieces of land. This method is optimal in the sense that among linear estimators it has minimal mean squared estimation error.

The variograms or the covariance functions must be known for kriging. The experimental cross variogram between times, however, cannot be computed using the method-of-moments estimator because for each point only

one value is known. If variation is bounded then the covariance functions can be used instead, and these can be estimated from any configuration of sampling points. If the autovariograms are unbounded then either the pseudo cross variogram must be used or the normal cross variogram is estimated using the suggested least-squares method. A theoretical analysis of the stationarity condition required for the existence of the pseudo cross variogram revealed that this function exists only if the differences between two random processes are second order stationary. This is more restrictive than the condition necessary for the existence of the normal cross variogram. In practice we can use the pseudo cross variogram if the autovariograms are either bounded or if they all increase at the same rate for large lag distances.

The geostatistical method can also be used for estimating global mean change over the whole monitoring plot. The plot is then subdivided into small blocks, the mean changes of these blocks are cokriged locally, and the global mean is obtained by combining the local estimates. A theoretical analysis showed that the model-based estimate should be at least as precise as any of the classical methods when repeatedly applied to many realizations of a random process. In practice, however, this criterion is hardly meaningful where one has a single realization only. It is more relevant to know which method is most precise when repeatedly applied to the same realization.

I studied the matter by generating two large fields of auto- and cross correlated data. In the simulation a statistical model similar to that found in a case study on short-term change of pH in a forest soil was used. Geostatistical global estimation was compared with stratified and systematic random sampling, and the effect of pairing the sampling locations on the two occasions was explored for the classical schemes. For each method 300 pairs of samples were randomly drawn from the fields. The geostatistical method was less precise than paired stratified and paired systematic sampling. The poorer precision was caused mainly by the effort necessary for estimating the variograms at short lag distances. Pairing the locations decreased the sampling error for both stratified and systematic designs, but the gain in efficiency was larger for stratified random sampling. This plan was nearly as precise as paired systematic sampling. For small sample

sizes paired stratified samples with two observations per stratum are to be preferred to paired systematic schemes because they are both precise and predict the sampling error accurately.

The analysis of the pH data essentially confirmed these results. Computing the expectations and the variances of the sampling errors under the random process model further suggested that the same results are likely to be found if other realizations of the same random processes are investigated.