research for
man and environment

RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEU
NATIONAL INSTITUTE OF PUBLIC HEALTH AND THE ENVIRONMENT

RIVM report 408657 003

Geostatistical interpolation of abiotic site conditions in the Netherlands

A method for reference mapping

A.M.F. Bio [1], R. Alkemade [2], A. Barendregt [1], J. Wiertz [2]

September 1999

[1] *Department of Environmental Studies, Utrecht University, P.O. Box 80115, 3508TC Utrecht, The Netherlands.*
[2] *Soil and Groundwater Research Laboratory, National Institute of Public Health and the Environment, P.O. Box 1, 3720 BA Bilthoven, The Netherlands.*

This investigation has been performed on request and for the account of the Dutch National Institute of Public Health and the Environment, within the framework of project 712910/408576, 'Environment and Nature'.

RIVM, P.O. Box 1, 3720 BA Bilthoven, telephone: 31 - 30 - 274 91 11; telefax: 31 - 30 - 274 29 71

Abstract

It is of great importance for environmental research and policy making to have a country-wide picture of abiotic conditions and their changes with time. For this reason available geographical and time-referenced data have to be interpolated to unsampled regions. Here, geostatistical interpolation methods are proposed and successfully applied for the production of abiotic maps on a national scale.

We used local ordinary block kriging to interpolate national abiotic maps for moisture, acidity/alkalinity, nitrogen and salinity conditions. This method allows us to produce unbiased estimates and their respective estimation errors on different geographical scales. Prior to interpolation, data were selected from two time periods and stratified on the basis of the dominant soil types underlying the vegetation records. The resulting maps were compared to other available information and discrepancies analysed.

Overall, our results agree with other abiotic maps and expert knowledge. Inaccuracies found in the interpolation results were caused mainly by sampling (design), stratification and the assumptions chosen for the interpolation. Given the methodological focus of this study, we suggest further detailed examination of the maps presented, and correction of obvious errors in the maps prior to applying them for reference and making inferences on the results, especially comparison of abiotic site conditions in time.

Contents

List of Figures

List of Tables

Samenvatting

Een landsdekkend beeld van standplaatsfactoren in Nederland voor verschillende tijdperioden is van groot belang voor vergelijkend onderzoek en beleidsvorming. Omdat een landsdekkende bemonstering en analyse van standplaatsfactoren in Nederland voor het verleden ontbreekt en voor het heden te kostbaar is, zijn we op de interpolatie van de beschikbare gegevens aangewezen om een schatting van deze factoren voor niet bemonsterde gebieden te verkrijgen.

Geostatistische analyse- en interpolatiemethodes zijn uitermate geschikt voor de interpolatie van ruimtelijke gegevens. In dit rapport passen wij ze toe om een landsdekkend beeld van de vocht, zuurgraad, nutriënten en saliniteit status van Nederland te verkrijgen. Als gegevens worden uit vegetatieopnamen afgeleide Ellenberg F, R, N en S indicatiewaarden gebruikt. Voor de ruimtelijke analyse en interpolatie zijn gegevens uit twee tijd perioden geselecteerd—een recente periode, tussen 1990 en 1997, en een historische periode tussen 1930 en 1970. Deze zijn vervolgens gestratificeerd in acht bodemtypen—zeven bodemsoorten en aquatische gebieden. De interpolatie is gebaseerd op de ruimtelijke afhankelijkheid in elk stratum. Strata worden derhalve apart geanalyseerd en geïnterpoleerd. De resultaten zijn samengevoegd in landsdekkende kaarten, met één kaartbeeld per tijd periode en standplaats factor. De gepresenteerde kaartbeelden hebben een ruimtelijke resolutie van 250×250 m^2 voor de recente tijdstap en 1000×1000 m^2 voor de historische periode. Naast de interpolatieresultaten worden ook de interpolatiefouten gepresenteerd.

De kaarten geven een goed algemeen beeld van de Nederlandse situatie. Ze stemmen redelijk overeen met gegevens uit andere ecologische bronnen. Het beste blijken de kaarten voor vocht, zuurgraad en nutriënten, die ook de meest belangrijk zijn in het kader van de verdrogings-, verzurings- en vermestingsproblematiek. Er zijn ook afwijkingen van de verwachte resultaten. Fouten blijken voornamelijk veroorzaakt te worden tijdens de bemonstering, stratificatie, ruimtelijke analyse en modellering, en de interpolatie.

De focus van deze studie ligt bij het toepassen en vastleggen van geostatistische methodes voor de interpolatie van standplaatsfactoren. De verkregen kaartbeelden zijn slechts een eerste stap in de ontwikkeling van referentie kaarten. Vóór het gebruik van deze kaartbeelden moet er nog een gedetailleerde analyse en, waar nodig, correctie plaats vinden.

Summary

It is of great importance for environmental research and policy making to have a country-wide picture of abiotic conditions and their changes with time. Because sampling on a national scale is economically unfeasible, available geographical and time-referenced data have to be interpolated to estimate site conditions for unsampled regions.

Geostatistical analysis and interpolation methods are particularly suited for the interpolation of spatial data. Here, we have used these methods to depict moisture, acidity/alkalinity, nitrogen and salinity conditions in the Netherlands. The abiotic data are represented by Ellenberg F, R, N and S values, averaged from plant species present in field samples. Prior to analysis and interpolation, data were selected from two time periods—a recent one dating from 1990 to 1997 and a historical one from 1930 to 1970. The data were stratified on the basis of eight dominant soil types underlying the vegetation records—seven soil types and aquatic regions. Interpolation is based on the spatial dependence found in each stratum. Strata are thus individually analysed and interpolated prior to combining the results into national maps (one map for each time period and stratum). Current and historical abiotic situations in the Netherlands are presented as grid maps of 250×250 m^2 and 1000×1000 m^2 resolution, respectively. Besides the interpolation results, interpolation errors are also presented .

The resulting maps give a good idea of the Dutch situation. Overall, our results agree with other abiotic maps and expert knowledge. There are, however, inaccuracies in the interpolation results, which are caused mainly by sampling (design), stratification and the assumptions chosen for the interpolation.

The results of this study, focused on application and description of geostatistical methods for the interpolation of site conditions, represent a mere first step in the development of reference maps. We suggest further detailed examination of the maps presented, and correction of obvious errors in the maps prior to their further application.

1. Introduction

To better assess, and set targets and define limits for future nature conservation and development, policy makers need proper reference for historical and actual environmental conditions in their regions or countries. Environmental measurments on a national scale are economically unfeasible. Therfore, available geographically and time referenced data have to be interpolated to estimate site conditions for unsampled regions.

Interpolation of site conditions could take place on the basis of environmental records considered representative for the area and period of interest. However, samples with actually measured environmental variables are generally scarce; especially for these *past times* so often chosen as desired reference conditions for nature restoration. Historical, and also the most actual, records are often a mere collection of species, without measured abiotic site conditions. In the absence of measured abiotic variables, it is possible to estimate likely abiotic site conditions from the plant species present in samples. Ellenberg attributed indicator values to numerous plant species of Central Europe, characterizing the ecological conditions under which these species usually occur (Ellenberg *et al.* 1991). These indicator values can be used as indirect semi-quantitative estimates of the abiotic site conditions (Wittig *et al.* 1985; Briemle 1986; Novákovå 1997; Ertsen *et al.* 1998).

In our study we will propose and apply geostatistical methods to interpolate abiotic reference maps of actual and historical abiotic environmental conditions in the Netherlands, on the basis of a vast collection of vegetation records. We will estimate moisture, acidity, trophic level and salinity conditions by averaging the Ellenberg indicator values of the species present at each sample site. These averaged values will be subsequently treated as abiotic site conditions and interpolated to the whole of the country (Ter Braak & Barendregt 1986; Ter Braak & Gremmen 1987; Ter Braak & Wiertz 1994).

The method proposed for the spatial interpolation of reference maps on a national scale, is called *local ordinary block kriging*. We opted for this geostatistical method because it uses the spatial dependence, or autocorrelation, structure between samples in an optimal way, producing unbiased interpolation estimates and variances for unsampled sites. Environmentally, sites that are close together are generally more similar than those further apart. Attributing weights to samples in the neighbourhood dependent on their distance from the unsampled site will hence improve our estimate for the latter. Methods, like inverse distance weighting are limited in the choice of weights used. Furthermore, they do not correct for the spatial dependence between samples themselves. An unsampled site, located, for instance, in between a single point in the east and a cluster of points in the west, will be unduly influenced by the cluster, as all of its samples are weighted as independent samples. Kriging not only allows for a variety of shapes of spatial dependency, but it also weights samples based on the distance between them as well as from the unsampled site. Weights are chosen according to a model of the spatial dependence between pairs of samples—the *semivariogram.*

Kriging is a smooth interpolation method, ideally suited for continuous phenomena. It should, therefore, only be applied to a reasonably homogeneous set of data, without local differences in their spatial dependence structure (i.e. satisfying a certain *stationarity* criterion). Considering all nationwide samples—even from a single time period—as belonging to one homogeneous population (or *random variable*, as considered in geostatistics) would be very unrealistic and in contrast to the fragmented nature of the Dutch landscape. Therefore, to distinguish between clearly different regions, like aquatic and terrestrial areas, and to better satisfy the stationarity requirements of the interpolation method (stating that the spatial dependence between data pairs is merely a function of their distance), data are stratified into one aquatic and seven soil classes prior to geostatistical modelling and interpolation.

Spatial analysis and interpolation are done separately for each of the eight strata, the four abiotic variables—moisture, acidity, nitrogen and salinity—and two distinct time periods. Samples collected from 1991 to 1997 are used to map the *actual* abiotic situation, samples collected from 1930 to 1970 for an *historical* situation. The results obtained per stratum are combined into national grid maps, one for each time period and site condition. The spatial resolution is $250 \times 250\ m^2$ for the more recent time-period maps and $1000 \times 1000\ m^2$ for the historical maps.

2. Data

2.1 Data pre-processing

Data were supplied in 53 DBASE databases (Appendix I.1; Schaminée *et al.* 1995-1999). These were combined into one total database (169,202 records) for further analysis and processing. Given the aim of the present study—the production of abiotic reference maps for determined time intervals on a grid of one square kilometre or finer—only records with accurate time (sample year) and location references ('Amersfoort' and/or 'Atlasblok' co-ordinates on the required scale) could be used. Of all the 169,202 records, and after some calculations and corrections (Appendix I.2), 129,092 were found to meet these criteria.

Some new attributes are deduced from the original ones. The sample year is extracted from the sample date, the habitat from the ecotype, and new co-ordinates (in metres) from 'Atlasblok' and 'Amersfoort' co-ordinates.

The available data differ in precision of spatial reference. For recent observations most samples are recorded on tenth kilometre co-ordinates. Historical data are usually recorded per square kilometre. Interpolation should ideally take place on the sample scale. Hence, we opted for a 250×250 m^2 grid for the most recent time period, and for a 1000×1000 m^2 grid for the historical period.

Based on the availability and quality of the data, two basic working sets were composed:

actual data	39,389 samples collected from 1991 to 1997 location specified at tenth kilometre precision fields: REL, YEAR, XCOOR, YCOOR, ECO, HAB, SOIL, Fnr, FM, Rnr, RM, Nnr, NM, Snr, SM
historical data	16,974 samples collected from 1930 to 1970 location specified at (at least) one kilometre precision fields: REL, YEAR, XCOOR, YCOOR, ECO, HAB, SOIL, Fnr, FM, Rnr, RM, Nnr, NM, Snr, SM

Where:
REL is the sample IDnumber; YEAR the sample year; XCOOR 'Amersfoort' co-ordinates in east-west and YCOOR in north-south direction (in metres); ECO is ecotype (see Appendix I.3); HAB the habitat (W = aquatic; T = terrestrial); SOIL the attributed soil class (0 = aquatic, 1 = sand poor, 2 = sand rich, 3 = sand calcareous, 4 = clay non-calcareous, 5 = clay calcareous, 6 = loam non-calcareous, 7 = peat and 8 = urban areas; see Appendix I.6 for the soil map used). Fnr is the number of species with Ellenberg F indicator value present in the sample; FM, the averaged indicator value for moisture for the sample; Rnr & RM, the same for acidity/alcalinity; Nnr & NM, for nitrogen; and, Snr & SM, for salinity. Estimated Ellenberg values were obtained for each site by averaging the indicator values of the vegetation present (Ter Braak & Barendregt 1986; Ter Braak & Gremmen 1987; Ellenberg *et al.* 1991), with the exception of large shrubs, trees and species indifferent to the factors considered (i.e. classified as × in Ellenberg *et al.* 1991). For the present study only sites with indicator values averaged from two or more species were considered for spatial analysis and interpolation.

Sample density differs geographically and in time. In Table 2.1, the number of samples per square kilometre is summarized for the twelve Dutch provinces and the two time periods considered. Sample distribution per time period, abiotic variable and stratum is summarized in Appendix I.7. Sample distribution along the abiotic gradients is normal to skewed. Skewness is most accentuated for salinity, with few to no data in the brackish and saline region.

***Table 2.1** Sample densities for the actual and historical data sets per province*

Province	**Area (km^2)**	**Samples × km^{-2}**	
		Actual (1991 to 1997)	**Historical** (1930 to 1970)
Groningen	2390	0.04	0.25
Friesland	3540	0.47	0.56
Drenthe	2680	0.16	0.37
Overijssel	3410	0.35	0.59
Gelderland	5130	0.60	3.56
Utrecht	1430	1.26	0.39
Noord-Holland	2820	0.81	0.84
Zuid-Holland	3070	0.62	1.08
Zeeland	1820	0.97	0.76
Noord-Brabant	5040	0.28	0.91
Limburg	2200	0.43	0.92
Flevoland	1470	0.05	0.08
Total	35,000		
Average		0.50	0.86

2.2 Actual data

The set of *actual data*, i.e. records collected from 1991 to 1997, contains 39,389 samples (Appendix I.4). According to the species present, each sample was assigned to one of seven ecological groups (coded in the ecotype ECO; see Appendix I.3): forest, grassland, pioneer vegetation, rough grassland, shrubs, waterside and water. The relative composition of the data is presented in Figure 2.1.

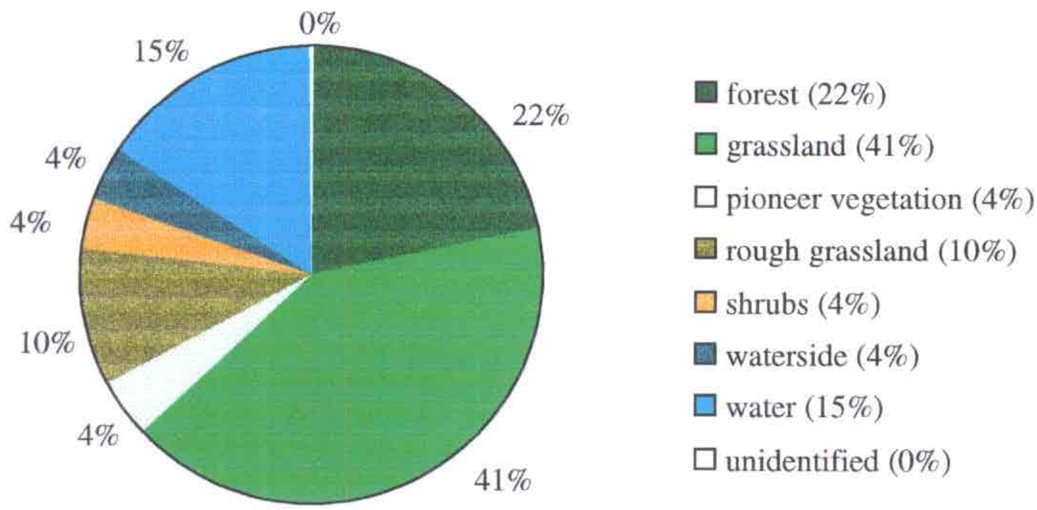

Figure 2.1. *Proportion of ecological groups for samples collected from* ***1991*** *to* ***1997****.*

Data stratification for spatial analysis and interpolation was based on the classes of the Dutch soil map (Appendix I.6): sand poor, sand rich, sand calcareous, clay non-calcareous, clay calcareous, loam non-calcareous, peat, aquatic and urban areas. Each sample site location was attributed the underlying soil or cover type. The distribution of soil types, aquatic and urban areas in the sample is shown in Figure 2.2.

Note, that the soil map displays the dominant soil type on a $250 \times 250\ m^2$ grid. Consequently, samples do not always correspond to the underlying soil type grid. It is impossible to correct for all inaccuracies, solely on the basis of the vegetation present. We did, however, avoid obviously faulty associations between aquatic and terrestrial systems, considering only terrestrial samples (forest, grassland, pioneer vegetation, rough grassland, shrubs) for the spatial analysis and interpolation for the terrestrial strata (sand, clay, loam and peat), and all aquatic samples (waterside and water) for the aquatic stratum. Urban and unidentified samples were disregarded. The *actual* data set contains 31,727 terrestrial and 7571 aquatic samples.

Abiotic site conditions of the *actual* data set, i.e. Ellenberg indicator values averaged from two or more species for each record are summarized in Table 2.2.

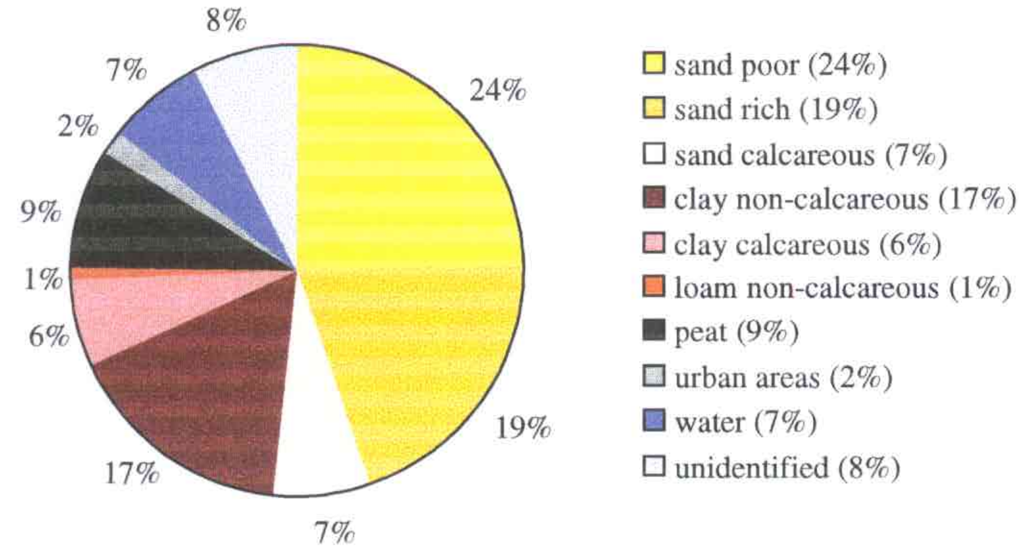

Figure 2.2. *Proportion of soil types, aquatic and urban areas for the samples collected from* ***1991*** *to* *1997.*

Table 2.2 *Summary of the averaged Ellenberg indicator values in the actual time period data*

	Ellenberg Indicator values			
	F	R	N	S
minimum	2.0	1.0	1.0	0.0
mean	7.2	5.6	5.3	0.4
maximum	12.0	8.7	8.5	9.0
number of records*	37,089	36,571	37,484	36,791
number of missing values**	2300	3818	1902	2598

* The number of records for each abiotic variable depends on the number of missing values.

** Samples with less than two species with respective indicator value.

2.3 Historical data

The *historical data* set contains 16,974 samples collected from 1930 to 1970 (Appendix I.5). Based on ecological groups, the relative composition of the sample is presented in Figure 2.3.

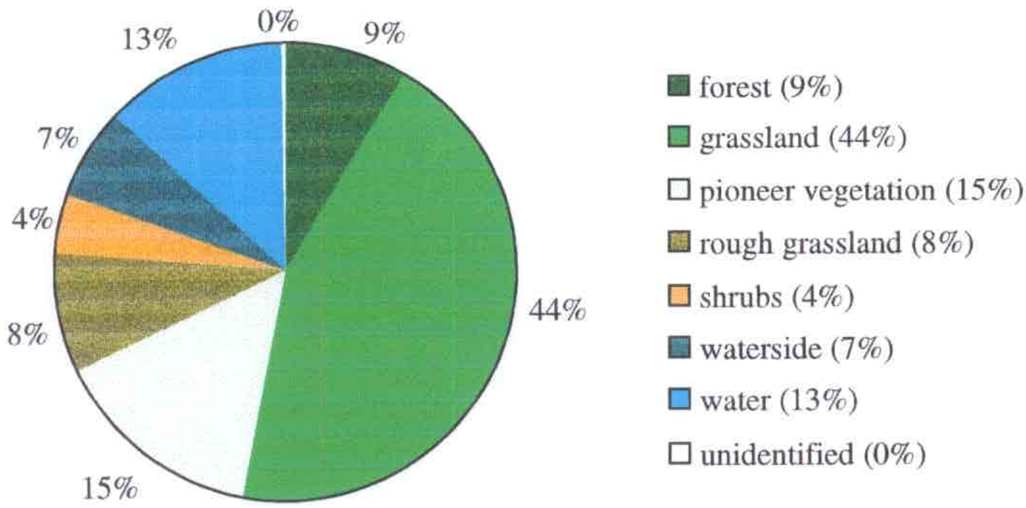

Figure 2.3. *Proportion of ecological groups for samples collected from **1930** to **1970**.*

Since most historical data have their sample locations defined on a kilometre scale, we opted for a two-step approach for stratification into soil types. Analogous to the *actual* data set, each sample with location defined at the tenth kilometre was assigned the respective soil type in the 250×250 m^2 grid map (Appendix I.6). For these 2278 samples the distribution of underlying soil types, aquatic and urban areas is shown in Figure 2.4.

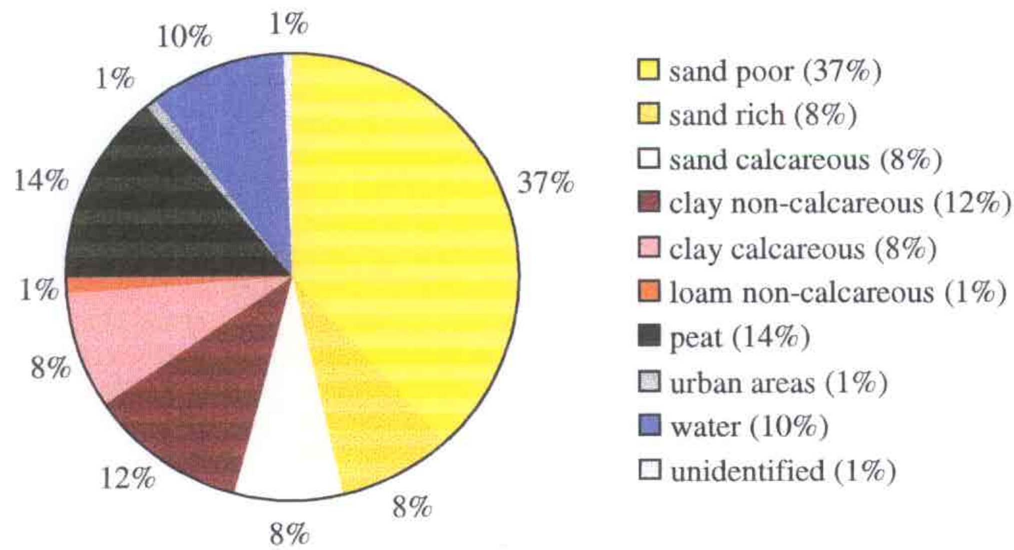

Figure 2.4. *Proportion of soil types, aquatic and urban areas for the samples with location defined at tenth kilometre, collected from **1930** to **1970**.*

Samples referring to square kilometre co-ordinates only (14,696 records), were attributed all soil types (terrestrial *or* aquatic) present in that respective square kilometre. The same sample can thus be assigned up to seven different classes. The relative occurrence of different soil types, aquatic and urban areas on a square kilometre grid is shown in Figure 2.5.

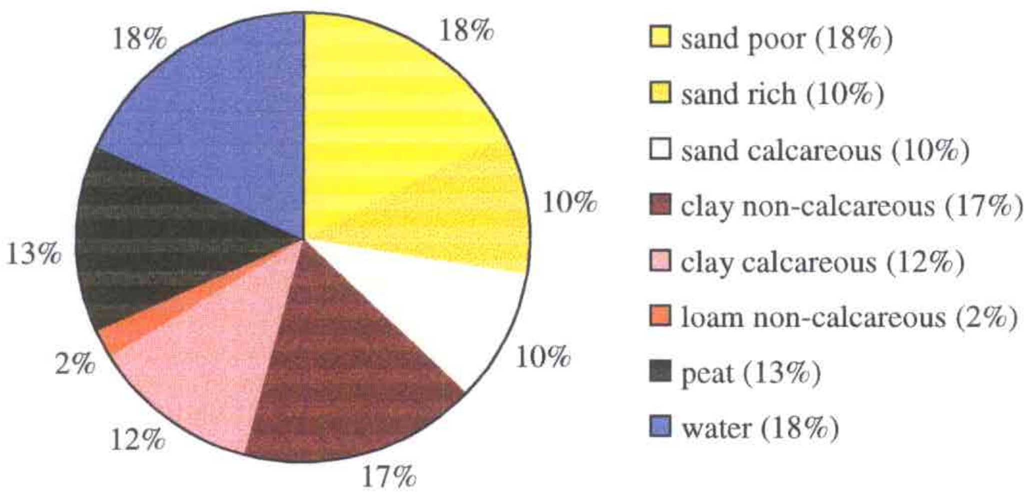

Figure 2.5. *Proportion of soil types, aquatic and urban areas for the samples with location defined at square kilometre only, collected from* ***1930*** *to* ***1970****.*

On a square kilometre resolution, the risk that samples will not conform to the attributed soil type is substantial. Yet, just as with the actual data, the only possible restriction was the distinction between terrestrial and aquatic habitats. Terrestrial samples were used exclusively for spatial analysis and interpolation for the terrestrial strata, and water and waterside samples for the aquatic stratum. Historical data were subdivided into 13,666 terrestrial and 3260 aquatic samples; 48 were undefined in their habitat. There was no surface unit (km^2) that was exclusively urban or unidentified; these classes were disregarded.

Abiotic site conditions of the *historical* data set, i.e. Ellenberg indicator values averaged from two or more species for each record, are summarized in Table 2.3.

Table 2.3 *Summary of the averaged Ellenberg indicator values in the historical time period data*

	Ellenberg Indicator values			
	F	R	N	S
minimum	2.3	1.0	1.0	0.0
mean	7.4	5.8	4.9	1.3
maximum	12.0	8.7	9.0	8.7
number of records*	16484	16387	16645	16674
number of missing values**	490	587	329	300

* The number of records for each abiotic variable depends on the amount of missing values.

** Samples with less than two species with respective indicator value.

3. Methods

3.1 3.1. Spatial analysis

To perform spatial interpolation of our estimated abiotic variables, we treated the averaged Ellenberg indicator values as a continuous variable. Considering the (greatly varying) sample size, which is usually small in comparison to the support (area) for interpolation, our samples are point-referenced data. The methods used in the present study originate from geostatistics and are based on the *regionalized variable theory* (Matheron 1971). In geostatistics, a set of observations, $z(\mathbf{s}_i)$, at known spatial locations, $\{\mathbf{s}_1, ..., \mathbf{s}_n\}$, is regarded as a sample from a potentially infinite number of measurements, $\{z(\mathbf{s}): \mathbf{s} \in D\}$, which could have been taken throughout the sample region D. The collection of potential measurements is in itself assumed to be a realization of the random field, or stochastic process, $\{Z(\mathbf{s}): \mathbf{s} \in \boldsymbol{D}\}$.

We will assume that the spatial variation of our abiotic variable can be expressed as the sum of two major components: a structural deterministic function, $m(\mathbf{s})$, the trend; and a random function, $\delta(\mathbf{s})$ (Cressie 1993; Burrough & McDonnell 1998). The value of a random variable Z at location $\mathbf{s}$ is modelled as

$$Z(\mathbf{s}) = m(\mathbf{s}) + \delta(\mathbf{s}) \quad (1)$$

According to the above-mentioned theory, our estimated abiotic site conditions originate from a single realization of the random field. To analyse their spatial dependence, we will have to impose some structure on the random field to compensate for the lack of replicates.

We will consider the spatial correlation of the regionalized variable $\delta(\mathbf{s})$ to be merely a function of separation distance, h, between pairs of data, independent of their orientation. We will model exclusively isotropic semivariograms, displaying the average spatial structure in all directions. Furthermore, we will assume $m(\mathbf{s})$ to be spatially constant. The component $\delta(\mathbf{s})$ is, then, a zero-mean intrinsically stationary random process (Cressie 1993) and the variance of differences between random variable pairs at distances h from each other, $\mathbf{s}$ and $\mathbf{s}+h$, will be

$$\mathrm{Var}[Z(\mathbf{s}) - Z(\mathbf{s}+h)] = \mathrm{E}\{[Z(\mathbf{s}) - Z(\mathbf{s}+h)]^2\} = \mathrm{E}\{[\delta(\mathbf{s}) - \delta(\mathbf{s}+h)]^2\} = 2\gamma(h) \quad (2)$$

where $2\gamma(h)$ is the *variogram*, which is a function of only the distance (or *lag*) h. The *semivariance* for points separated by distance h, is $\gamma(h)$. Under the conditions mentioned above, $\gamma(h)$ can be estimated from observations $z(\mathbf{s}_i)$ by

$$\hat{\gamma}(\bar{h}) = \frac{1}{2N_h}\sum_{i=1}^{N_h} [z(\mathbf{s}_i) - z(\mathbf{s}_i+h)]^2 \quad (3)$$

where $\bar{h}$ is the average distance of the respective lag and N_h the number of data pairs separated by the distance h.

A plot of the semivariances, $\hat{\gamma}(h)$, against the increments, h, the *sample* (or *empirical*) *semivariogram*, provides information about the spatial dependence between data pairs with increasing distance between them. The sample semivariogram can be modelled by fitting a theoretical semivariogram to the semivariances. The estimator used for the semivariogram is optimal, when the random variable is normally distributed. Departures from normality and outliers will influence estimation (Haining 1990). Variable distribution should ideally be normal or, less ideally, at least symmetrical.

In order to use the spatial structure displayed in the empirical semivariogram for spatial interpolation, a (parametric) model is fitted to the semivariances. There is a suit of useful semivariogram models (Cf. Kitanidis 1983; Haining 1990) characterized by their shape and behaviour near zero lag. From the definition of the semivariogram, the semivariance is zero for a zero distance.

Sample variograms, however, generally display a positive value at the origin (Köhl & Gertner 1997); a discontinuity called the *nugget effect* (C_0), which is attributed to the sum of residual, spatially uncorrelated noise, such as measurement error and spatial variation on spatial scales smaller than the smallest sample spacing.

For a stationary process, the semivariogram is bounded; i.e. the semivariance increases with increasing lag distance till it reaches a plateau, the *sill*. The distance, at which the sill is reached, the *range* (a), indicates the maximum distance of spatial dependence between sample pairs. The most commonly used model of this type is the *spherical* model. Some variogram models, like the *exponential* and the *Gaussian* ones, are weakly non-stationary, increasing asymptotically towards a sill. Others, e.g. the *linear* model, are strongly non-stationary.

The choice of semivariogram model was done *ad hoc*, dependent on the apparent distribution of empirical semivariances. The following semivariogram models were used:

Spherical model with nugget effect

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ C_0 + C_1\left[\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right] & \text{for } 0 < h \leq a \\ C_0 + C_1 & \text{for } h > a \end{cases} \qquad (4)$$

Gaussian model with nugget effect

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ C_0 + C_1\left[1 - \exp\left(-\frac{h}{a}\right)^2\right] & \text{for } h > 0 \end{cases} \qquad (5)$$

Exponential model with nugget effect

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ C_0 + C_1\left[1 - \exp\left(-\frac{h}{a}\right)\right] & \text{for } h > 0 \end{cases} \qquad (6)$$

Linear model with nugget effect

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ C_0 + m \times h & \text{for } h > 0 \end{cases} \qquad (7)$$

where h is the lag, a the range, C_1 the partial sill (or spatial variance) and C_0 the nugget of the fitted semivariogram model, with sill = $C_0 + C_1$. Gaussian and exponential models are asymptotic with an approximate sill and range. The linear model is defined by its intercept at zero lag, b, and the slope m (Pebesma 1997).

Nugget models were avoided, as they imply absolute spatial independence between samples, even at near zero distances—which is unrealistic for abiotic site variables. At zero distance the semivariogram tends towards a value equivalent to uncorrelated noise. This value is essentially equal to the *measurement error* or, in our case, the estimation error of the averaged Ellenberg values. In practice, sample pairs at very small distances from each other are scarce, hindering an accurate estimate of the nugget. The nugget of our empirical semivariogram models will therefore contain variance at small distances next to the estimation errors.

Empirical semivariograms tend to fluctuate at greater lag distances. This is due to the correlation caused by repeated use of the same samples for successive lags. A rule of thumb is to consider distances up to half the maximum distance between samples for variogram analysis (Journel & Huijbregts 1978). We opted for a more local approach, restricting spatial analysis to a maximum distance of 40 km between sample pairs. Stratification was based exclusively on the soil or land-cover class. Yet abiotic values can differ due to other factors. Reducing the range of the neighbourhood used for interpolation will reduce the effect of possible trends and non-stationarities not distinguished in our stratification (e.g. in North-South direction).

Semivariograms were calculated for the four abiotic variables for each stratum, using Gstat (Pebesma 1997; Pebesma & Wesseling 1998). A semivariogram model was fitted for each empirical semivariogram using weighted least squares, with weights dependent on the number of data pairs per lag. Each semivariogram was plotted and summarized. An example of GSTAT command files and an estimates file is given in Appendix II.1.

3.2 Geostatistical interpolation

Abiotic site conditions were interpolated for unsampled sites by local ordinary block kriging. Interpolation is done, on the basis of the values of samples within the considered neighbourhood (a radius of 40 km) with weights determined by the respective semivariogram model. For each stratum, the trend is constant—being the mean of the samples of that stratum.

Because samples are markedly smaller than the interpolated grid cells (especially the historical sample), we opted for *block kriging* (Cressie 1993; Burrough & McDonnell 1998), a method that produces the best linear unbiased predictor for an area. This will provide us with an averaged value per grid cell, the block mean value, and the respective estimation variance (using block instead of point kriging will, generally, result in a smoother predicted surface and in much smaller estimation variances).

The average value of z over a block *B* is given by

$$z(B) = \int_B \frac{z(\mathrm{s})}{\mathrm{area}B} \tag{8}$$

estimated by

$$\hat{z}(B) = \sum_{i=1}^{n} \lambda_i \cdot z(\mathrm{s}_i) \tag{9}$$

with the sum of all weights, λ_i, equalling one, and under minimization of the estimation variance $\hat{\sigma}^2(B)$. The minimum estimation variance is given by

$$\hat{\sigma}^2(B) = \sum_{i=1}^{n} \lambda_i \bar{\gamma}(s_i, B) + \phi - \bar{\gamma}(B, B) \tag{10}$$

obtained when

$$\sum_{i=1}^{n} \lambda_i \gamma(s_i, s_j) + \phi = \bar{\gamma}(x_j, B) \qquad \text{for all } j. \tag{11}$$

The estimation variance can be used to calculate confidence intervals for each block estimate. The 95% confidence interval is then $[\hat{z}(B) - 2\hat{\sigma}, \hat{z}(B) + 2\hat{\sigma}]$ and a 99% confidence interval $[\hat{z}(B) - 3\hat{\sigma}, \hat{z}(B) + 3\hat{\sigma}]$.

The available data differ in their precision on spatial reference. The actual data set contains exclusively samples recorded on tenth kilometre co-ordinates. Historical data are mostly recorded per square kilometre. The actual interpolation took place with block-kriging on a 250×250 m^2 grid for the most recent time period, and on a 1000×1000 m^2 grid for the historical period. An example of a GSTAT command file is shown in Appendix II.1).

In the historical maps, we chose to omit the eastern and southern part of *Flevoland* from our interpolation because it is a relatively new region, drained between 1950 and 1968 and the few available historical samples are not likely to reflect a stable abiotic situation.

3.3 Mapping

National maps of the kriging results—interpolated values and respective kriging variances—were produced, for each time period, and abiotic variable. The kriging results of the strata belonging to one time period and one variable were imported into the GRID module of ARCINFO and combined into one national map. Maps were produced using GEOVIEW 4.0 (ARIS[1], RIVM/CIM[2]). An example of a map producing (krt-) file is shown in Appendix II.2.

For the actual time period, each stratum represents a distinct subgrid of the total map. Strata do not overlap and are easy to recombine into a single grid map. For the historical data, however, we are dealing with overlapping strata, because each stratum is interpolated for every kilometre grid where it does occur. To present all results without having to display one map per soil type, we display these abiotic values—though interpolated on a square kilometre grid—within the class limits of the original soil map, i.e. on a 250×250 m^2 grid. The block results of each interpolated kilometre cell are thus only displayed within the smaller cells of the respective original soil subgrid.

[1] Consultant for Spatial Information Systems, Utrecht
[2] RIVM/Department for Environmental Information Systems

4. Results

4.1 Spatial analysis

The empirical semivariograms obtained differ markedly for different time periods, variables and soil types (Figures 4.1 to 4.8). Most of them display a clear range of spatial dependence within the maximum distance between sample pairs considered (40 km). The features of the fitted semivariogram models are summarized in Appendix III.1.

When studied per stratum, the empirical semivariograms for peat and the aquatic stratum seem the most consistent in time and for the various abiotic site conditions. Sand strata show accentuated to very modest spatial structure, whereas the empirical semivariograms of non-calcareous loam have the tendency to be very scattered.

For the actual data set, spatial dependence seems unsubstantial for sand-poor and sand-rich strata, for all abiotic variables studied; clay and peat strata show a more accentuated partial sill; non-calcareous loam has a scattered semivariogram for all variables, and the aquatic stratum performs well for all variables except salinity. The models obtained for salinity are consistently less satisfactory than the ones for the remaining variables.

With reference to the historical data, we find a well-defined spatial dependence model for all sand strata and all abiotic variables, except for poor and rich sand and salinity. Spatial dependence is well defined for clay, although semivariances are fairly scattered for moisture and acidity. The semivariograms for non-calcareous loam are also very scattered, especially at distances above 25 km. The aquatic stratum has, just as for the actual data set, an accentuated spatial autocorrelation, which is somewhat scattered for salinity.

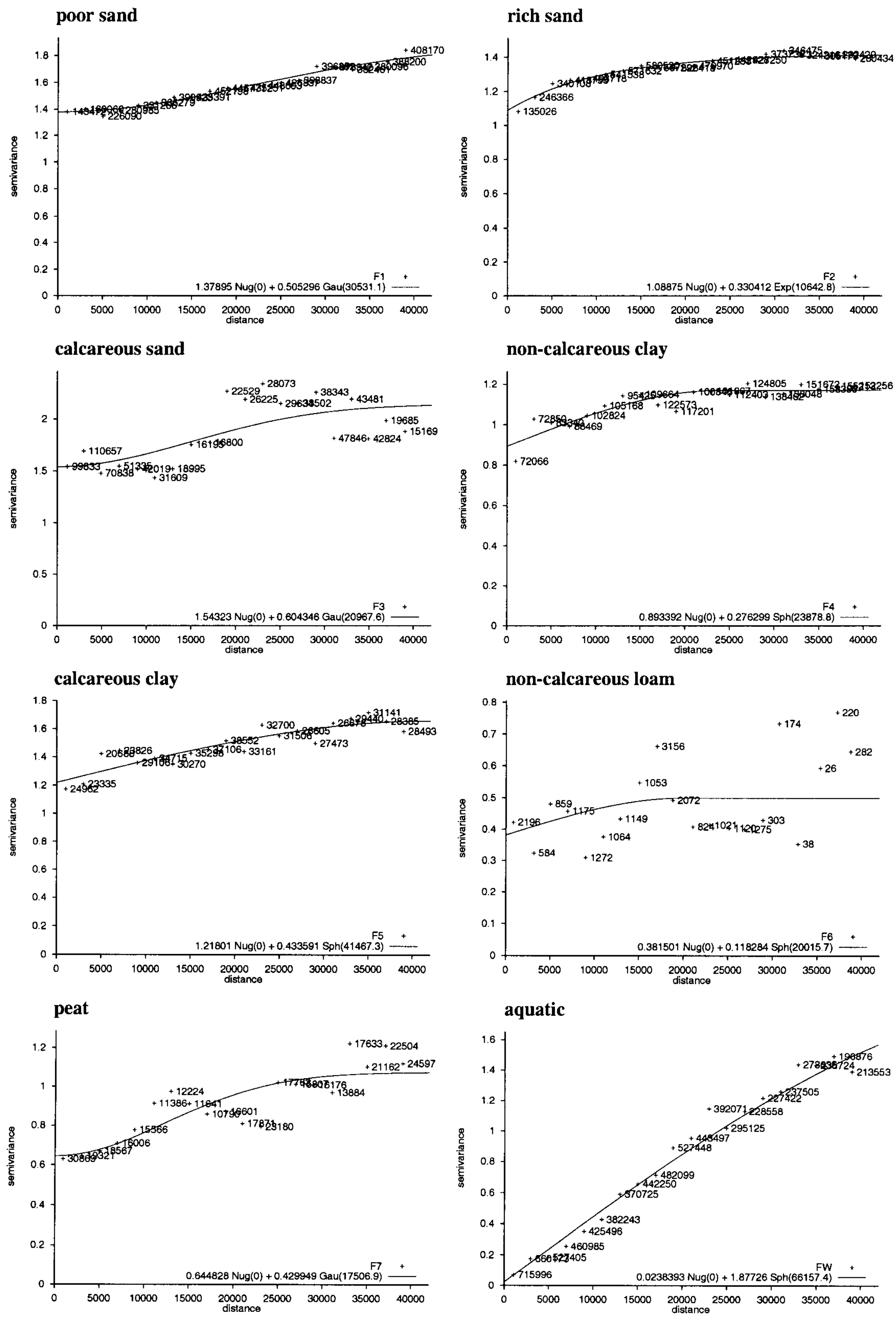

Figure 4.1. *Semivariances (+) and semivariogram models (—) of the averaged Ellenberg* ***F*** *values for samples collected from* ***1991*** *to* ***1997*** *per soil type. Model type and parameters are presented in the lower right-hand corner; numbers of sample pairs used for calculating semivariances are presented for each lag.*

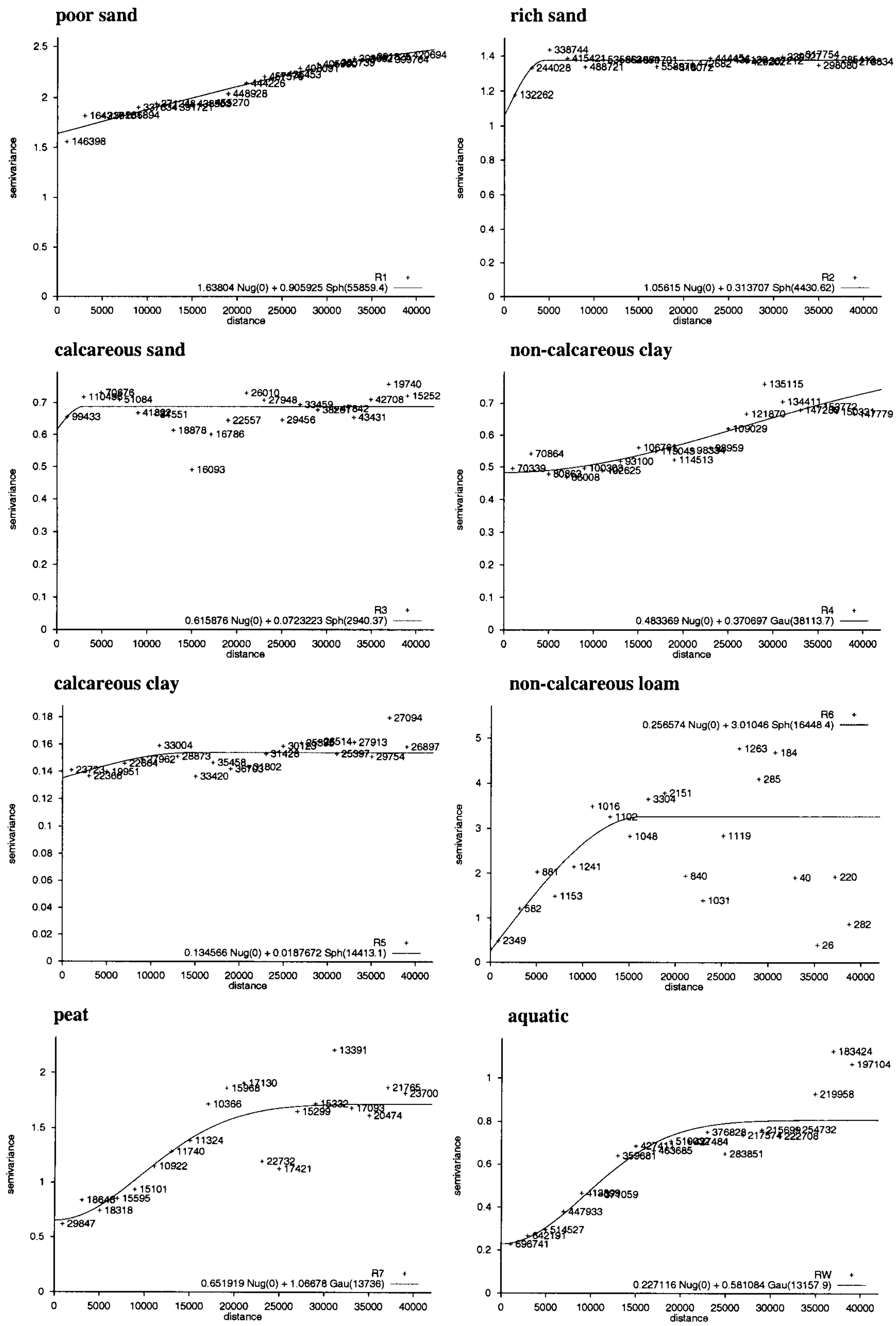

*Figure 4.2. Semivariances (+) and semivariogram models (—) of the averaged Ellenberg **R** values for samples collected from **1991** to **1997** per soil type. Model type and parameters are presented in the lower right-hand corner; numbers of sample pairs used for calculating semivariances are presented for each lag.*

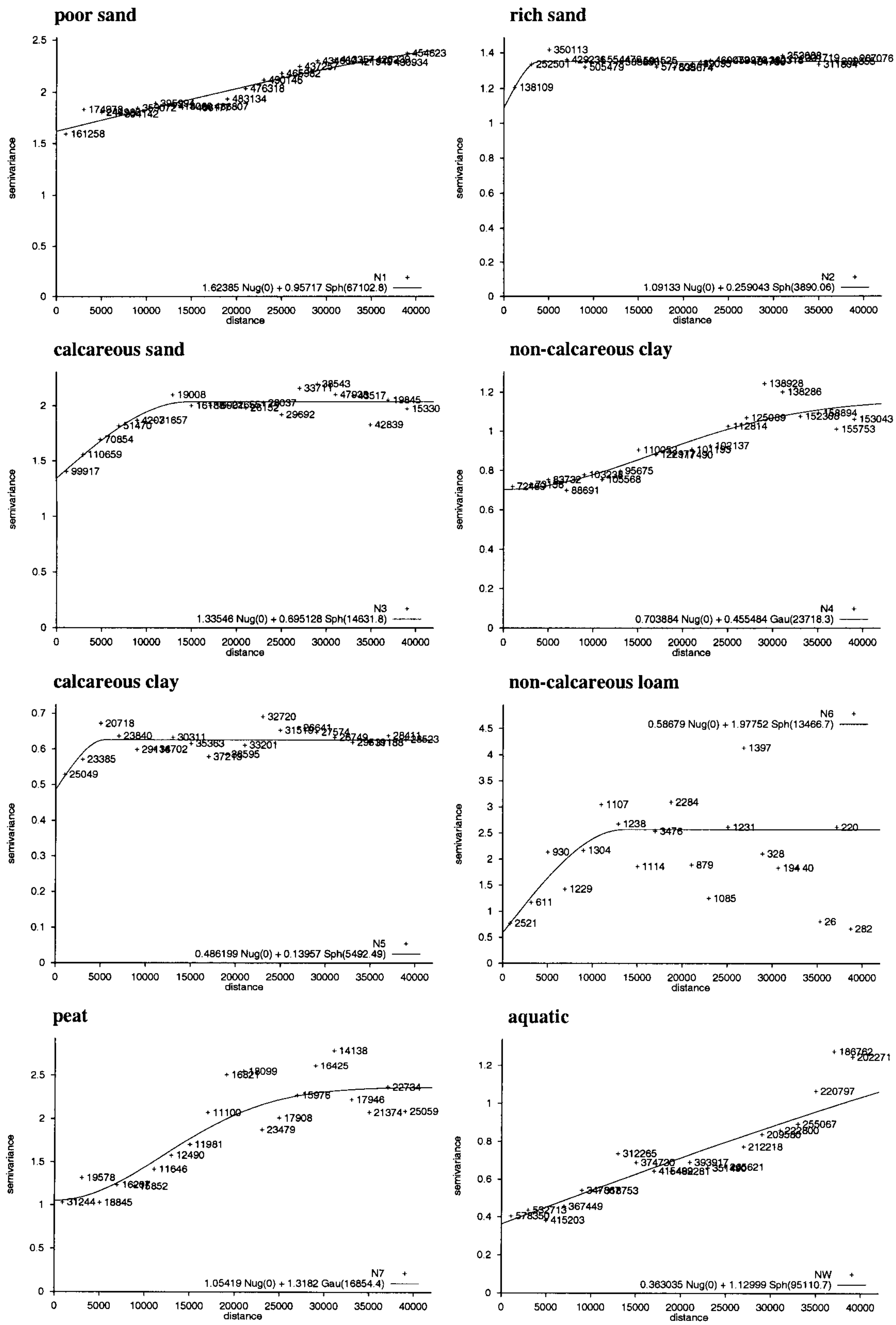

Figure 4.3. *Semivariances (+) and semivariogram models (—) of the averaged Ellenberg N values for samples collected from* ***1991*** *to* ***1997*** *per soil type. Model type and parameters are presented in the lower right-hand corner; numbers of sample pairs used for calculating semivariances are presented for each lag.*

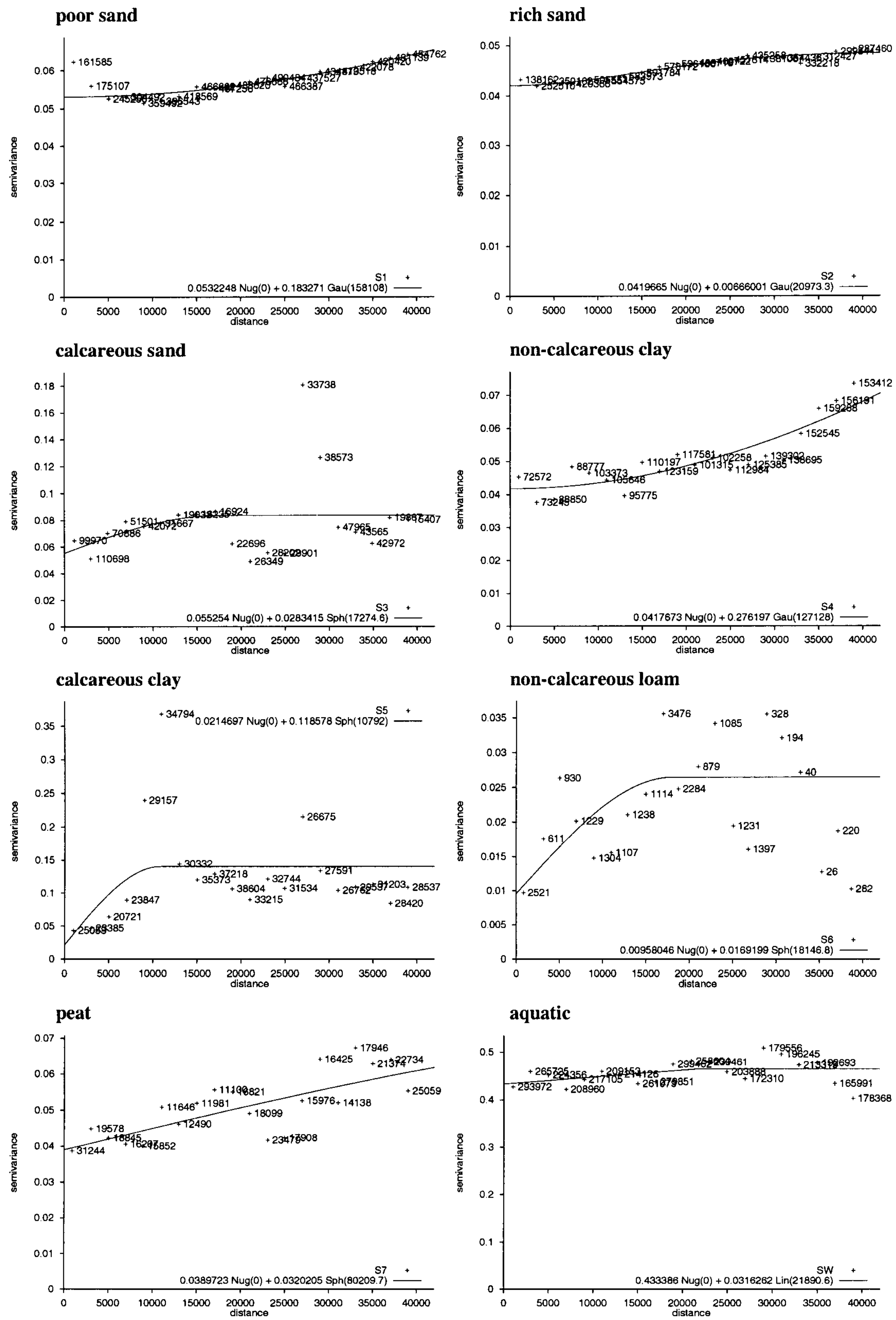

*Figure 4.4. Semivariances (+) and semivariogram models (—) of the averaged Ellenberg **S** values for samples collected from **1991** to **1997** per soil type. Model type and parameters are presented in the lower right-hand corner; numbers of sample pairs used for calculating semivariances are presented for each lag.*

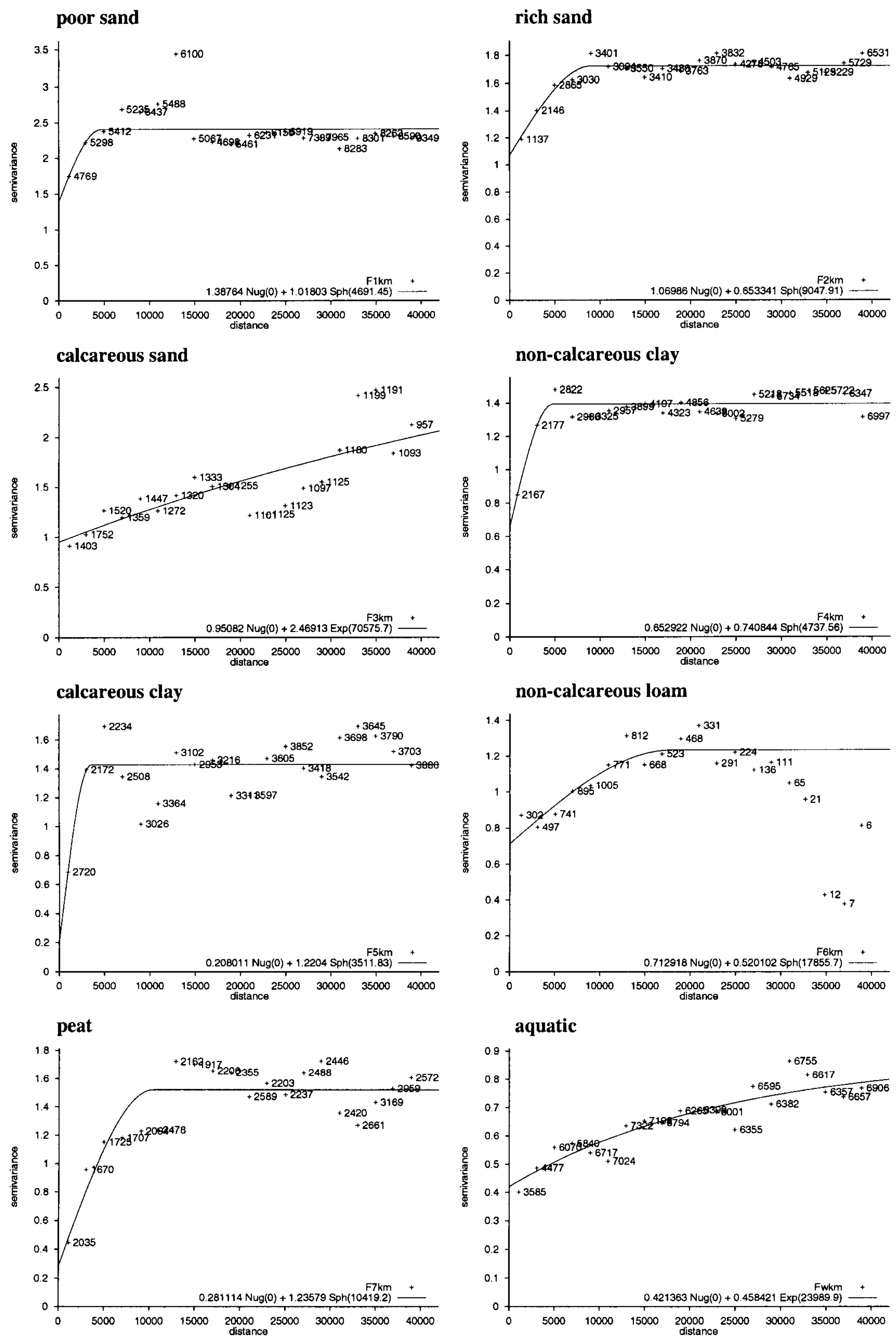

Figure 4.5. *Semivariances (+) and semivariogram models (—) of the averaged Ellenberg* ***F*** *values for samples collected from* ***1930*** *to* ***1970*** *per soil type. Model type and parameters are presented in the lower right-hand corner; numbers of sample pairs used for calculating semivariances are presented for each lag.*

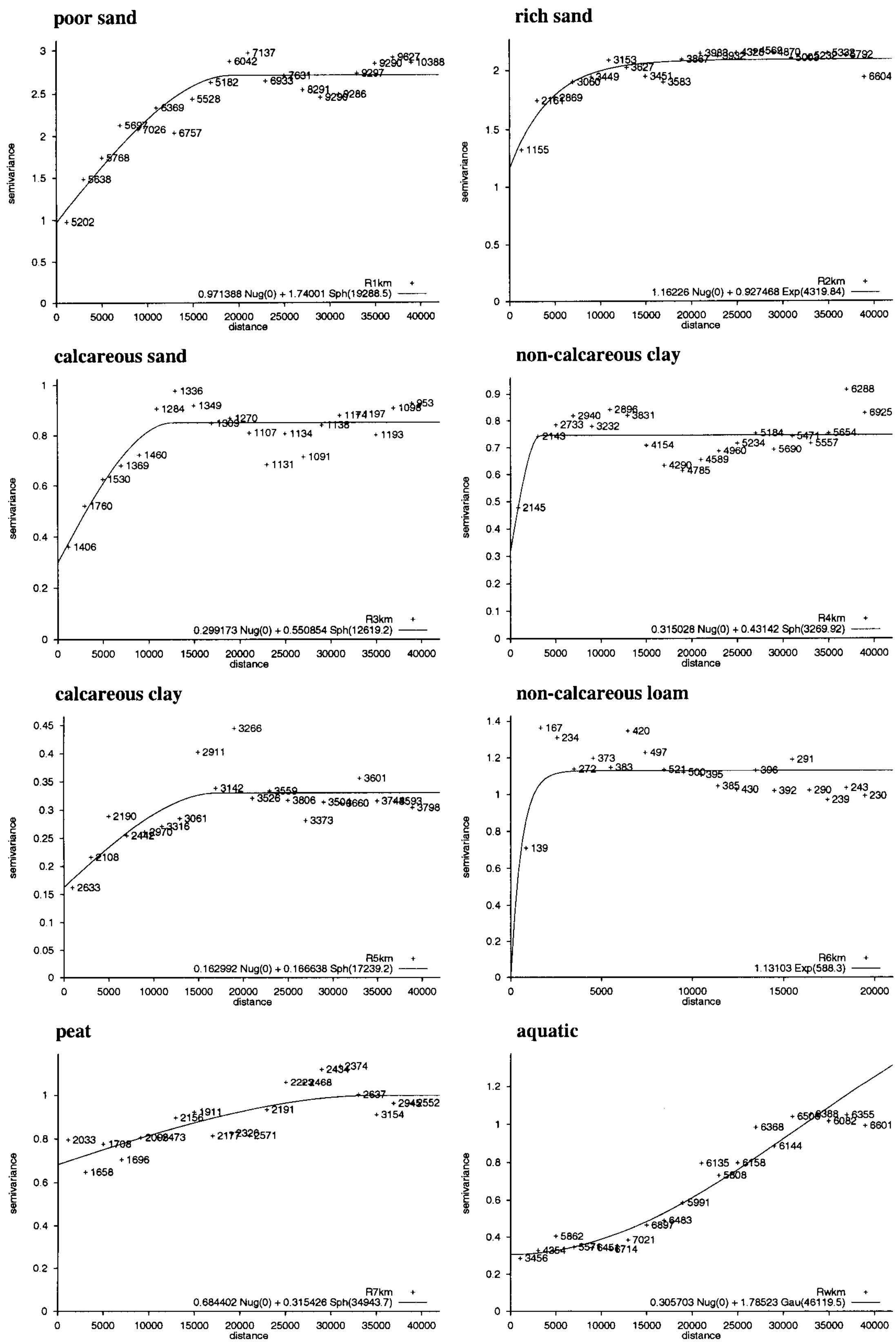

Figure 4.6. *Semivariances (+) and semivariogram models (—) of the averaged Ellenberg **R** values for samples collected from **1930** to **1970** per soil type. Model type and parameters are presented in the lower right-hand corner; numbers of sample pairs used for calculating semivariances are presented for each lag.*

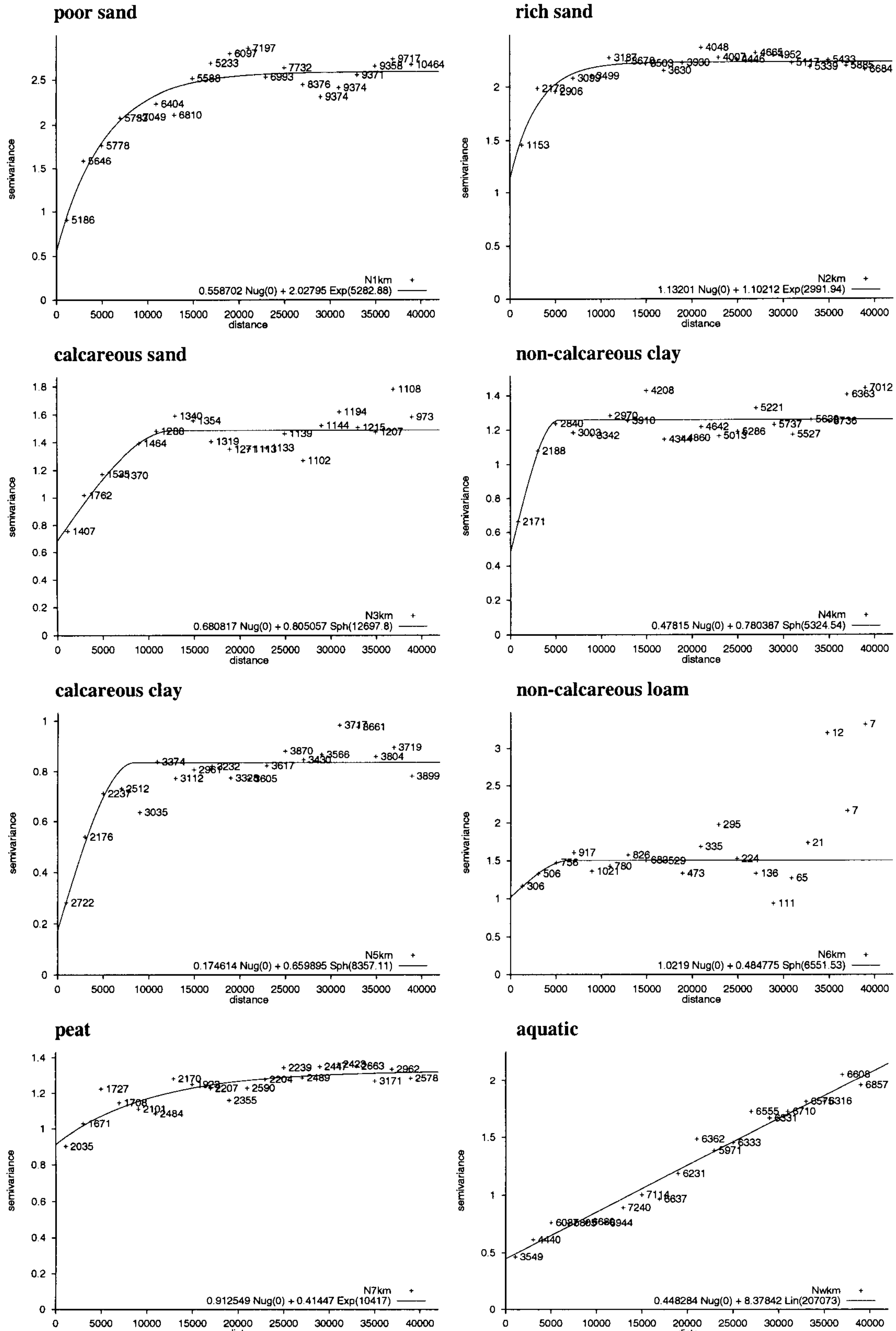

Figure 4.7. *Semivariances (+) and semivariogram models (—) of the averaged Ellenberg N values for samples collected from **1930** to **1970** per soil type. Model type and parameters are presented in the lower right-hand corner; numbers of sample pairs used for calculating semivariances are presented for each lag.*

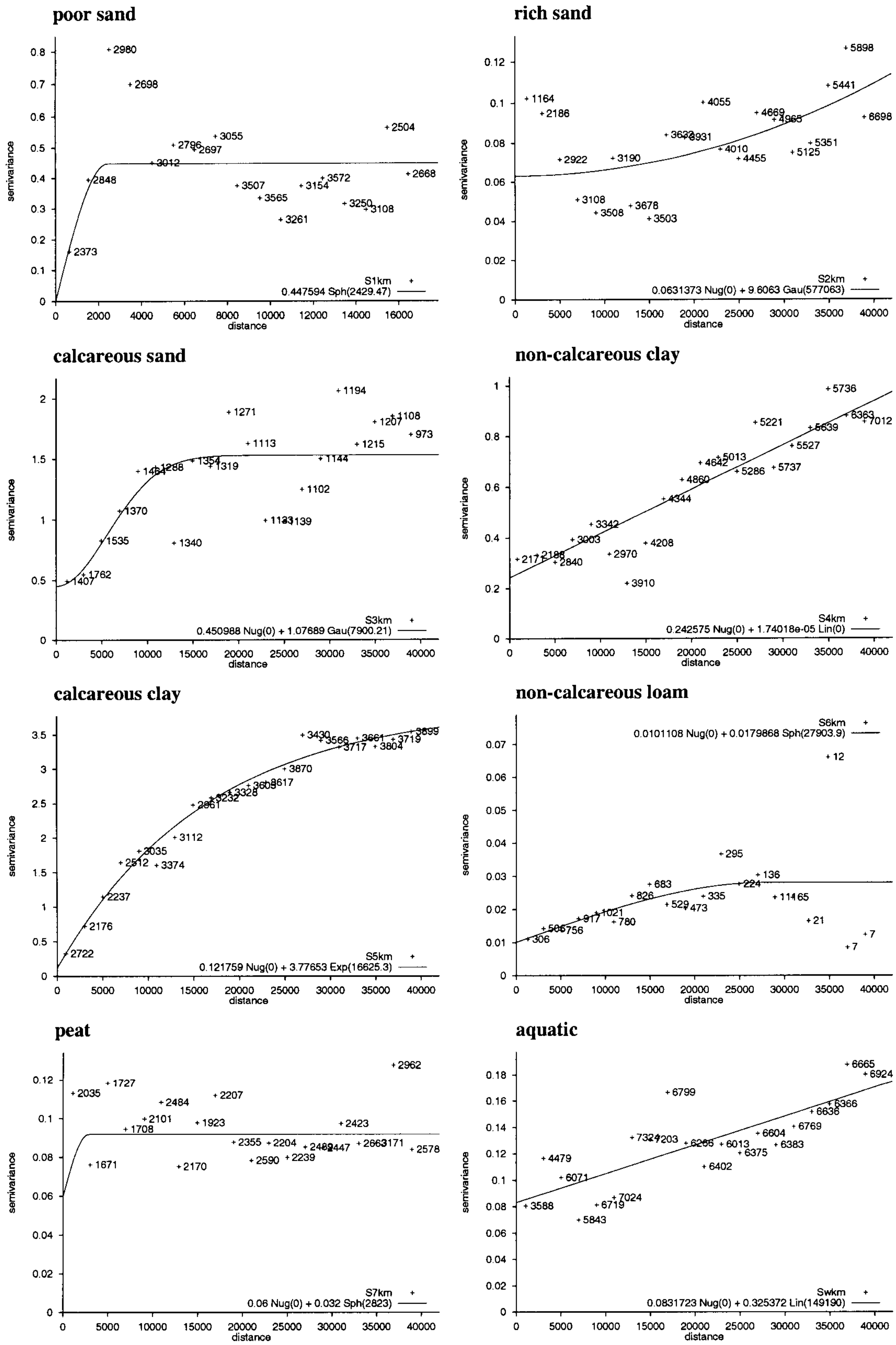

Figure 4.8. *Semivariances (+) and semivariogram models (—) of the averaged Ellenberg **S** values for samples collected from **1930** to **1970** per soil type. Model type and parameters are presented in the lower right-hand corner; numbers of sample pairs used for calculating semivariances are presented for each lag.*

4.2 Interpolated maps

The geostatistical interpolation of abiotic site conditions, carried out per stratum, was combined in gridded actual and historical maps for moisture, acidity, nitrogen and salinity, and their respective kriging variances (Figures 9 to 24).

Interpolated values are summarized per abiotic variable and per stratum in Figures 25 & 26 (and Appendix III.2). The kriging variances range, for the actual data set, from 0.00 to 2.68, with an average of 0.16 for moisture; from 0.00 to 5.73, with an average of 0.17 for acidity/alkalinity; from 0.01 to 3.76, with an average of 0.20 for nitrogen, and from 0.00 to 0.26, with an average of 0.02 for salinity. For the historical data set, the kriging variance ranges from 0.02 to 2.43, with an average of 0.62 for moisture; from 0.00 to 2.20, with an average of 0.41 for acidity; from 0.01 to 2.35, with an average of 0.65 for nitrogen, and from 0.00 to 5.04, with an average of 0.30 for salinity. All digital records associated with the maps are cited in Apendix III.3.

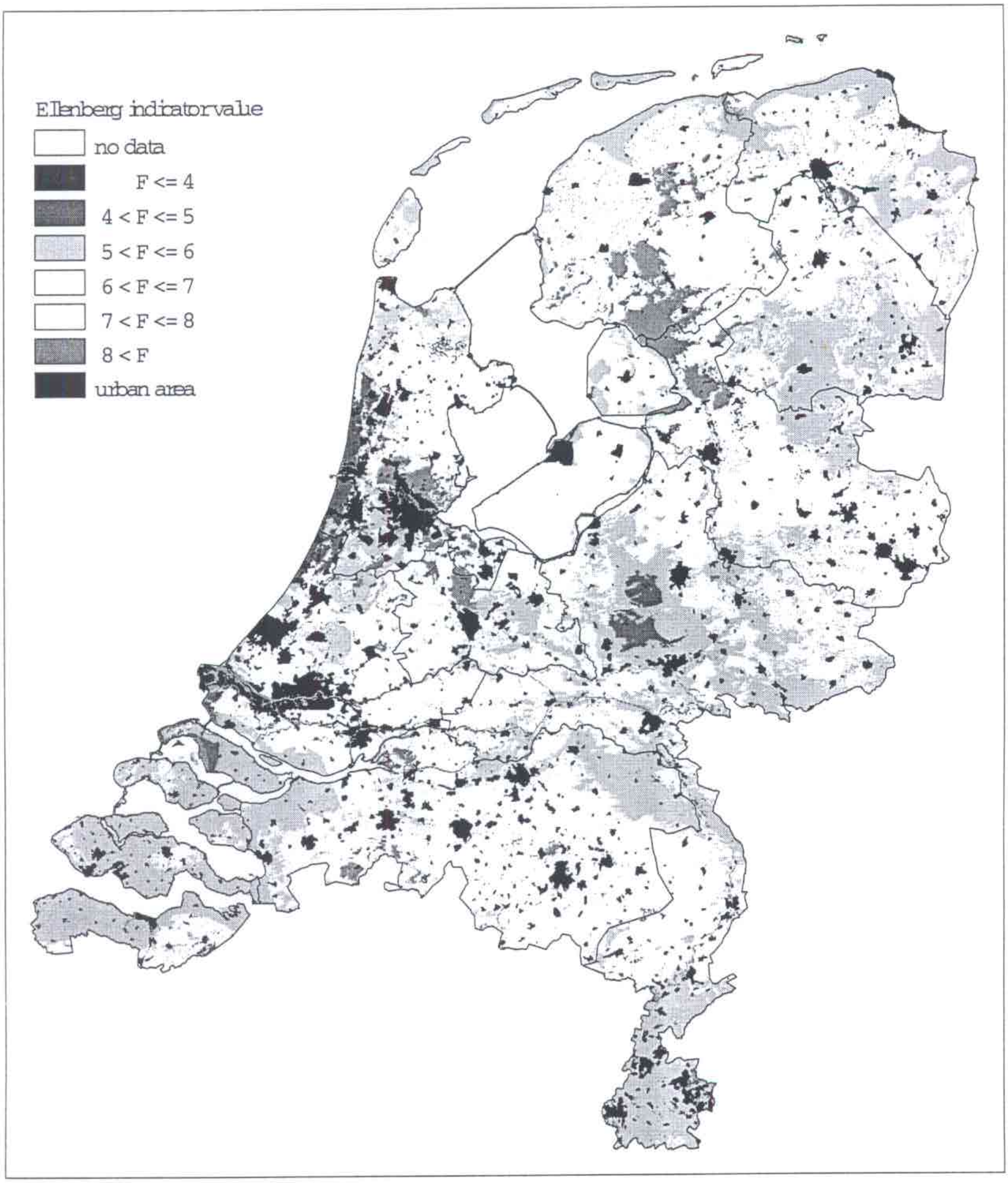

Figure 4.9. *Ellenberg **F** values, interpolated per soil type on a 250×250 m² grid and based on data collected from **1991** to **1997** (actual situation).*

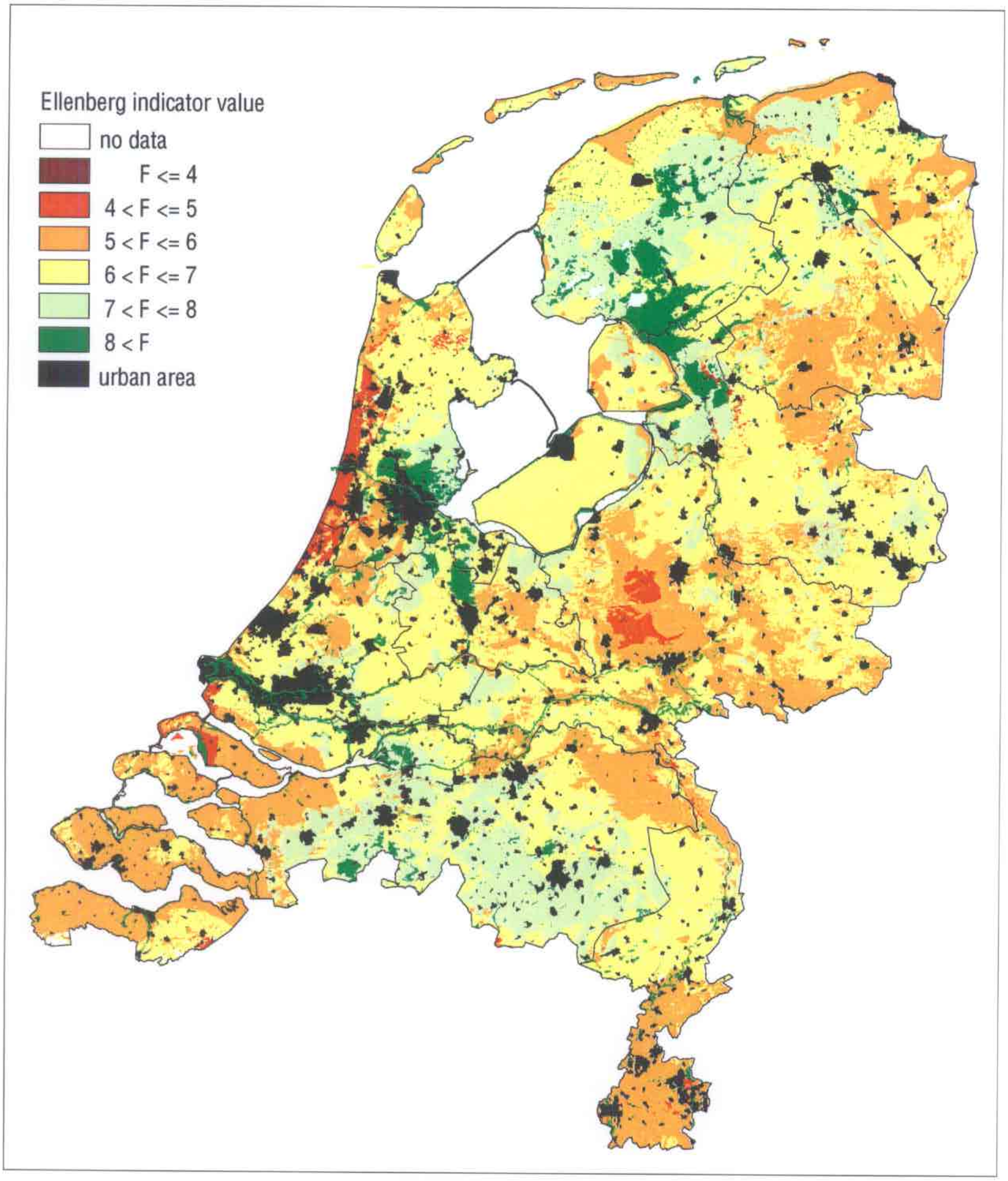
Ellenberg indicator value
no data
F <= 4
4 < F <= 5
5 < F <= 6
6 < F <= 7
7 < F <= 8
8 < F
urban area

Fig 4.9

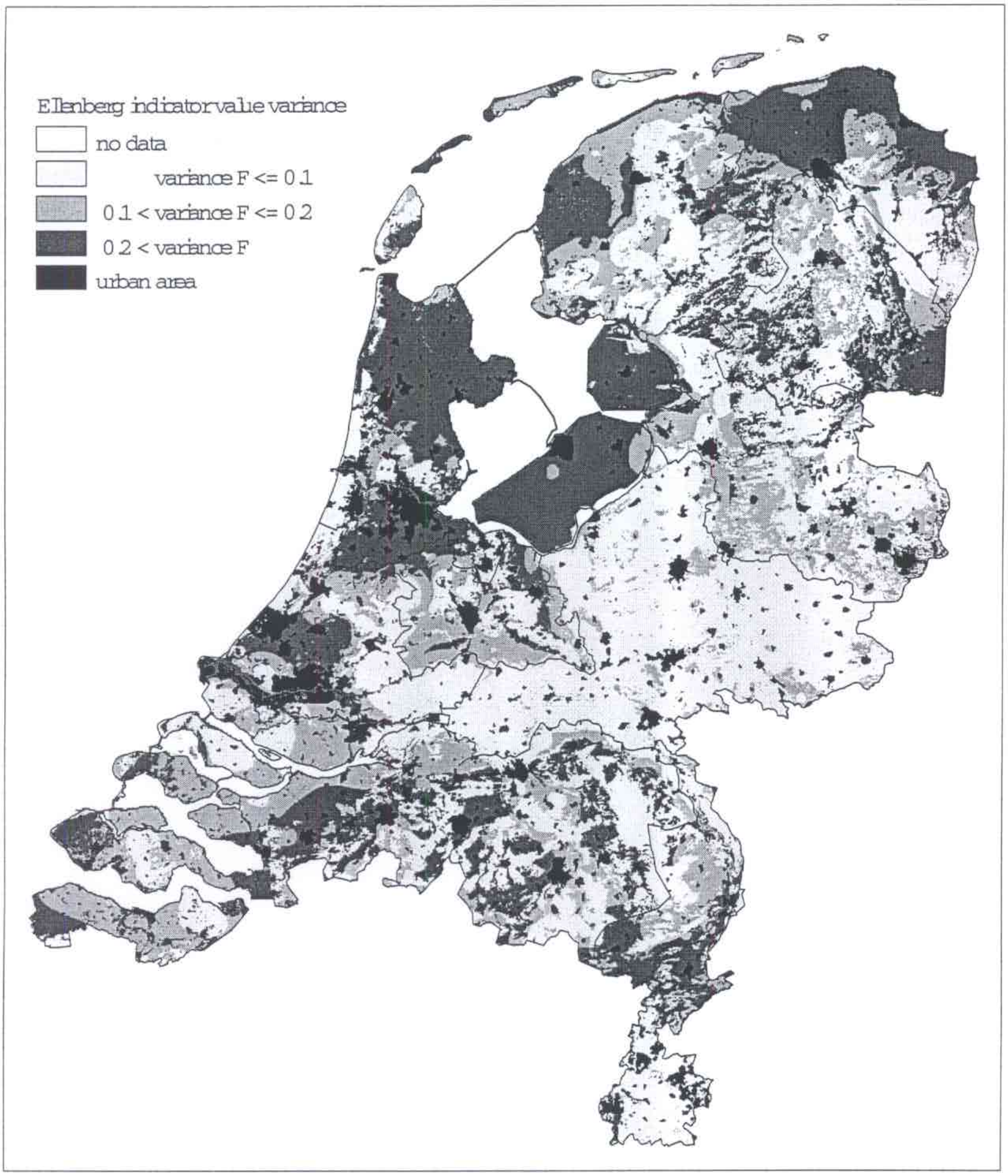

Figure 4.10. *Kriging variance of Ellenberg* ***F*** *values , interpolated per soil type on a 250×250 m² grid and based on data collected from* ***1991*** *to* ***1997*** *(actual situation).*

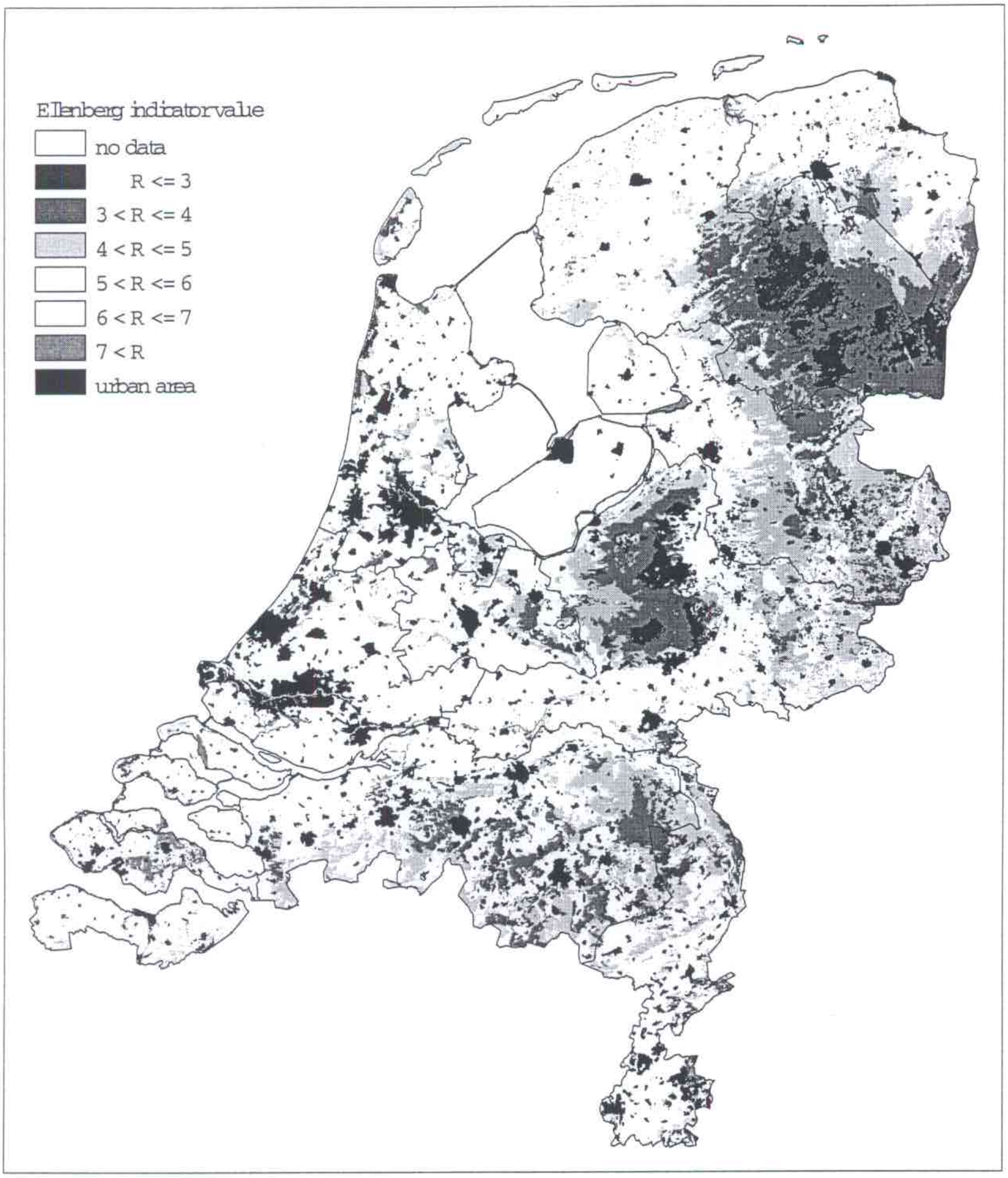

Figure 4.11. *Ellenberg* ***R*** *values, interpolated per soil type on a 250×250 m² grid and based on data collected from* ***1991*** *to* ***1997*** *(actual situation).*

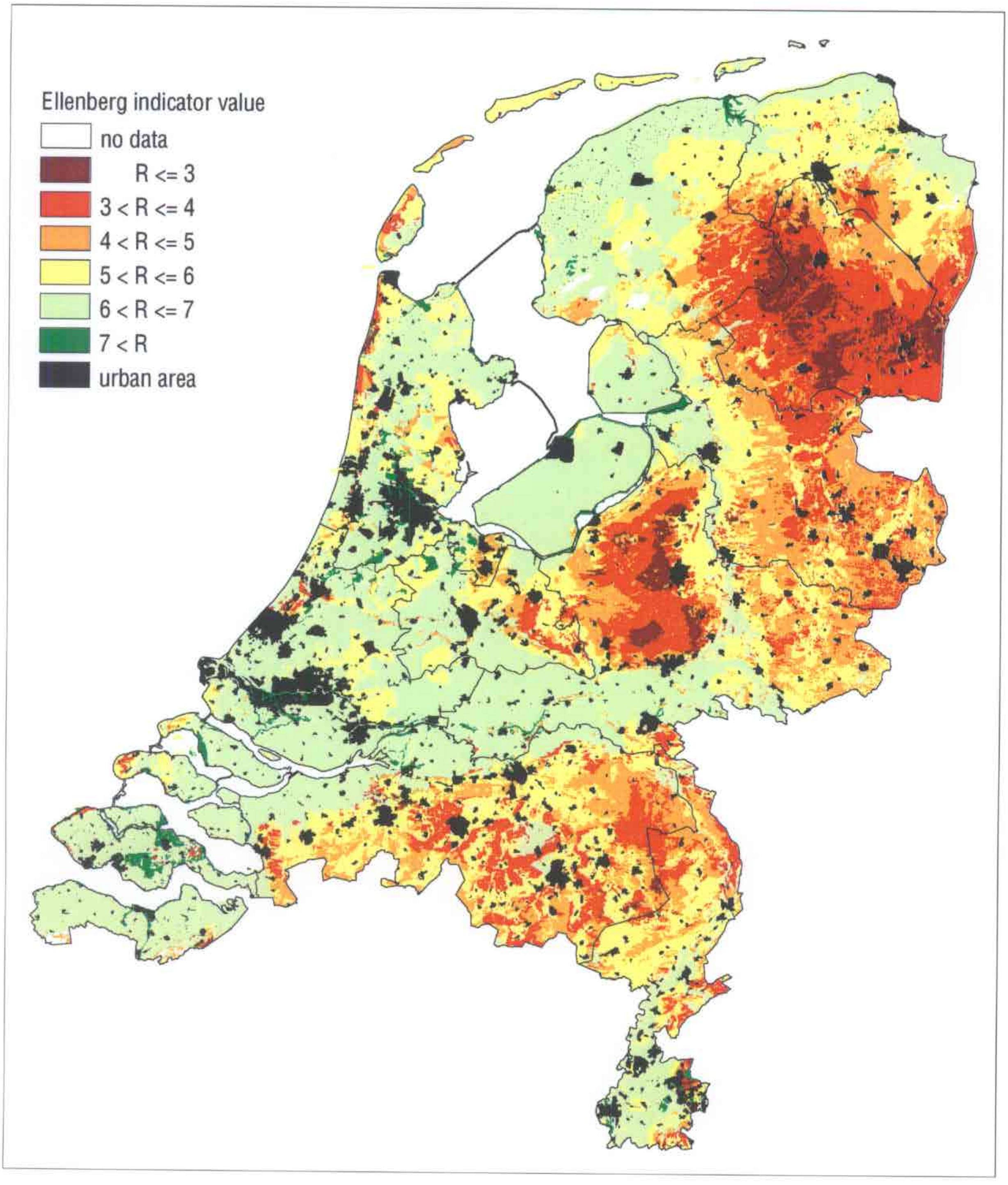
Ellenberg indicator value
no data
R <= 3
3 < R <= 4
4 < R <= 5
5 < R <= 6
6 < R <= 7
7 < R
urban area

Fig 4.11

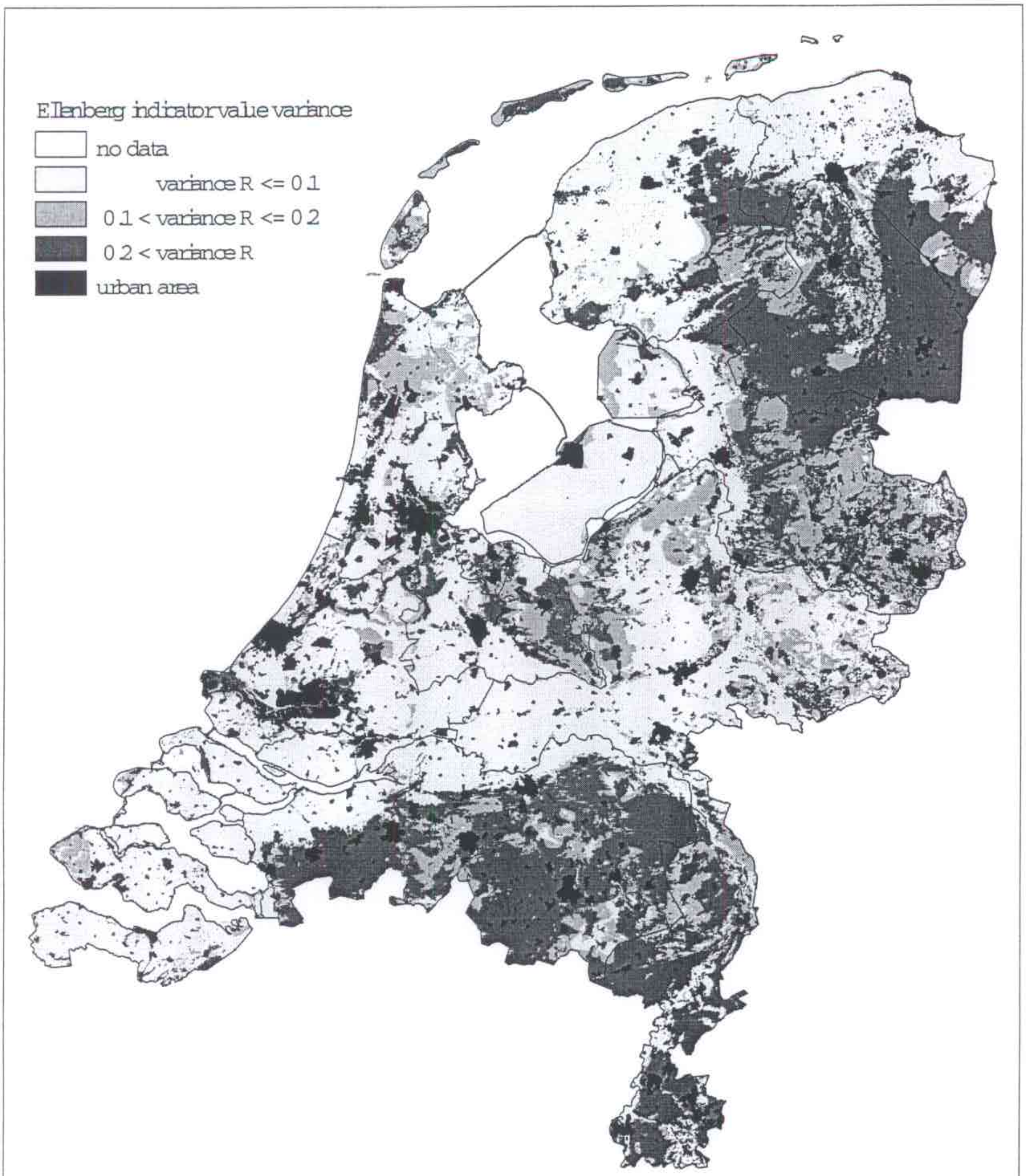

Figure 4.12. *Kriging variance of Ellenberg **R** values, interpolated per soil type on a 250×250 m^2 grid and based on data collected from **1991** to **1997** (actual situation).*

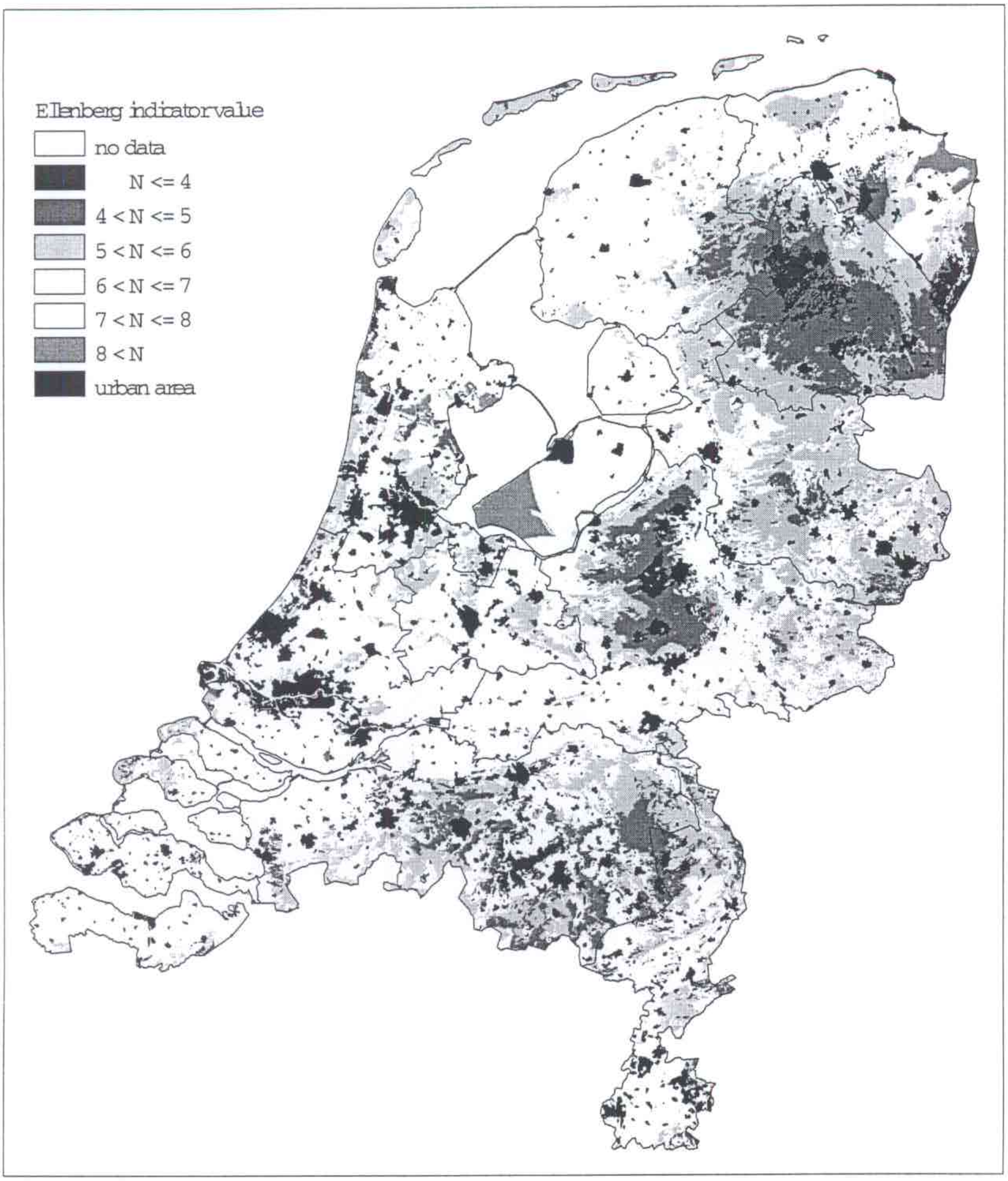

Figure 4.13. *Ellenberg N values, interpolated per soil type on a 250×250 m^2 grid and based on data collected from* ***1991*** *to* ***1997****; actual situation.*

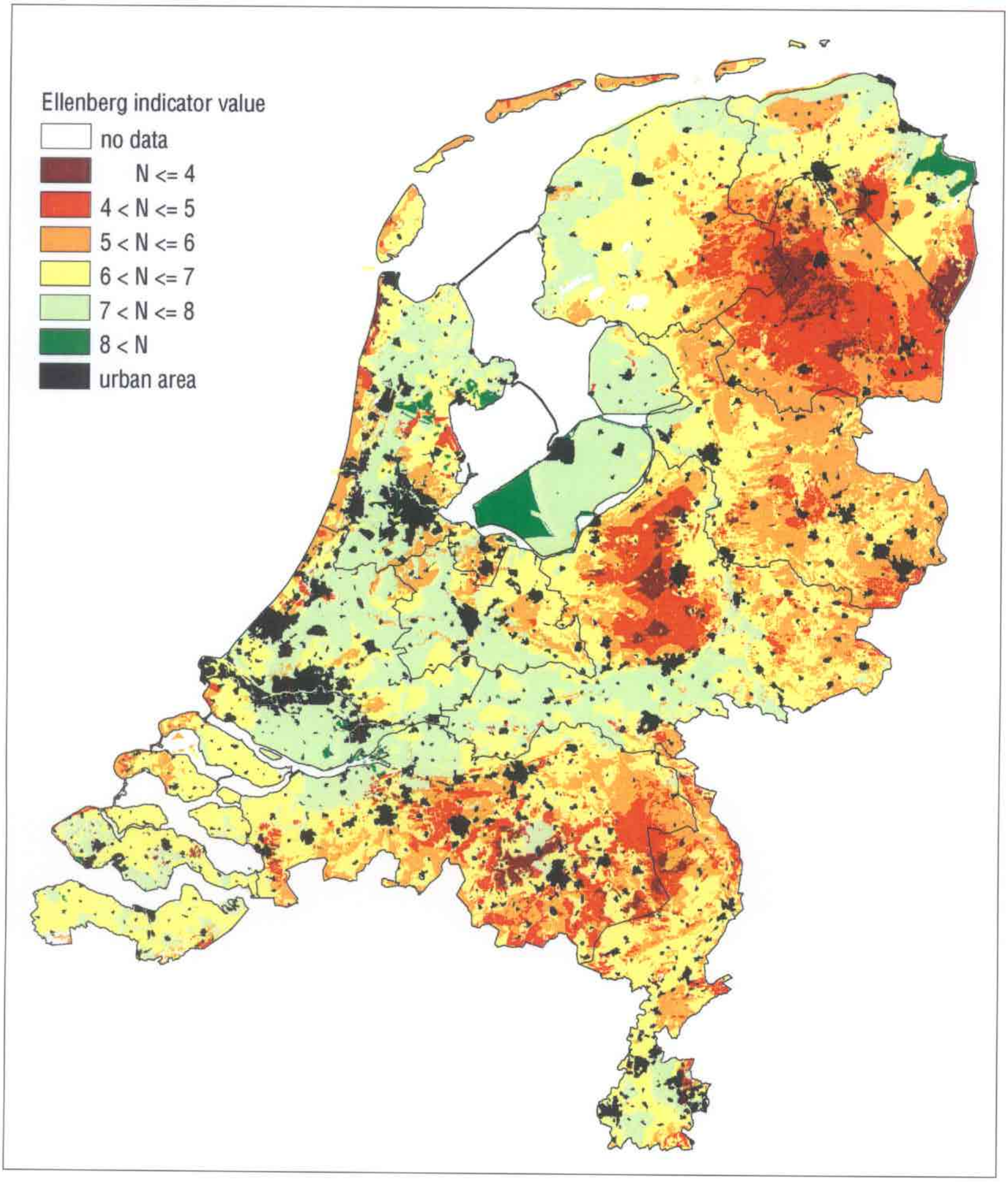
Ellenberg indicator value
no data
N <= 4
4 < N <= 5
5 < N <= 6
6 < N <= 7
7 < N <= 8
8 < N
urban area

Fig 4.13

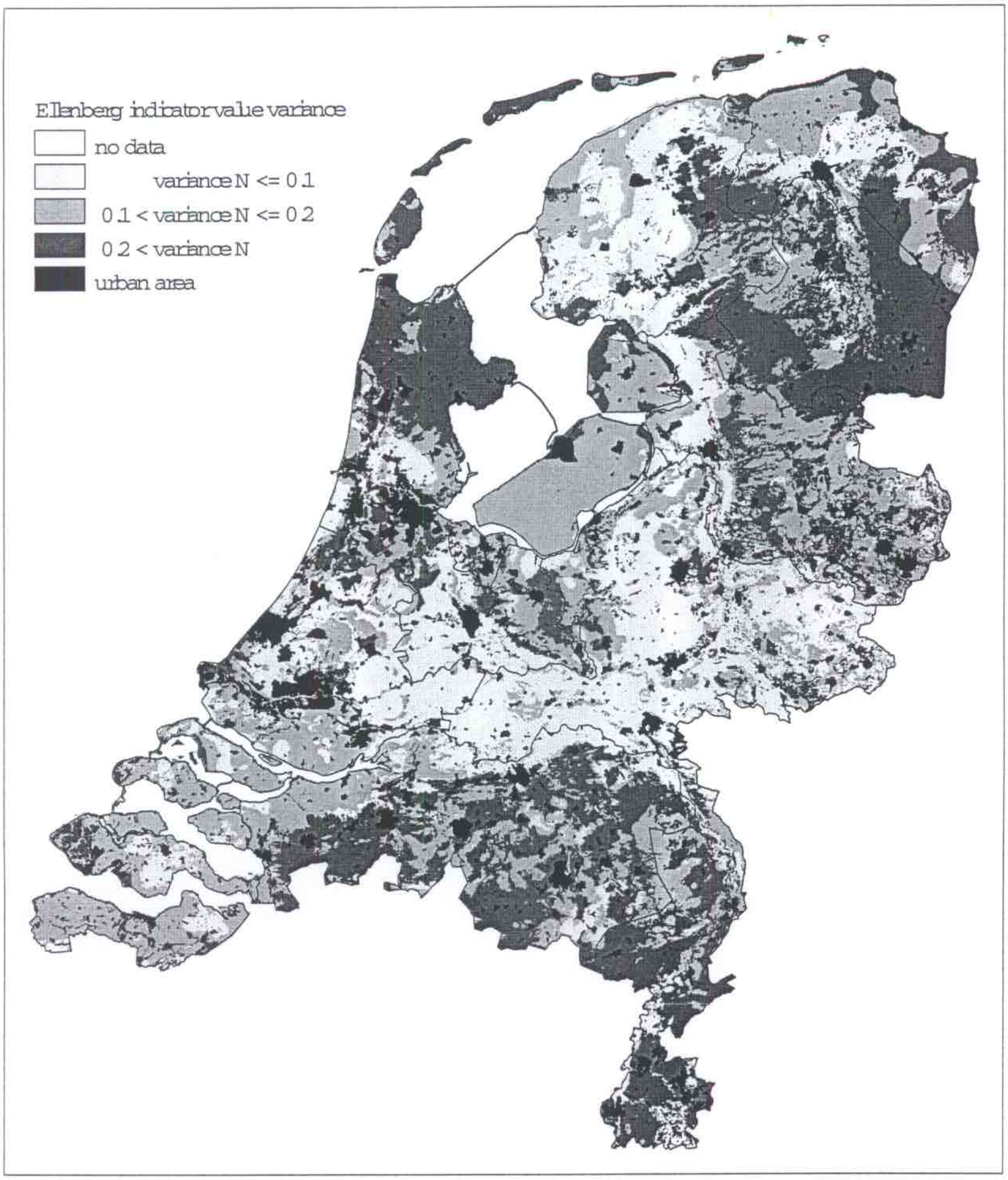

Figure 4.14. *Kriging variance of Ellenberg N values, interpolated per soil type on a 250×250 m^2 grid and based on data collected from* ***1991*** *to* ***1997*** *(actual situation).*

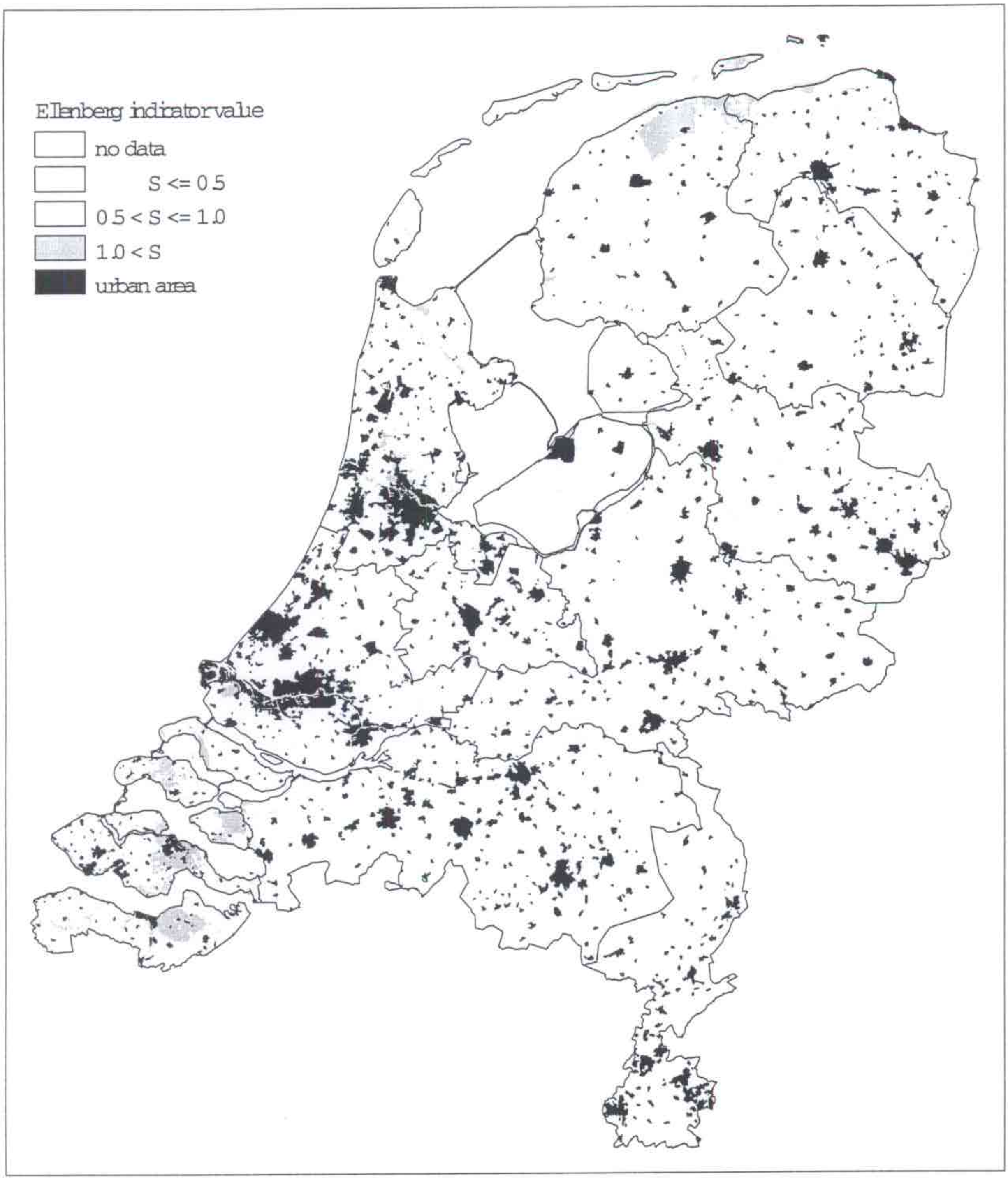

Figure 4.15. *Ellenberg **S** values, interpolated per soil type on a 250×250 m² grid and based on data collected from **1991** to **1997**; actual situation.*

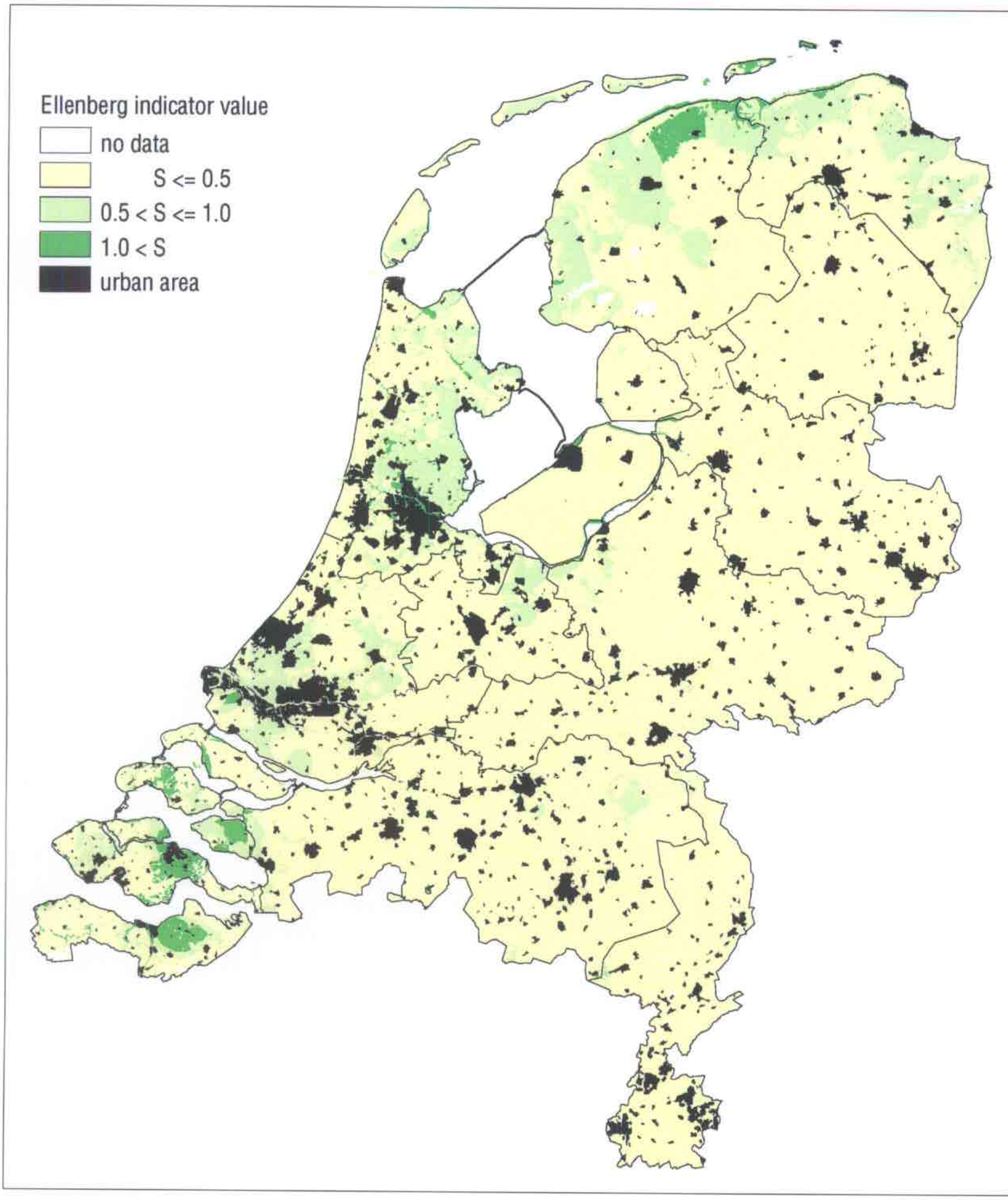

Fig 4.15

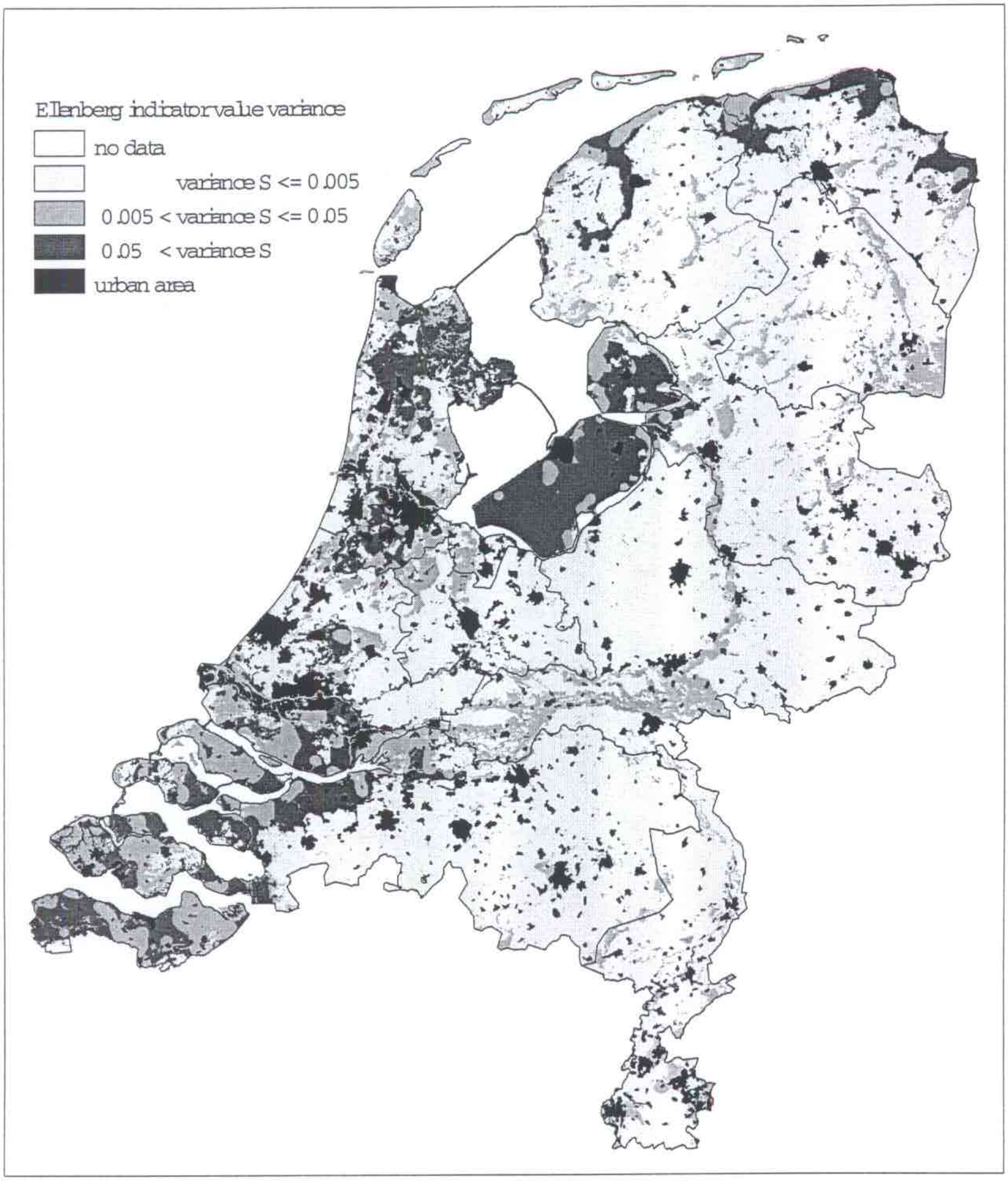

Figure 4.16. *Kriging variance of Ellenberg* ***S*** *values, interpolated per soil type on a 250×250 m^2 grid and based on data collected from* ***1991*** *to* ***1997*** *(actual situation).*

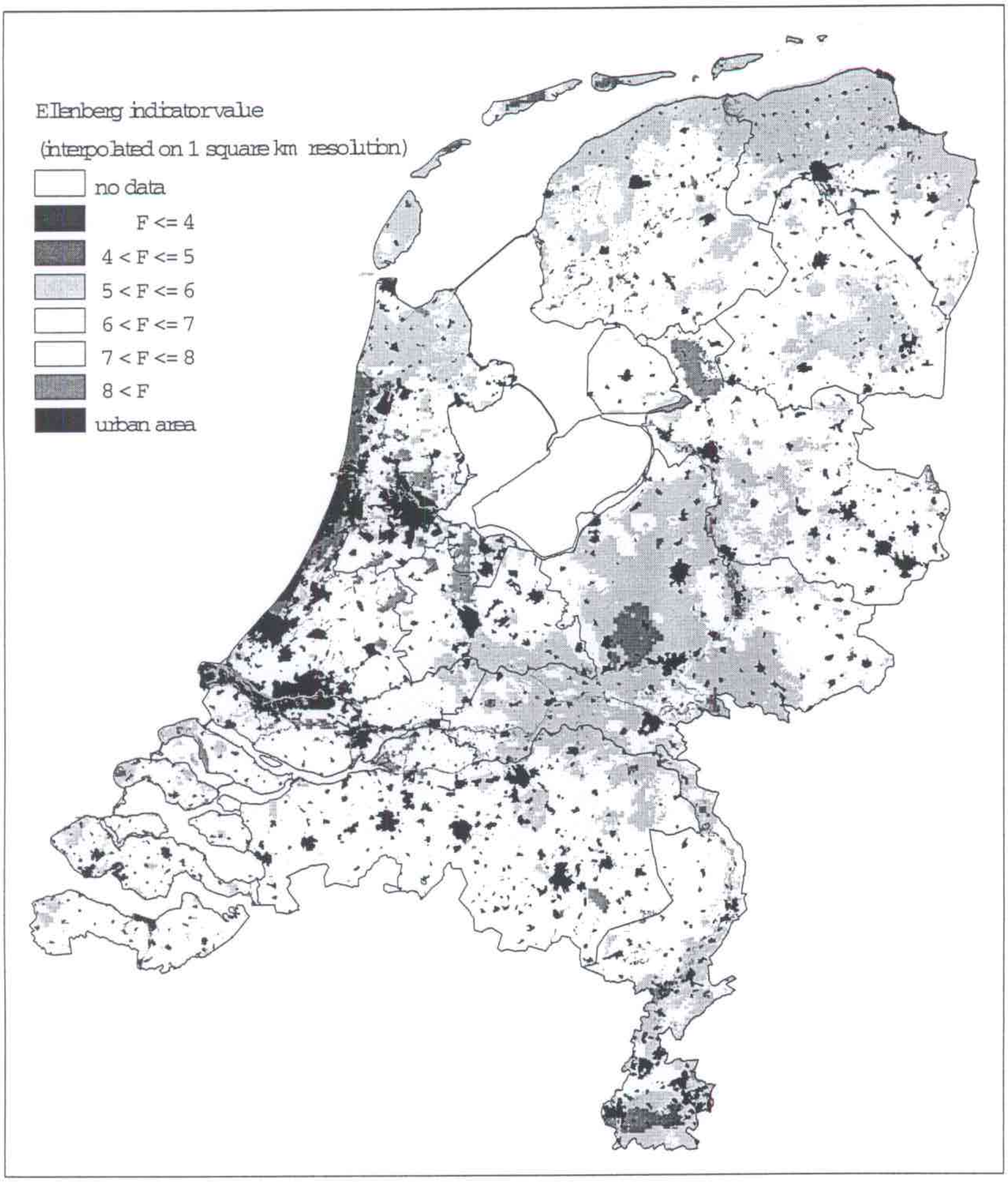

Figure 4.17. *Ellenberg **F** values, interpolated per soil type on a 1000×1000 m² grid and based on data collected from **1930** to **1970** (historical situation). N.B. For better visualization data are presented within the original 250×250 m² grid.*

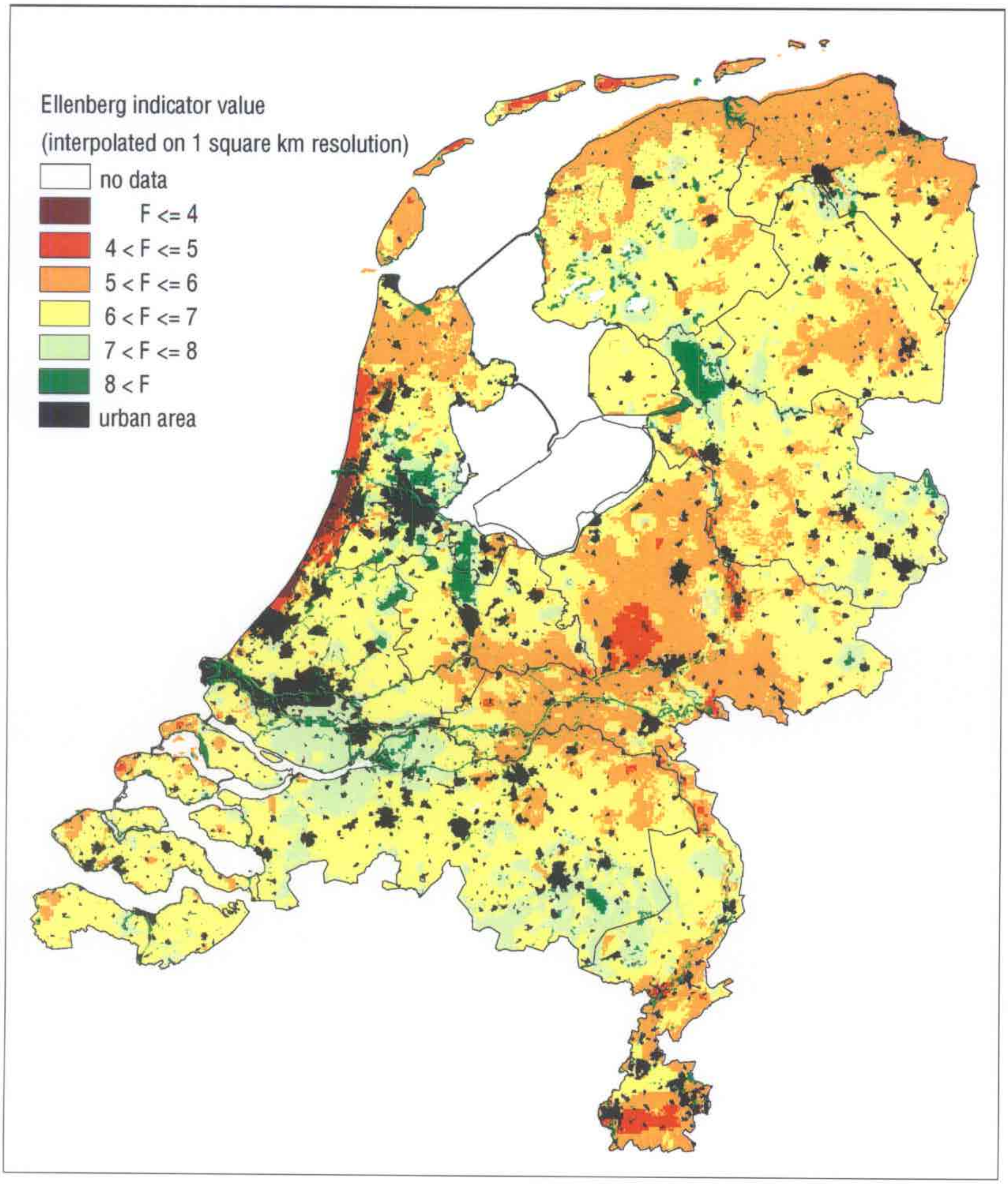
Ellenberg indicator value
(interpolated on 1 square km resolution)
no data
F <= 4
4 < F <= 5
5 < F <= 6
6 < F <= 7
7 < F <= 8
8 < F
urban area

Fig 4.17

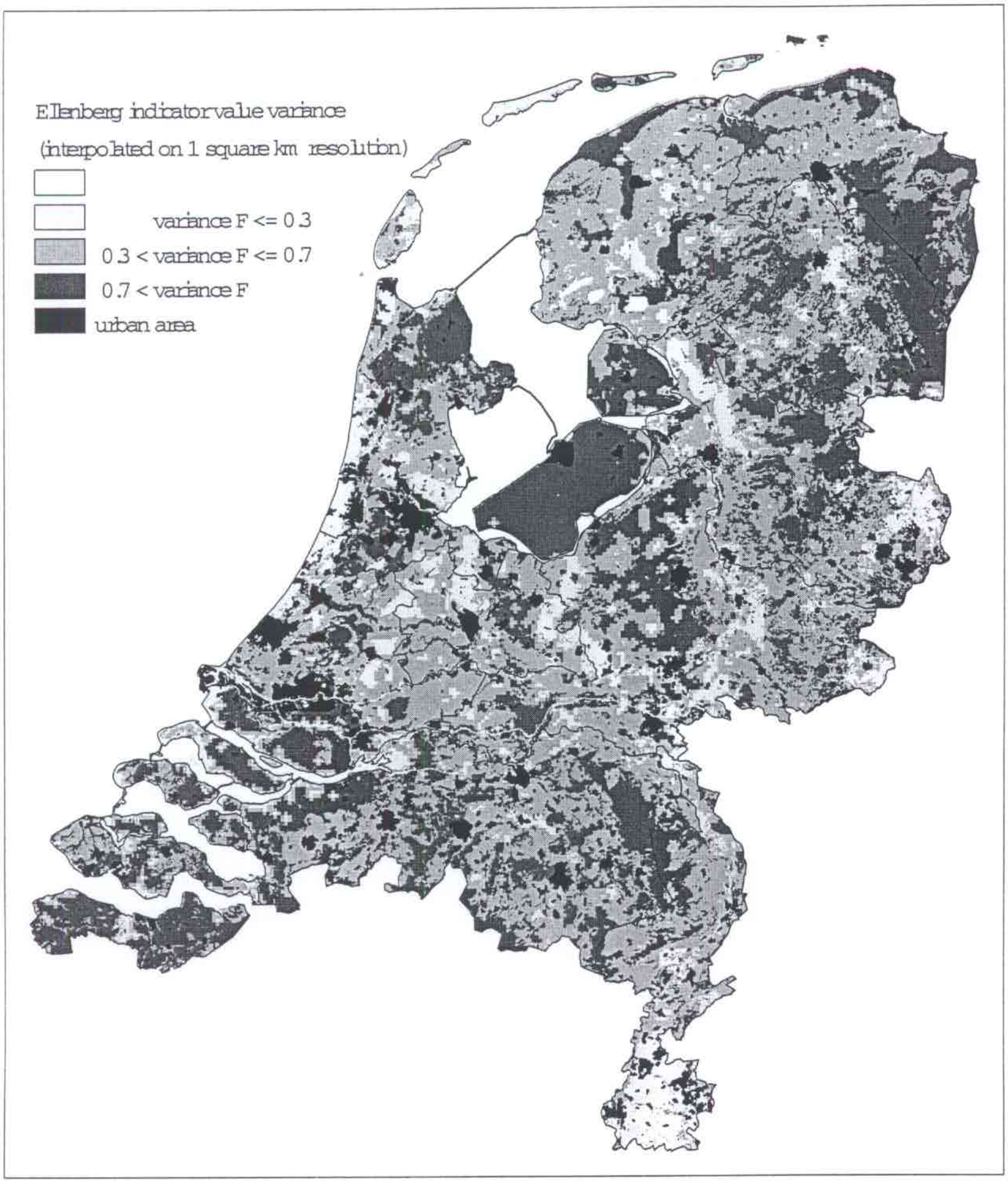

Figure 4.18. *Kriging variance of Ellenberg* ***F*** *values, interpolated per soil type on a 1000×1000 m² grid and based on data collected from* ***1930*** *to* ***1970*** *(historical situation). N.B. For better visualization data are presented within the original 250×250 m² grid.*

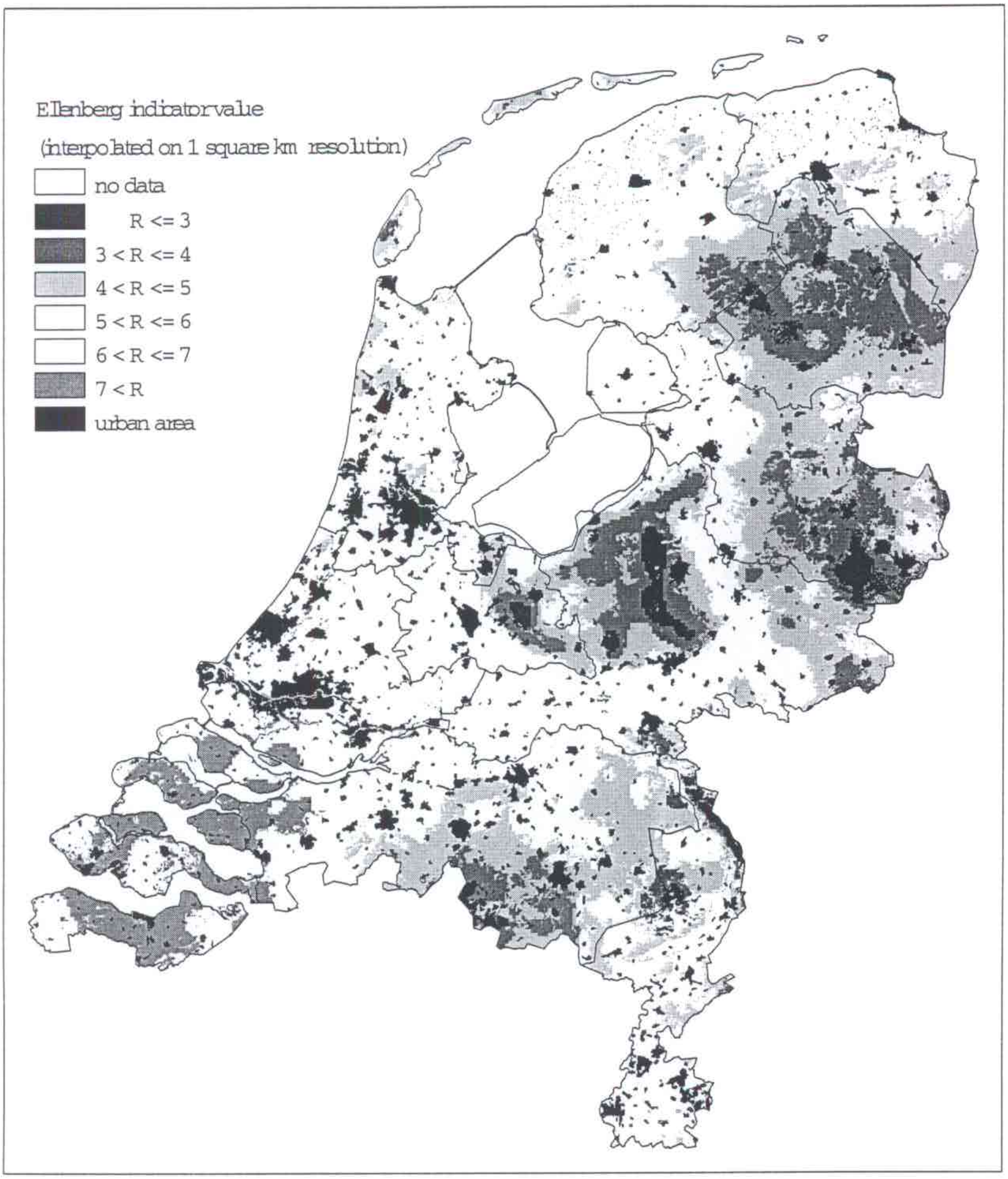

Figure 4.19. *Ellenberg* ***R*** *values, interpolated per soil type on a 1000×1000 m^2 grid and based on data collected from* ***1930*** *to* ***1970*** *(historical situation). N.B. For better visualization data are presented within the original 250×250 m^2 grid.*

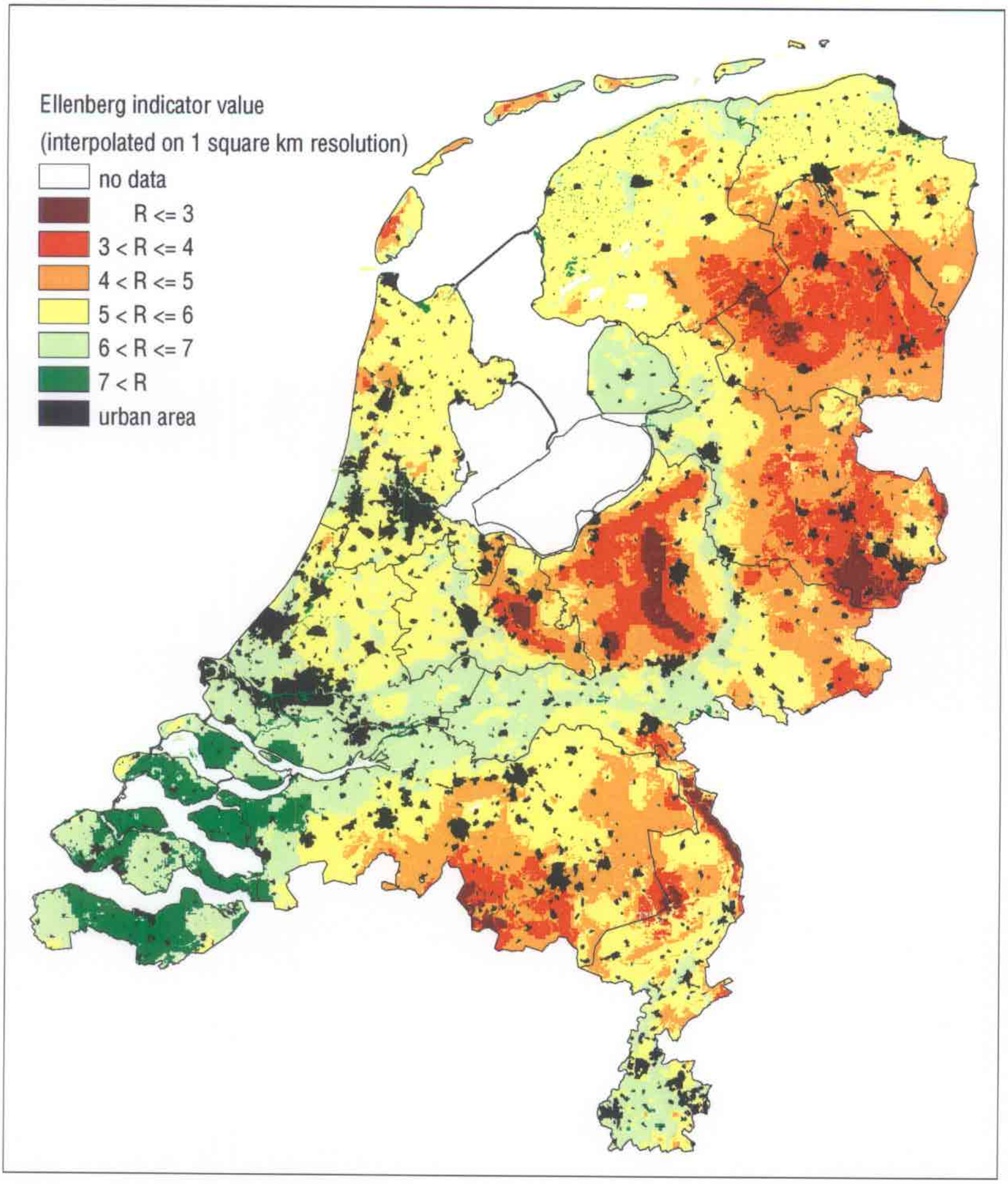
Ellenberg indicator value
(interpolated on 1 square km resolution)
no data
R <= 3
3 < R <= 4
4 < R <= 5
5 < R <= 6
6 < R <= 7
7 < R
urban area

Fig. 4.19

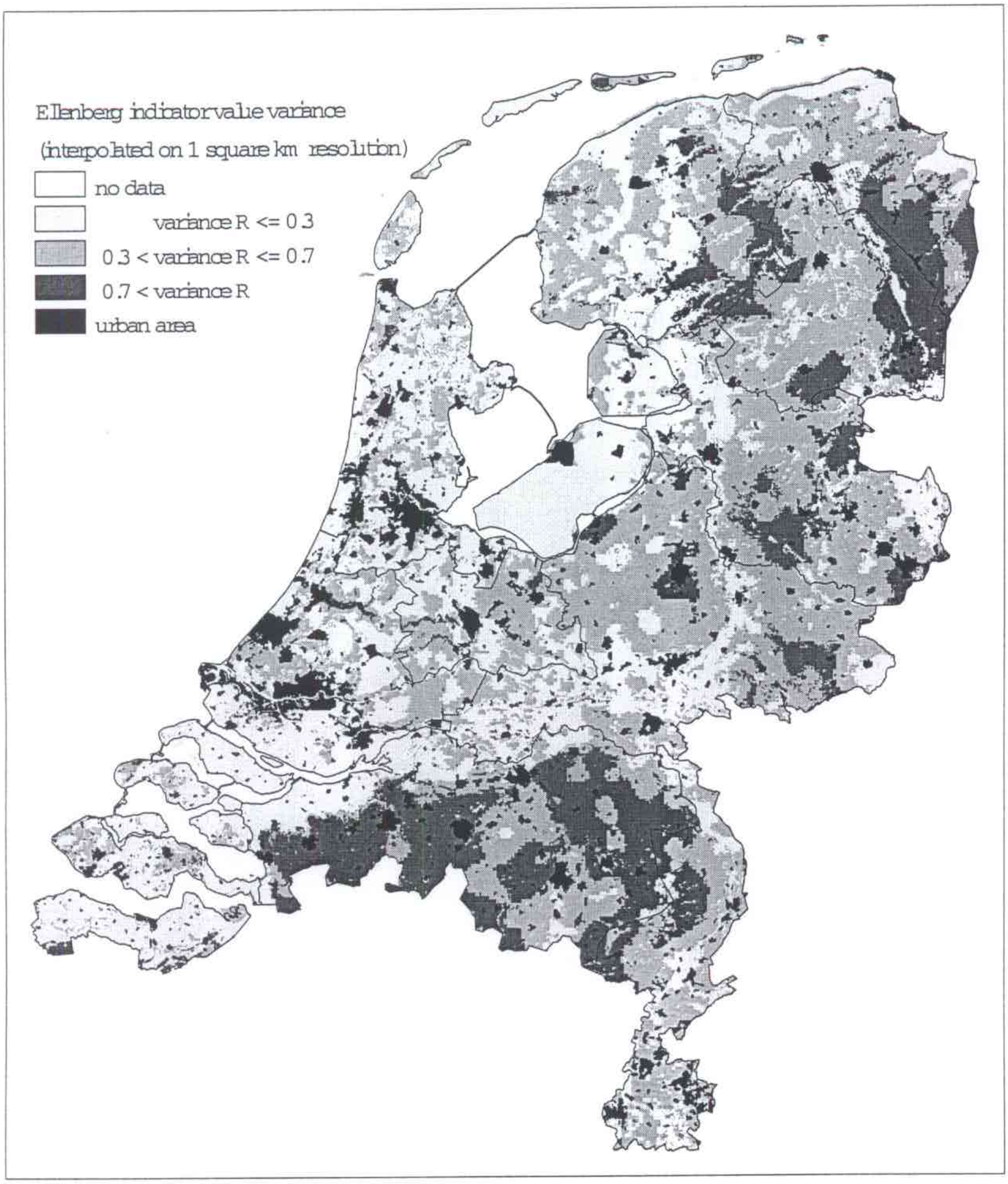

Figure 4.20. *Kriging variance of Ellenberg* ***R*** *values, interpolated per soil type on a 1000×1000 m^2 grid and based on data collected from* ***1930*** *to* ***1970*** *(historical situation). N.B. For better visualization data are presented within the original 250×250 m^2 grid.*

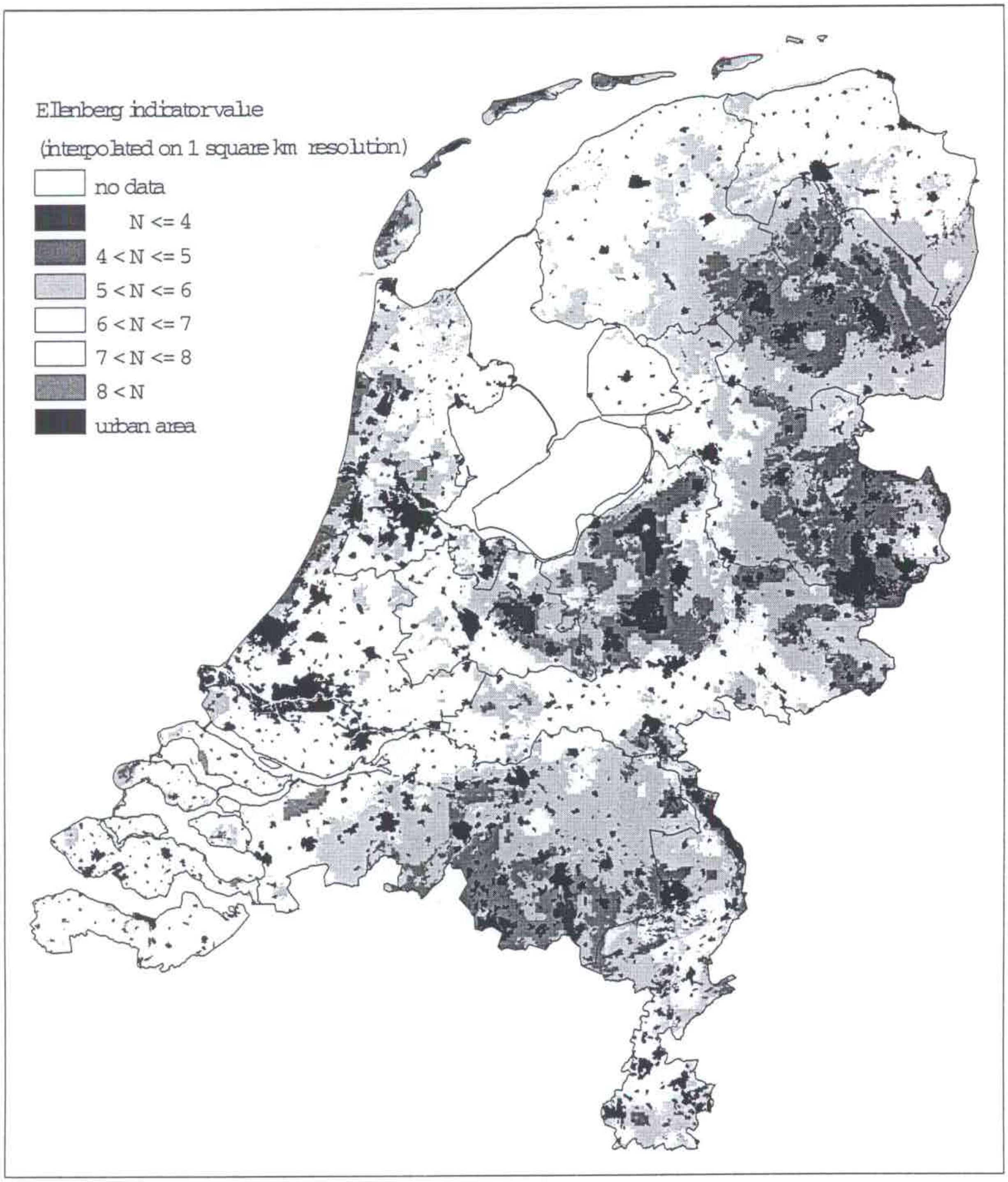

Figure 4.21. *Ellenberg N values, interpolated per soil type on a 1000×1000 m² grid and based on data collected from* ***1930*** *to* ***1970*** *(historical situation). N.B. For better visualization data are presented within the original 250×250 m² grid.*

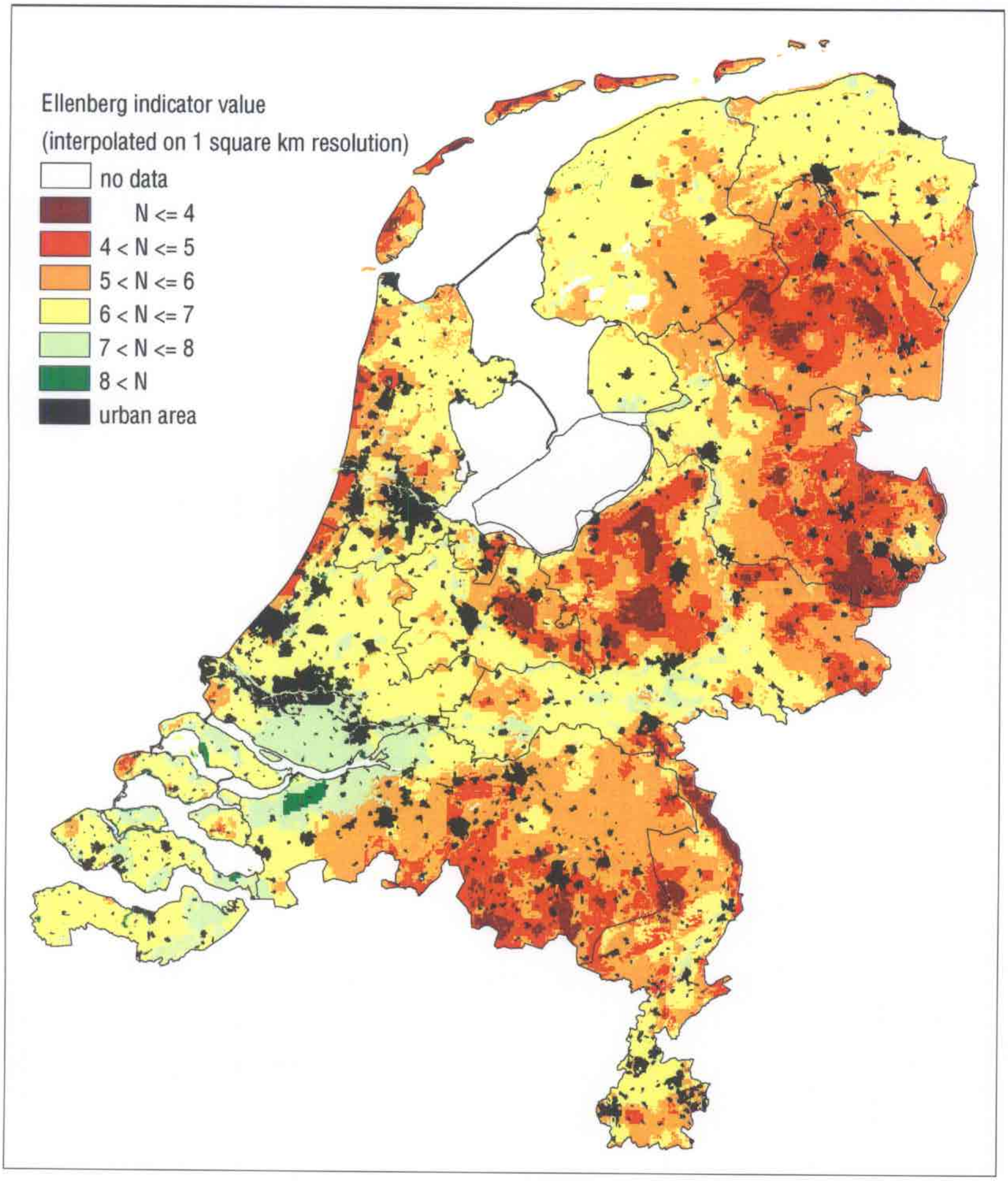
Ellenberg indicator value
(interpolated on 1 square km resolution)
no data
N <= 4
4 < N <= 5
5 < N <= 6
6 < N <= 7
7 < N <= 8
8 < N
urban area

Fig. 4.21

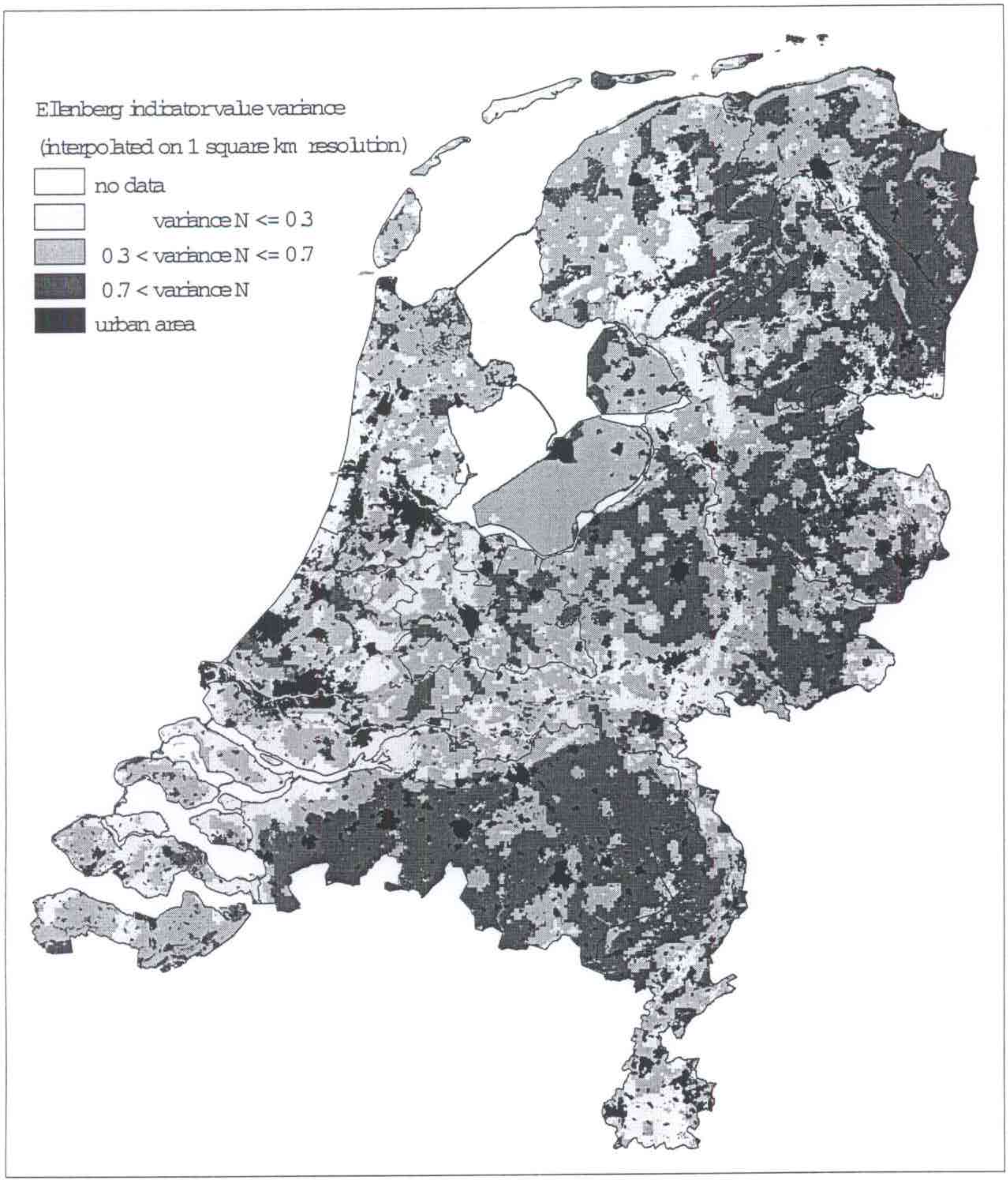

Figure 4.22. *Kriging variance of Ellenberg* ***N*** *values, interpolated per soil type on a 1000×1000 m^2 grid and based on data collected from* ***1930*** *to* ***1970*** *(historical situation). N.B. For better visualization data are presented within the original 250×250 m^2 grid.*

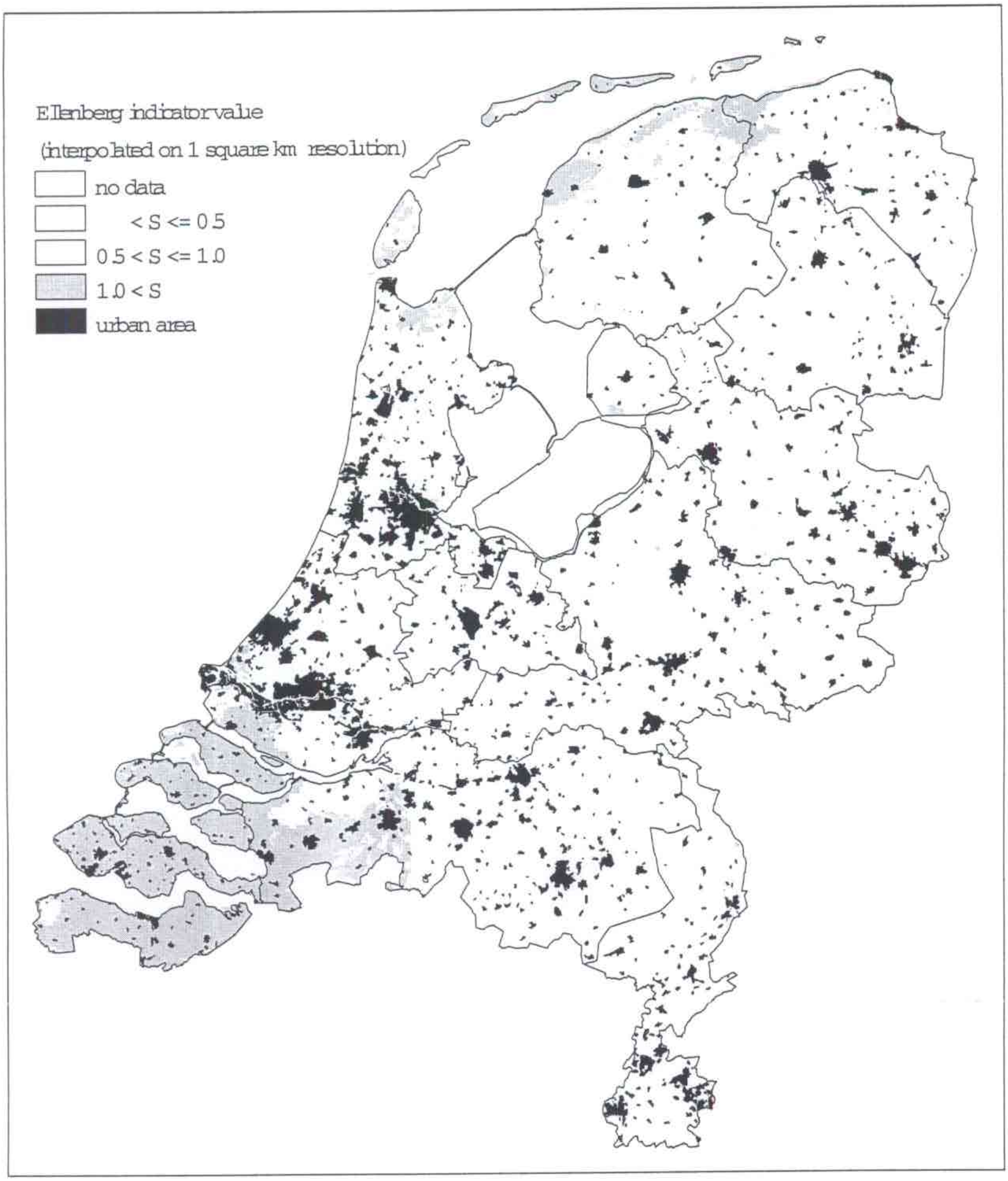

Figure 4.23. *Ellenberg* ***S*** *values, interpolated per soil type on a 1000×1000 m^2 grid and based on data collected from* ***1930*** *to* ***1970*** *(historical situation). N.B. For better visualization data are presented within the original 250×250 m^2 grid.*

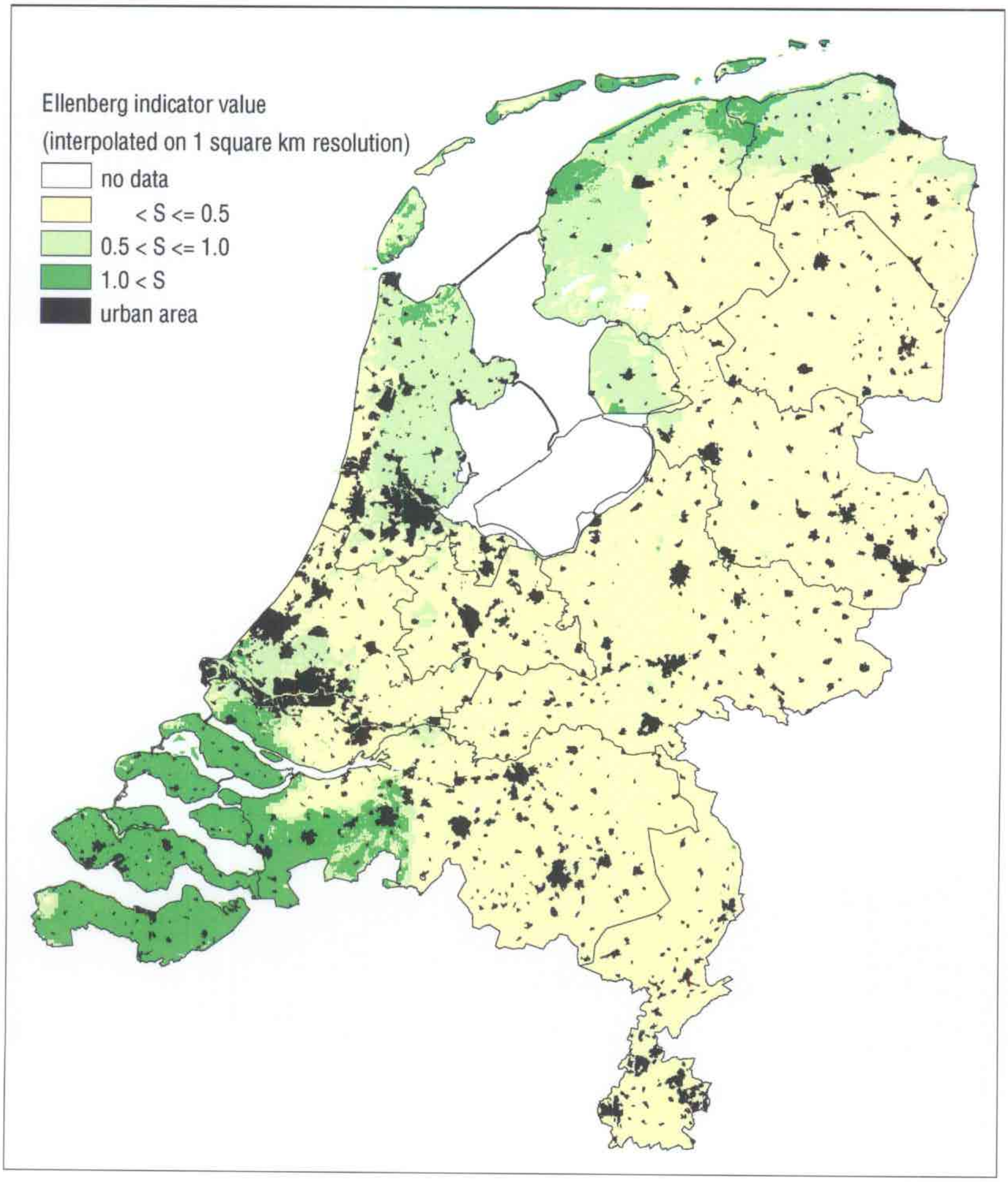
Ellenberg indicator value
(interpolated on 1 square km resolution)
no data
< S <= 0.5
0.5 < S <= 1.0
1.0 < S
urban area

Fig. 4.23

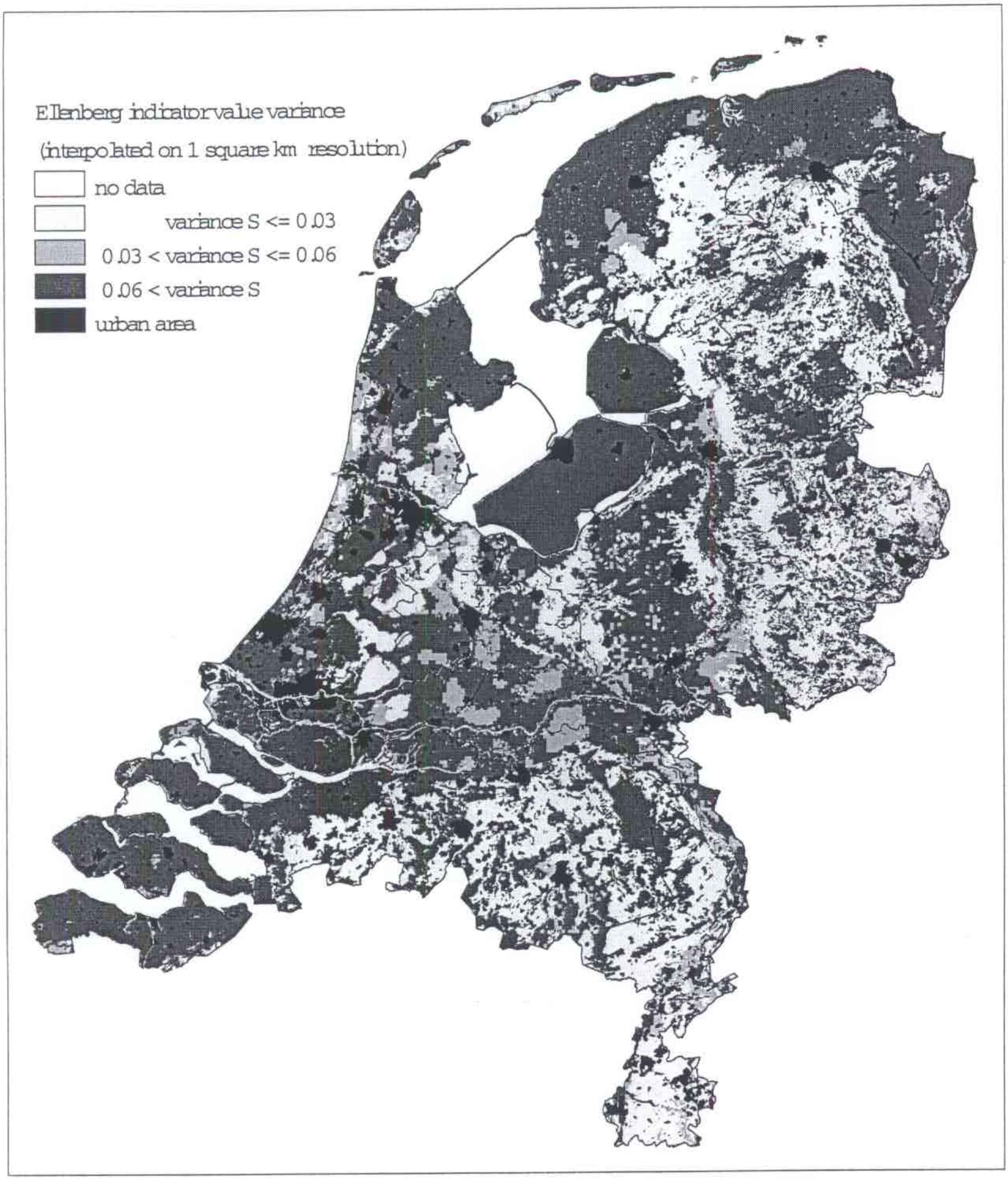

Figure 4.24. *Kriging variance of Ellenberg **S** values, interpolated per soil type on a 1000×1000 m^2 grid and based on data collected from **1930** to **197D** (historical situation). N.B. For better visualization data are presented within the original 250×250 m^2 grid.*

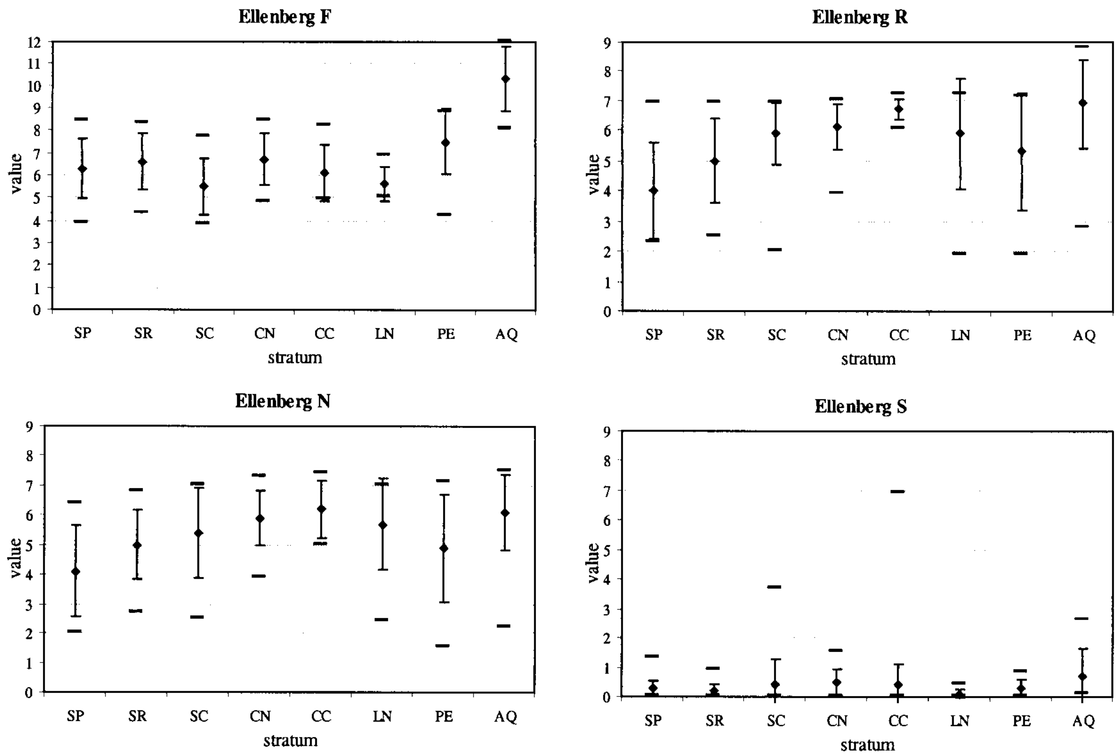

Figure 4.25. Mean (♦), minimum and maximum (–), and twice standard error bars (I) of the interpolated Ellenberg F, R, N and S values per soil type (SP–sand poor, SR–sand rich, SC–sand calcareous, CN–clay non-calcareous, LN–loam non-calcareous, PE–peat, AQ–aquatic) for data collected from ***1991*** *to* ***1990****.*

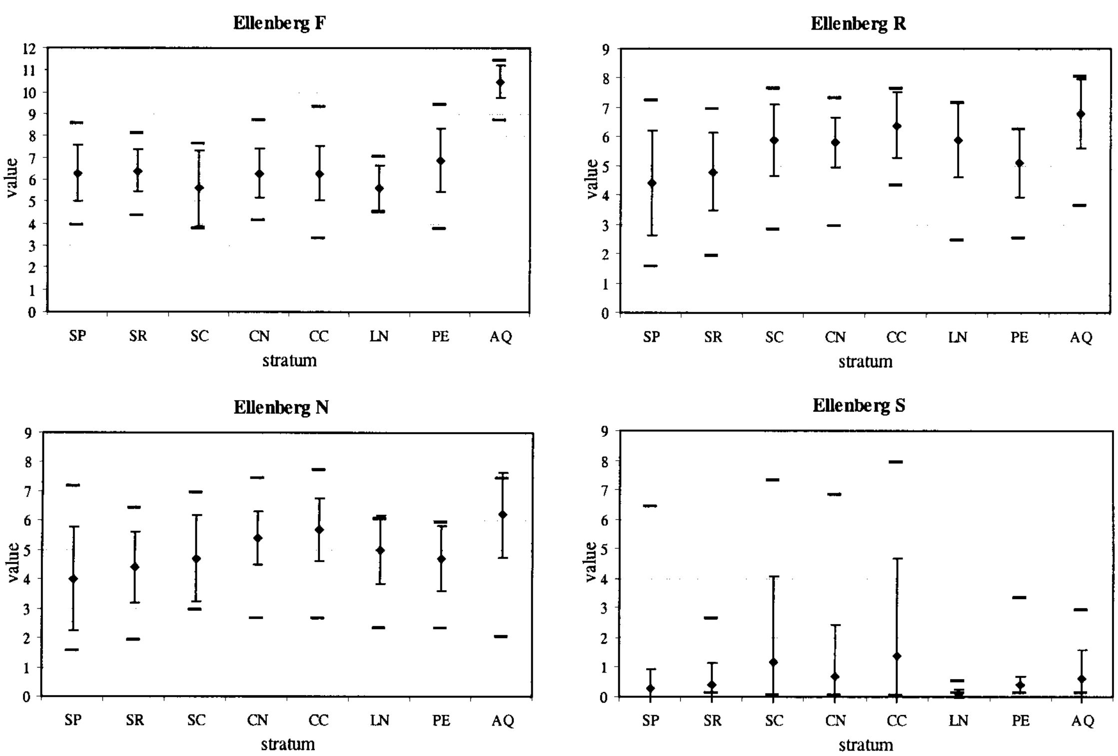

Figure 4.26. Mean (♦), minimum and maximum (–), and twice standard error bars (I) of the interpolated Ellenberg F, R, N and S values per soil type (SP–sand poor, SR–sand rich, SC–sand calcareous, CN–clay non-calcareous, LN–loam non-calcareous, PE–peat, AQ–aquatic) for data collected from ***1930*** *to* ***1970****.*

5. Discussion

Our results suggest that stratified geostatistical interpolation is a feasible and sensible method for the production of abiotic maps on a national scale. Analysis of the spatial structure in our Dutch data revealed moderate to strong spatial dependence in the estimated Ellenberg indicator values, differing according to time periods and strata. This information was successfully used to estimate abiotic site conditions for unsampled sites.

The reference maps produced generally agree with other available map material, like the ecotope map in Witte & Van der Meijden (1992)—which describes environmental acidity—and the ecoseries map of Klijn *et al.* (1992). Reference maps for nitrogen and acidity are the most satisfactory. This is not surprising, as these abiotic variables are well characterized by the soil types used for stratification. Moisture seems less accurately described, although some features are very nicely exposed (for instance, the complex of streams in the NE (Appendix IV.1, Figure A12; reference R1) is clearly visible in the moisture map for the actual time period). It would seem that as terrestrial soil types are not characterized by their moisture, they are not an ideal stratification for this variable. The aquatic stratum works fine, exhibiting excellent semivariograms and adequate estimates. Salinity maps seem locally unreliable, but outstanding features are clearly visible—like the impact of the great flood (in 1957) causing high salinity values for the south-eastern delta area for the historical data. At first sight, reference maps for historical data seem to perform better than those for actual data. Actual maps display more detail but also more interpolation artefacts than the historical ones. Historical maps are on a larger grid, reducing the effect of extreme samples and thus (block) variance. Comparing the historical and actual situations, we notice an apparent reduction in moisture and increase in nutrients. Acidity seems to increase in the SW and decrease in the NW of the Netherlands.

Discrepancies in the site conditions found in other studies (e.g. Witte & Van der Meijden 1992; Klijn *et al.* 1992) can originate in one of the following study phases: *sampling*, *stratification*, *spatial modelling* or *interpolation*. We will present a few illustrative examples of problems found and their possible causes. The geographical location of these examples—referenced as R1 to R14—and the provinces mentioned are found in Appendix IV.1, Figure A12. Readers interested in assessing sample density may copy the sample distribution maps (Appendices I.4 & I.5) onto a transparent sheet to use as an overlay for the abiotic and soil maps.

Sampling: The present study is based on a very large set of floristic samples collected under different conditions. Besides differences in sampling period (season, year), samples also show variable scales (sample size) and densities (in time and per region). Furthermore, samples were gathered by different entities and with different objectives or interests. Records will tend to come from spots attractive to the researcher, because of both their floristic character and accessibility. This is especially true for data collected in past times (in recent decades, there has been an effort to design sampling, aiming at random or stratified random samples). This variability will affect interpolation accuracy. Since the estimation error will, for instance, be larger for floristically less interesting areas, comparison in time will need some correction for the difference in sampling design.

Noord Brabant (Figure A12, R2) is a good example to show errors due to unbalanced sampling. The interpolation result for this region is excessively moist, probably because the most interesting and hence most sampled features are small streams and their margins. *Drenthe* (R3) and *Zeeland* (R4), on the other hand, are estimated to be less moist than expected, possibly also due to unbalanced sampling. In the northern part of the *Veluwe* (R5) a dry region is missing in the actual moisture map. This is apparently due to samples concentrated around the dry area and following some water lines. We expect small-scale variation in the moisture gradient to be most relevant for sandy strata, as these vary topographically within a few hundred metres. In contrast, peat and the aquatic stratum are very homogeneous on small scales and hence likely to be better interpolated.

A detail not captured by the actual salinity map is a salty to brackish area in the NW (R6), which lies next to an area dominated by very fresh water (R7). There are no samples in that particular area, so that its salinity must be determined by surrounding regions only. The actual nitrogen map, on the other hand, seems to miss some agricultural nutrient-rich zones in the east, possibly due to over-sampling of natural environments.

Stratification: Geostatistical interpolation is based on the spatial structure in the data. This structure may differ for different regions, making some form of trend analysis necessary. In the absence of additional information, trend analysis is generally done on the spatial sample co-ordinates. Kriging is a smooth interpolation method, ideal for continuous processes. Kriging on a national scale would lead to interpolation without distinction of dissimilar sites. The result would be a very smoothed, averaged image of abiotic site conditions in the Netherlands, blurring differences between distinct geographical features, like, for instance, aquatic and terrestrial habitats. The Netherlands, by contrast, has very fragmented nature areas. To keep some detail in the abiotic maps and also to better meet the stationarity requirements for the kriging procedure, we require some form of stratification before interpolation.

Stratification was done in time and per soil type. We aggregated samples from a series of years to produce data sets (actual and historical) belonging to a certain time period. Variations within each time period were disregarded. Each data set was subsequently stratified by soil type. Here, misclassifications are unavoidable, especially for the historical data set. For the historical reference maps, samples at only one kilometre precision were attributed all suitable (aquatic *or* terrestrial) soil types occurring in that square kilometre. This use of the same samples for different strata is arguable. Each sample is obviously related to one main stratum—we just do not know which one. In using one sample for several strata, its information is blown up and interpolation variances (probably) underestimated. Other approaches are possible for this problem of scale. We could resample the Dutch soil map on a square kilometre grid and attributing each sample the most frequent dominant soil type of that grid. Just as with the method adopted here, most of the attributed soil types would probably be mistaken, and we would sacrifice the more detailed information available for the samples collected at tenth kilometre precision. We could carry out a conditional simulation (Burrough & McDonnell 1998) using the distribution of soil types in each square kilometre as input—a very time consuming operation considering the amount of grids cells to be interpolated. There is a lack of information (detail) for our historical data which is difficult to make up for.

The soil-type dependent stratification also has its drawbacks. It is a very sensible stratification for acidity/alkalinity and nutrient conditions. After all, these conditions form criteria for the definition of the soil types used. There is no clear relationship between soil type and salinity. Another stratification for this abiotic variable and, possibly, also for moisture might produce better maps.

Spatial modelling: There are several aspects in the spatial modelling process that may have contributed to inaccuracies in the abiotic maps. Firstly, there is the assumption that data of the one time period and soil type are homogeneous in their spatial structure. This generalization is not always appropriate. The semivariogram for non-calcareous loam, for instance, is generally scattered at lags larger than 15 kilometres. This soil type is mainly restricted to the southern *Limburg* region (R8), occupying a small, connected area. The diverging semivariances are caused by a few scattered patches occurring at further distances, which seem to have a different character. Restricting the spatial analysis to data from the southern area would probably improve our semivariogram but also reduce the number of sample pairs, which is already critical for some lags, as semivariances calculated with less than 30 sample pairs are generally considered unreliable (Haining 1990).

Then there are possible directional variations in spatial structure—anisotropy—which are not considered. Spatial structure was modelled on the basis of isotropic semivariogram analysis; i.e. all our semivariograms were calculated omnidirectionally, averaging the spatial structure over all directions. On local scale anisotropy is surely important, for example for linear features like riversides. At a national scale and with the rough stratification used, directional variation in spatial dependence is averaged out, as it is different in different parts of each stratum. Possible anisotropy will contribute to the interpolation error.

Finally, there is the choice of variogram model. For each empirical semivariogram there are frequently several possible and plausible semivariogram models, differing in shape and range. Whenever two or more semivariogram models were equally well fitted, we chose the most informative on our problem scale. Since we aimed at local interpolation, we preferred extracting shorter distance spatial structure and personal experience. Pure nugget variograms were avoided all together, as they imply no spatial dependence at all and would lead to attribution of the stratum average to all samples—which is not very sensible for our purpose.

Interpolation: Interpolation assumed a certain homogeneity of spatial structure per stratum. To reduce the effect of possible trends and regional variances not distinguished in our stratification (e.g. north to south), we opted for local kriging interpolation (the limited radius and maximum number of neighbours also helped to keep calculation time reasonable). The example of southern *Limburg* (R8) cited above demonstrates that interpolation was not always local enough. Other examples are some areas of calcareous clay—in NE *Groningen* (R9), SE *Flevoland* (R10) and northern *Noord Holland* (R11)—which are undersampled (as they are of little floristic interest) and display rather elevated nitrogen estimates. These estimates probably originated from the samples in the *Biesbosch* (R12) and borders of *Flevoland* (R13). Nitrogen in *Flevoland* itself is a good example of interpolation artefacts, showing a clear-cut class limit caused by a few sample points and the interpolation neighbourhood radius chosen.

For normally distributed data, the linear kriging predictor is optimal and the estimation errors are normally distributed. When non-normally, or even non-symmetrically, distributed data are submitted to ordinary kriging, their error is averaged and variances change along the data range. Our data—especially the estimate for salinity—are frequently skewed. Consequently, reference maps for salinity are not very reliable. This is not the place to discuss the nature of Ellenberg indicator values, or to argue about their applicability as (continuous) estimates of environmental variables. Ellenberg values are defined on an ordinal scale. They are not continuous and calibration results suggest that their relation to associated environmental factors is not necessarily linear (Wittig *et al.* 1985; Ertsen *et al.* 1998). This is especially true for Ellenberg S in relation to salinity. To reduce skewness, a transformation, e.g. using the logarithm, could be considered. But we have to keep in mind that (block) estimates and variances are difficult to backtransform to the original scale.

With respect to the kriging variance, we noticed that departure from the expected values was not always accompanied by elevated kriging variances. For instance, moisture estimates for the elevated boggy ('hoogveen') areas in the NE (R14) were rather dry. Yet, although this region is scarcely sampled in both time periods, kriging variance is small. The small variances in this area belong to the stratum 'sand poor'. This stratum, which is the most abundant, displays little variation in moisture (Figures 4.25 and 4.26), and its respective semivariogram is dominated by a huge nugget effect in the actual data set, and has a small range in the historical data set. All these aspects lead to a low estimated error when the mire region is 'filled up' during interpolation. Apparently wet samples are underrepresented in that stratum.

Finally, a note on the future use of this study and its results. The aim of the present study was to propose and apply a geostatistical method for interpolation of abiotic site conditions estimated from a large set of vegetation records. Consequently, it is mainly focussed on the methodology for spatial analysis and geostatistical interpolation. Application of the resulting maps and possible comparative studies of the temporal changes of abiotic site conditions in the Netherlands should therefore be done with care and after correcting for differences in scale, sampling and interpolation artefacts. It is clear that our reference maps—although overall satisfactory—need further detailed examination, using additional available information from other sources (regional or local abiotic data). Such an analysis may also improve stratification, resulting in a more accurate image of the fragmented reality of Dutch environment.

But even after corrections we must not forget that these maps are interpolated. It is current practice to use estimated averaged Ellenberg indicator values in the context of ecological (regression-based) vegetation models. And it is tempting to apply interpolated values—like those produced in this study—directly to model vegetation distribution on a national scale. But these values are (interpolation) estimates of (Ellenberg indicator value) estimates, with a more-or-less accentuated error component defined by the spatial model in which the data themselves are considered to be one of many possible realizations of a particular spatial process. To develop ecological models based on such interpolated maps, one should, for instance, perform simulation studies, defining the possible distribution of abiotic values for each site and the respective distribution of vegetation response.

Summarizing

- Local ordinary (block) kriging is a feasible and sensible geostatistical method for spatial interpolation of environmental data on a national scale. It allows us to produce estimates and their respective errors (variances) on different scales.
- The abiotic reference maps produced are overall satisfactory, but discrepancies found should be corrected prior to any future application.
- Accuracy of the interpolation results is mainly influenced by sampling design, stratification, and the assumptions and choices made for spatial modelling. Hence, inferences about the results and, especially, comparison of abiotic site conditions in time will require detailed analysis and, if necessary, corrections for these factors.

Acknowledgements. This study was funded by the Dutch National Institute of Public Health and Environment as part of the project 'Multistress Effects in Ecology' (no. 712910). It took place in coorperation with the Department of Environmental Sciences of the Utrecht University. We are indebted to Arjen van Hinsberg for his thorough analysis and comparison of the results with current knowledge of the Dutch environmental situation. His remarks form a substantial part of the discussion. Furthermore, we would like to thank Camiel Heunks for his GIS support and the production of most of the maps in this report. Finally, we would like to thank J.H.J. Schaminée and S.M. Hennekens, who kindly allowed us to use their vast database of relevés.

References

Cressie, N.A.C. (1993). *Statistics for spatial data.* Revised edition. John Wiley & Sons.

Briemle, G. (1986) Vergleich der Stickstoff-Mineralisation mit der N-Zahl Ellenberg's am Beispiel einer Streuwiese im Alpenvorland. Erste Erfahrungen mit zweijaerigen N min-Untersuchungen. *Natur und Landschaft* 61: 423-427.

Burrough, P.A. & McDonnell, R.A. (1998) *Principles of geographical information systems.* Oxford University Press, Oxford.

Ellenberg, H.; Weber, H.E.; Düll, R.; Wirth, V.; Werner, W. & Paulißen, D. 1991. Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica* 18: 1-248.

Ertsen, A.C.D.; Alkemade, J.R.M.; Wassen, M.J. (1998) Calibrating Ellenberg indicator values for moisture, acidity, nutrient availability and salinity in the Netherlands. *Plant Ecology* 135: 113-124.

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences.* Cambridge University Press.

Journel, A.G. & Huijbregts, C.J. (1978) *Mining geostatistics.* Academic Press, London.

Kitanidis, P.K. (1983) Statistical estimation of polynomial generalized covariance functions and hydrological applications. *Water Resources Research* 19: 909-921.

Klijn, F.; Ten Harmsel, A. & Groen, C.L.G. (1992) Ecoseries 2.0. Naar een ecoserieclassificatie t.b.v. het ecohydrologisch voorspellingsmodel DEMNAT-2. Centrum voor Milieukunde Rijksuniversiteit Leiden, CML-rapport nr.85.

Köhl, M. & Gertner, G. (1997) Geostatistics in evaluating forest damage surveys: considerations on methods for describing spatial distributions. *Forest Ecology and Management* 95: 131-140.

Matheron, G. (1971) The theory of regionalized variables and its applications. *Cahiers du Centre de Morphologie Mathématique de Fontainebleau*, No. 5, Fontainbleau, France.

Nováková, J. (1997) Agricultural impact on nonlinear vegetation formations: Species richness—stand trophy relations. *Ekológia (Bratislava)* 16: 233-241.

Pebesma, E.J.; Wesseling, C.G. (1998) GSTAT: a program for geostatistical modelling, prediction and simulation. Computers & Geosciences 24: 17-31

Pebesma, E.J. (1997) Gstat user's manual. Available at: http://www.geog.uu.nl/gstat/.

Schaminée, *et al.* (1995-1999) *De vegetatie van Nederland.* Parts 1 to 5. Opulus press, Leiden.

Ter Braak, C.J.F. & Barendregt, L.G. (1986) Weighted averaging of species indicator values: its efficiency in environmental calibration. *Mathematical Biosciences* 78: 57-72.

Ter Braak, C.J.F. & Gremmen, N.J.M. (1987) Ecological amplitudes of plant species and the internal consistency of Ellenberg's indicator values for moiture. *Vegetatio* 69: 79-87.

Ter Braak, C.J.F. & Wiertz, J. (1994) On the statistical analysis of vegetation change: a wetland affected by water extraction and soil acidification. *Journal of Vegetation Science* 5: 361-372.

Witte, J.P.M. & Van der Meijden, R. (1992) Verspreiding en natuurwaarden van ecotoopgroepen in Nederland. Landbouwuniversiteit Wageningen & Rijksherbarium / Hortus Botanicus, Leiden.

Wittig, R., Ballach, H.-J. & Brandt, C.J. (1985) Increase of number of acid indicators in the herb layer of the Millet Grass-Beech forest of the Westphalian bight. *Angewandte Botanik* 59: 219-232.

Appendix I.1 Original databases

Table A.1. *Original 53 DBASE data bases supplied by the IBN (Schaminée* et al. *1995-1999)*

ariette.dbf	groninge.dbf	overijss.dbf
bes.dbf	hettie1.dbf	piet.dbf
bosstat.dbf	hettie2.dbf	provinci.dbf
brabant.dbf	jan1.dbf	pwn.dbf
caroline.dbf	jan2.dbf	ron1.dbf
charlot.dbf	jan3.dbf	rws_md.dbf
cml.dbf	joop.dbf	sandra.dbf
deltaarc.dbf	kiwa.dbf	sbb.dbf
dijken.dbf	limburg.dbf	stephan.dbf
divers1.dbf	londopq.dbf	vechtpl.dbf
ecoburo1.dbf	marcel1.dbf	voornepq.dbf
ecoburo2.dbf	marcel2.dbf	wegberm.dbf
ecoburo3.dbf	marcel3.dbf	werf.dbf
ed.dbf	marian.dbf	westhof4.dbf
eddy.dbf	marniks.dbf	westhoff.dbf
gelderla.dbf	mayendel.dbf	zuidholl.dbf
giesgeur.dbf	mayendpq.dbf	_oevers.dbf
gooi.dbf	nat_monu.dbf	

Each of these databases contains:

TVABUND — identified plant species; fields:

RELEVE_NR, SPECIES_NR, COVER_CODE, COVER_PERC

TVHABITA — site attributes and estimated abiotic site conditions; fields:

RELEVE_NR, REFERENCE, COVERSCALE, PROJECT, AUTHOR, DATE, KM_HOK_X, KM_HOK_Y, BLOCK, SYNTAXON, NEW_SYNTAX, LENGTH, WIDTH, SURF_AREA, EXPOSITION, INCLINATIO, COV_TOTAL, COV_TREES, COV_SHRUBS, COV_HERBS, COV_MOSSES, COV_ALGAE, COV_LITTER, TREE_HIGH, TREE_LOW, SHRUB_HIGH, SHRUB_LOW, HERB_HIGH, HERB_LOW, HERB_MAX, MOSS_IDENT, PQ, TRANSECT, SBBCODE, REMARKS, LIJNVORM, REGIO, ECO1, ECO2, ECO3, ECO4, N_AANT_H, N_GEM_H, N_GEM_ALL, PH_AANT_H, PH_GEM_H, PH_GEM_ALL, V_AANT_H, V_GEM_H, V_GEM_ALL, L_AANT_H, L_GEM_H, L_GEM_ALL, S_AANT_H, SAL_GEM_H, SAL_GEM_AL, IPI

REMARKS — incomplete & irrelevant for the present study

Appendix I.2 Data pre-processing: corrections and selections

Table A.2. *Corrections applied to the original IBN data*

Relevé	Data subset	Field / Original	Field / Corrected
72205	roadside	date / 1907	year / 1987
85002–86097	dykes	date / NA	year / 1999 *
220936	rws_md	date / 1919	year / 1992
220942	〃	date / 1919	year / 1992
220945	〃	date / 1919	year / 1992
220952	〃	date / 1919	year / 1992
221947	〃	date / 1920	year / 1992
222175	〃	date / 1920	year / 1992
222243	〃	date / 1921	year / 1992
222250	〃	date / 1921	year / 1992
224646	〃	date / 1919	year / 1990
224651		date / 1919	year / 1990
222402	rws_md	date / 10-9-92	year / 1992
307740–308412	ecoburo3	date / 1901	*omitted*
301070	ecoburo3	date / 1909	year / 1989
401344	limburg	date / 19	year / 1991
95482	divers1	x, y / 409.8, 86.5	x, y / 86.5, 409.8
95483	〃	x, y / 409.8, 86.5	x, y / 86.5, 409.8
95533	〃	x, y / 535.0, 214.4	x, y / 214.4, 535.0
56059	ecoburo1	y / 929.2	y / 429.2
95514	divers1	y / 800.8	y / 600.8
54027–54279	ecoburo1	x, y / *in hundreds of km*	x, y / *in km (i.e. ×100)*
54281–54306	〃	〃	〃
54308–54358	〃	〃	〃
54455–54458	〃	〃	〃
28121	jan2	x, y / 1, 0	x, y / 0, 0
7482	marcel1	x, y / 102, 610	x, y / NA
94382	divers1	x, y / 674.1, 566.5	x, y / NA *(Germany)*

* Collected after 1990; coded '1999'; Notes are in *italics*.

For a fair number of samples the 'Atlasblok' co-ordinates do not correspond to the 'Amersfoort' kilometre co-ordinates. Differences of up to 14 km were found. Furthermore, while the last two digits of 8-digit block co-ordinates are generally the tenth kilometre in WE and NS direction (i.e. x, y), there are numerous samples with inverted 7th and 8th digits (i.e. y, x). The decisions to accept one or the other location specification or to dismiss both, as well as the interpretation of the tenth kilometre co-ordinates for 8-digit block-co-ordinates were made by Stefan Hennekens (IBN; tel. 0317-477908). These are summarized in text files collected under `ibncorre.zip`, and were all processed in `tvana.mdb` (tables `tvana` & `tvana2`).

The resulting tables in ACCESS files `turboveg.mdb` and `tvana.mdb` are:

`tvbasis`	52 data subsets (excl. banks and linear features) 160,252 records original data
`tv_tot`	53 data subsets (incl. banks; excl. linear features) 169,202 records original data
`tvana`	53 data sub sets (incl. banks; excl. linear features) corrected for typing errors; selected columns; YEAR (instead of DATE) 169,202 records
`tvana2`	from tvana X & Y (in meters); combined from block and km-co-ordinates * omitted missing values for YEAR and co-ordinates new columns: HAB = habitat (aquatic; terrestrial) HUM = humidity NUT = nutrients 129,090 records

* The centre co-ordinates of the respective square kilometre are assigned to samples with location specified with kilometre precision only.

Appendix I.3 Ecotypes

Table A.3. *Ecotypes coded in field ECO*

Chlorinity	**Vegetation structure**	**Moisture**	**Nutrients & acidity**	**Additional attributes**
– fresh **b** brackish s salt	**P** pioneer veg. **G** grassland **R** rough grassl. **S** shrubs **B** forest **V** waterside v. **W** water veg.	**1** aquatic **2** wet **3** very moist **4** moist **5** moderately moist **6** dry	**1** nutr. poor/acid **2** nutr. p. / slightly acid **3** nutr. p./alkal. **4** nutr. poor **5** nutr. poor/acid **6** sl. nutr. rich / alkal. **7** moder. n. rich **8** very n. rich **9** nutr. rich	**ho** **st** **dw** **tr** **ro** **sa**
prefix *	letter	1^{st} digit	2^{nd} digit	suffix *

* Optional

Appendix I.4 — Sample distribution for the actual data set

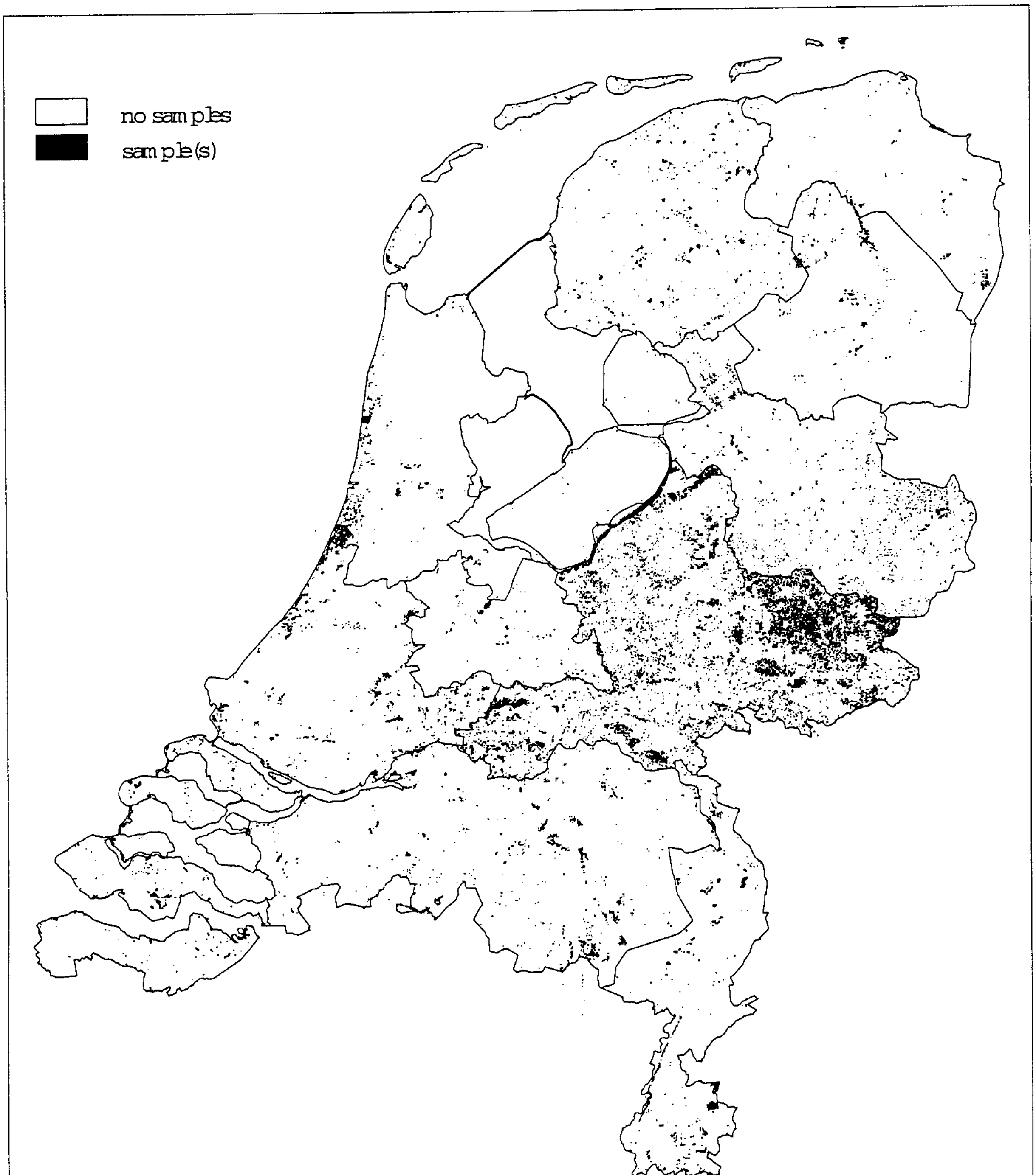

Figure A.1. Sample sites of the actual data set (collected from 1991 to 1997). Samples are displayed on a 250×250 m^2 grid. Each grid can contain numerous samples.

Appendix I.5 — Sample distribution for the historical data set

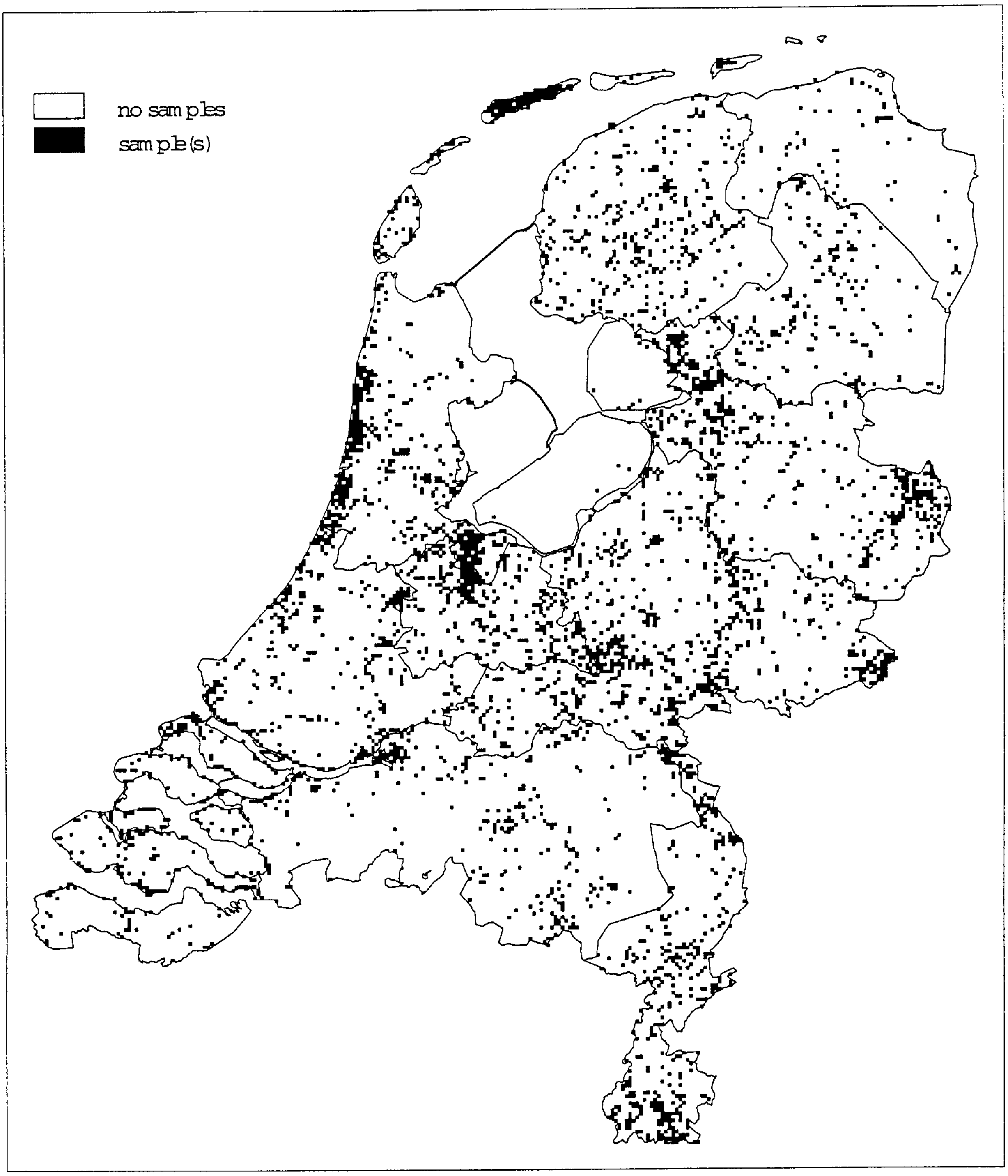

Figure A.2. *Sample sites of the historical data set (collected from 1930 to 1970). Samples are displayed on a 1000×1000 m^2 grid. Each grid can contain numerous samples.*

Appendix I.6 — Soil map used for data stratification

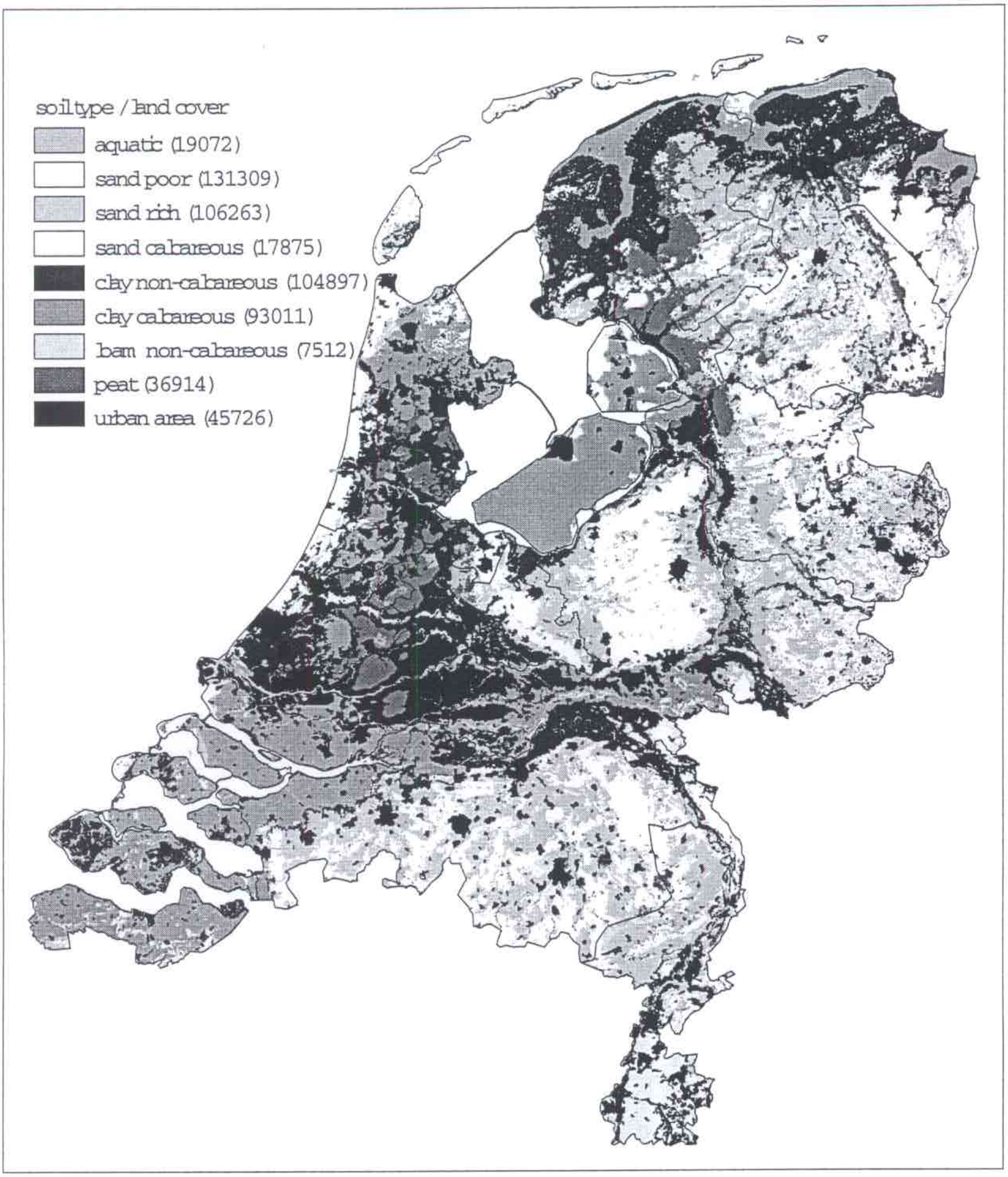

Figure A.3. *Soil map of the Netherlands on a 250×250 m² grid, with the strata used for spatial interpolation of abiotic site condition. Soil types include seven terrestrial and one aquatic stratum, as well as urban areas (with the respective number of grid cells in parentheses). The original coverage, received from the DLO-Staring Centrum in 1991, was gridded and adapted by Michel Bakkenes & Mariette van Esbroek (RIVM) in 1998.*

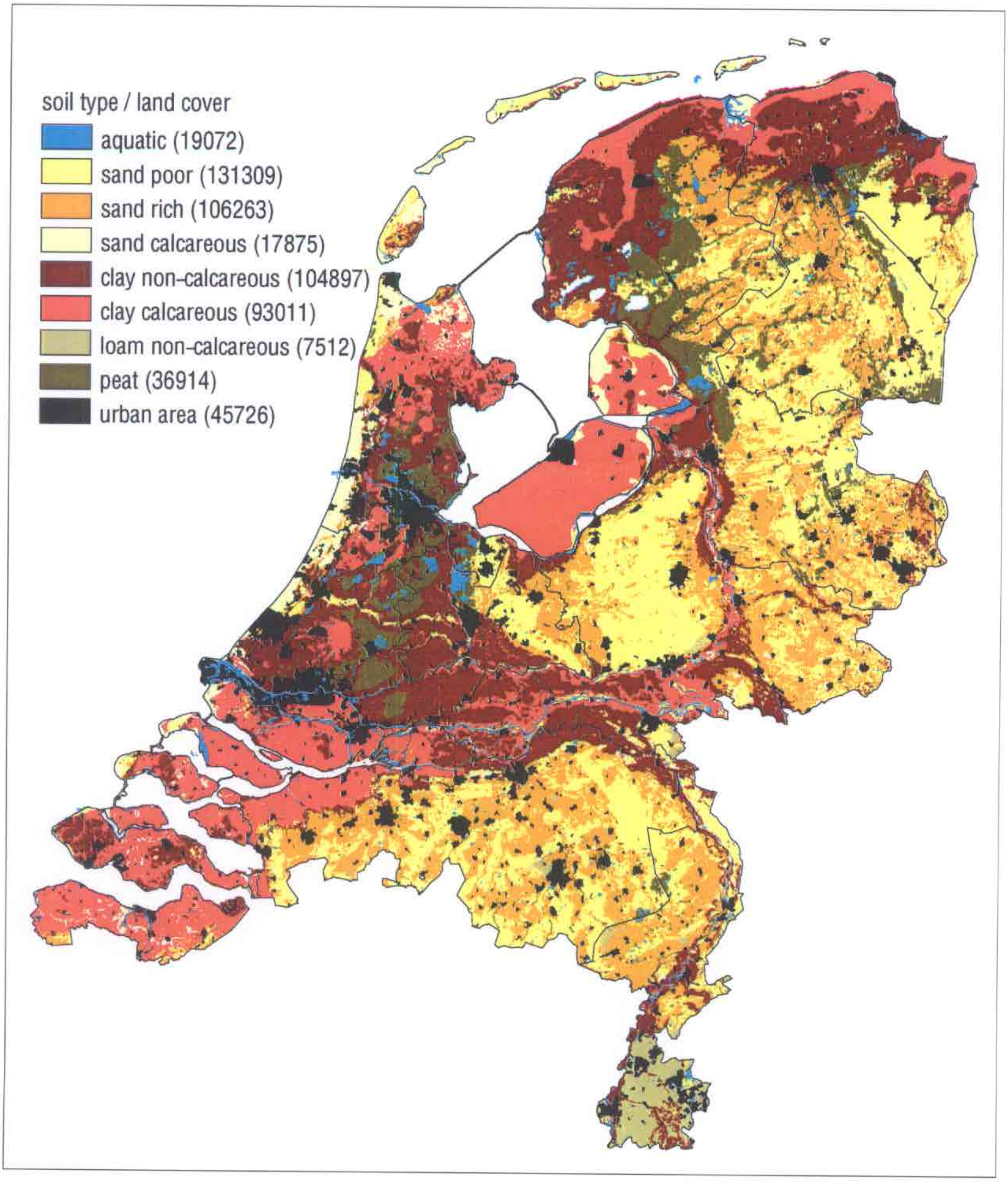
soil type / land cover
aquatic (19072)
sand poor (131309)
sand rich (106263)
sand calcareous (17875)
clay non-calcareous (104897)
clay calcareous (93011)
loam non-calcareous (7512)
peat (36914)
urban area (45726)

Fig. A3

Appendix I.7 — Samples used for interpolation

Table A.4. *Number of records used for the interpolation of the abiotic reference maps, distinguished per time period, soil type and abiotic variable*

Time period	Stratum	Ellenberg indicator value
actual data = 39,389	sand poor = 8945	F = 8115 N = 8624 R = 8157 S = 8628
	sand rich = 7043	F = 6779 N = 6894 R = 6682 S = 6901
	sand calcareous = 2563	F = 2524 N = 2542 R = 2517 S = 2544
	clay non-calcareous = 5416	F = 5340 N = 5362 R = 5248 S = 5367
	clay calcareous = 2177	F = 2177 N = 2178 R = 2136 S = 2179
	loam non-calcareous = 337	F = 311 N = 321 R = 314 S = 321
	peat = 2859	F = 2810 N = 2839 R = 2740 S = 2840
	aquatic = 7571	F = 6697 N = 6359 R = 6476 S = 5643
historical data = 16,974	sand poor = 4928	F = 4645 N = 4799 R = 4723 S = 4808
	sand rich = 2474	F = 2394 N = 2440 R = 2410 S = 2441
	sand calcareous = 2948	F = 2860 N = 2880 R = 2865 S = 2889
	clay non-calcareous = 4408	F = 4303 N = 4338 R = 4258 S = 4343
	clay calcareous = 3450	F = 3375 N = 3399 R = 3321 S = 3406
	loam non-calcareous = 632	F = 603 N = 616 R = 614 S = 616
	peat = 2681	F = 2633 N = 2624 R = 2642 S = 2655
	aquatic = 3260	F = 3202 N = 3189 R = 3123 S = 3201

The distribution of the records used for the interpolation of the abiotic reference maps is presented in the following histograms according to time period, abiotic variable and soil type.

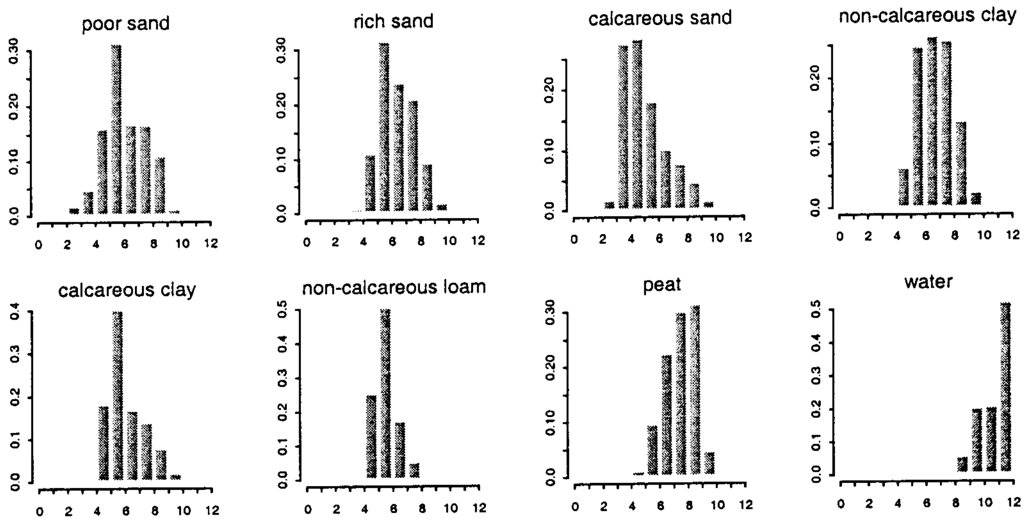

Figure A.4. *Relative occurrence of the averaged Ellenberg F value for samples of the actual data set (collected from 1991 to 1997).*

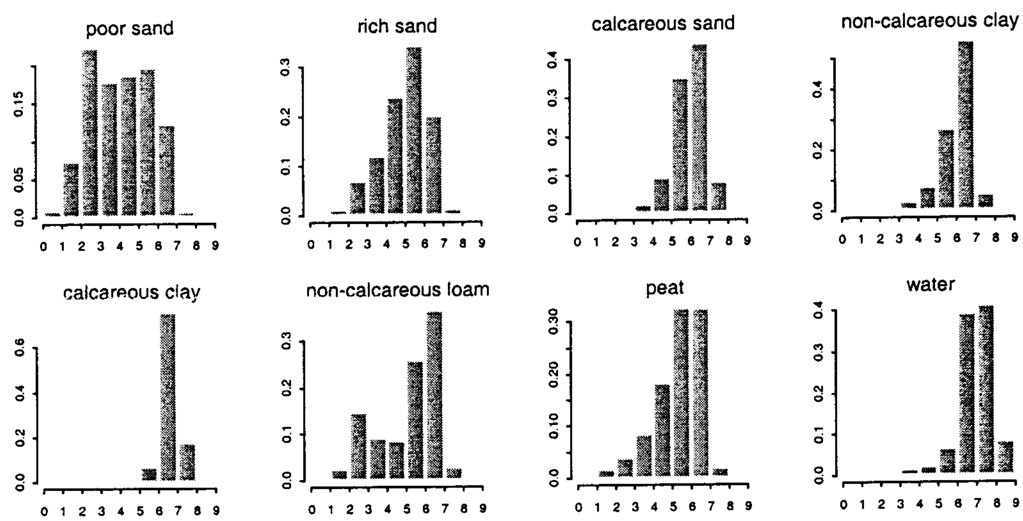

Figure A.5. *Relative occurrence of the averaged Ellenberg R value for samples of the actual data set (collected from 1991 to 1997).*

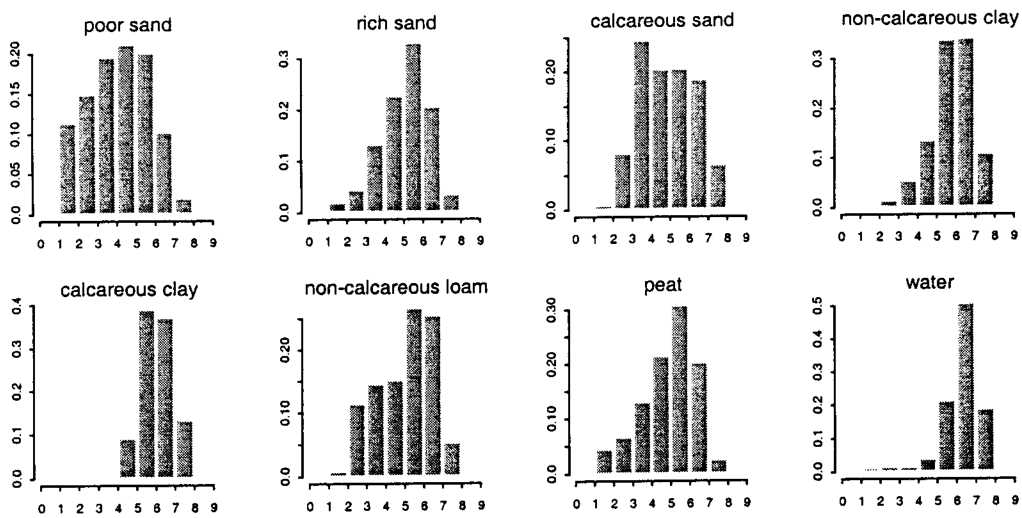

Figure A.6. Relative occurrence of the averaged Ellenberg N value for samples of the actual data set (collected from 1991 to 1997).

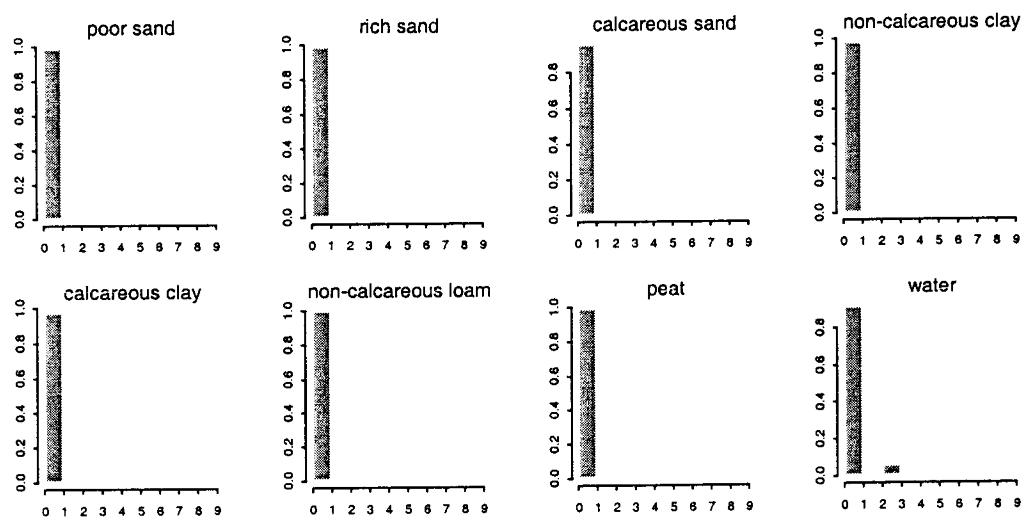

Figure A.7. Relative occurrence of the averaged Ellenberg S value for samples of the actual data set (collected from 1991 to 1997).

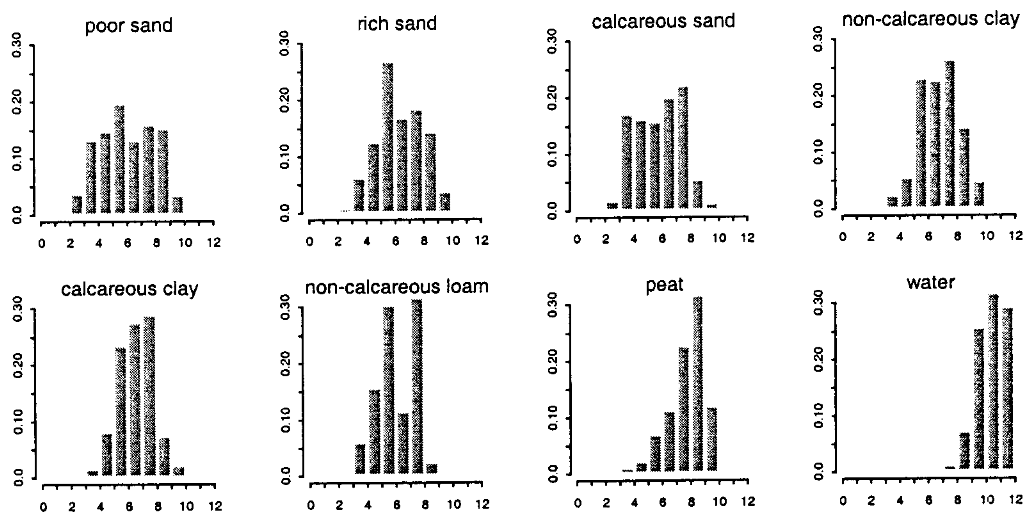

Figure A.8. *Relative occurrence of the averaged Ellenberg F value for samples of the historical data set (collected from 1930 to 1970).*

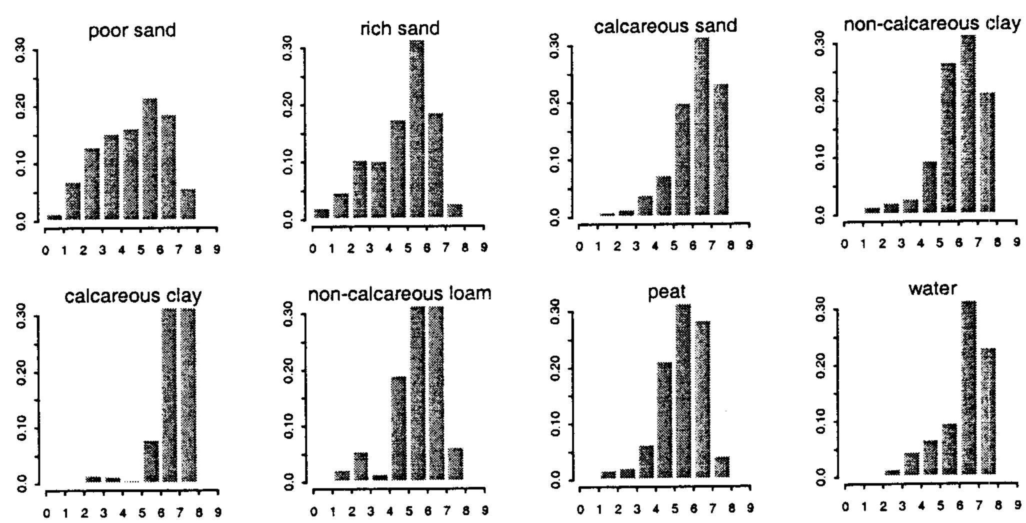

Figure A.9. *Relative occurrence of the averaged Ellenberg R value for samples of the historical data set (collected from 1930 to 1970).*

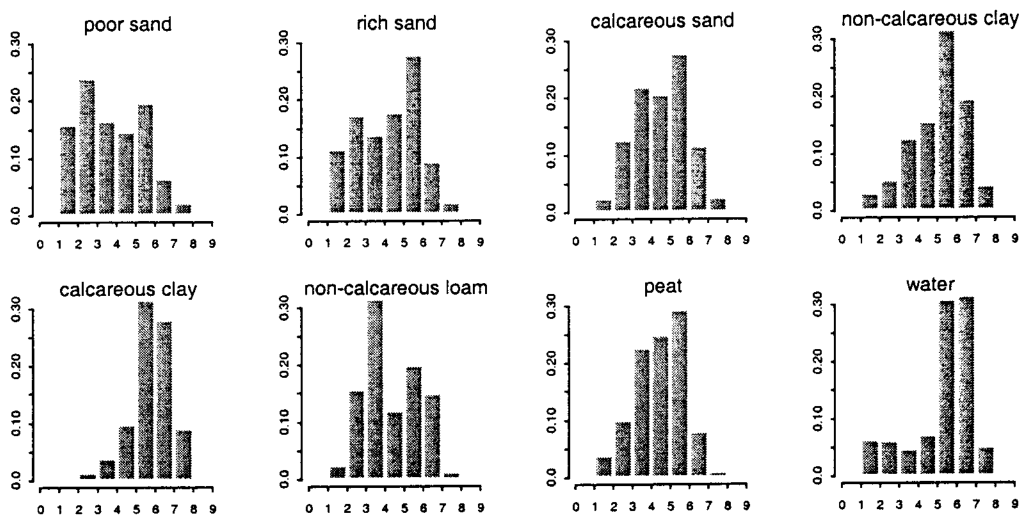

Figure A.10. *Relative occurrence of the averaged Ellenberg N value for samples of the historical data set (collected from 1930 to 1970).*

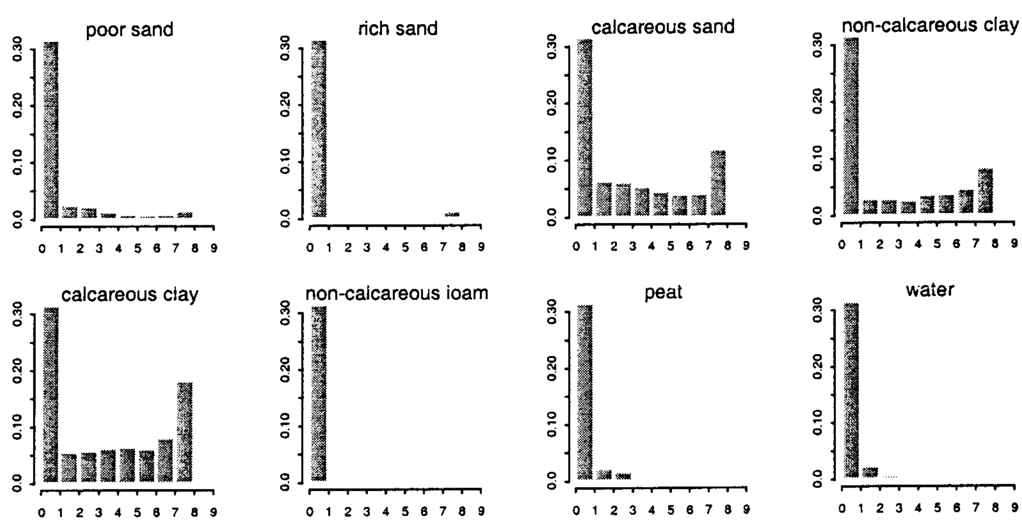

Figure A.11. *Relative occurrence of the averaged Ellenberg S value for samples of the historical data set (collected from 1930 to 1970).*

Appendix II.1 — Examples of GSTAT command and output files

Example of Gstat's command file for semivariogram modelling; semivariogram model for Ellenberg F value for poor sand soil:

```
#
# gstat command file, HP-UX version 2.0f (July 1998)
# Mon Oct 26 16:29:12 1998
#
data(F1): 'f1', x=2, y=3, v=4, average;
variogram(F1): 1.37895 Nug(0) + 0.505296 Gau(30531.1);
set cutoff = 40000;
set fit = 2;
set width = 2000;
```

Example of Gstat's estimates output; semivariogram estimates for Ellenberg F value for poor sand soil:

```
#gstat HP-UX 2.0f (July 1998) [gstat start.cmd]
#sample semivariogram
#Mon Oct 26 16:29:11 1998
#data(F1): 'f1', x=2, y=3, v=4, average;
#[1] mean: 6.11399 variance: 1.97906
#cutoff: 40000 interval width: 2000
#direction: total
#   from      to  n_pairs  av_dist semivariogram
       0    2000   143472  1049.99  1.38039
    2000    4000   160060  3059.93  1.39825
    4000    6000   226090   5052.2  1.34653
    6000    8000   280985  7031.05  1.39098
    8000   10000   331266  9014.55   1.4292
   10000   12000   366279    11017   1.4494
   12000   14000   390923    13010  1.48959
   14000   16000   435391  15005.3  1.48976
   16000   18000   452198  17007.5   1.5353
   18000   20000   445771  18992.3  1.55615
   20000   22000   435251  20997.6  1.55788
   22000   24000   448603  23007.1  1.57545
   24000   26000   426837  24984.2  1.59738
   26000   28000   398837  26990.9  1.61625
   28000   30000   396852  28993.4  1.72183
   30000   32000   373942  30991.6  1.71579
   32000   34000   382461    33006  1.69975
   34000   36000   380096  34996.9  1.72043
   36000   38000   388200  37008.9  1.75773
   38000   40000   408170  39011.7  1.84365
```

Example of Gstat's command file for local ordinary block kriging; interpolation of Ellenberg F value for poor sand soil; on a 250 m^2 grid:

```
#
# gstat command file, HP-UX version 2.0f (July 1998)
# Mon Oct 26 18:25:33 1998
#
data(F2): 'f1', x=2, y=3, v=4, average, max=50, radius=35000;
variogram(F1): 1.37895 Nug(0) + 0.505296 Gau(30531.1);
data():  'bod1', x=1, y=2;
blocksize: dx=250, dy=250;
set output = 'F1.kri';
```

Appendix II.2 — Example of a GEOVIEW krt-file

GEOVIEW krt-file for mapping the actual situation of Ellenberg F value:

```
/*****************************************************************************
/* Ellenberg-F, actual time period
/* kriging interpolation per soil type
/*****************************************************************************
kaart                    = "kaart 1"
  titel                  = "IBN Data 1991 t/m 1997 (250x250m)"
  subtitel               =
  altmapex               =
  datum                  = nee
  kaartkader             = ja
  kaartnr                =
  versie                 = 4.0
  taal                   = nederlands
  layout                 =
  logo                   =
  noordpijl              = nee
  schaalbalk             = nee
  concept                = nee
  berekeningen           = nee
  sort                   = user
  legendatitel           = "Ellenberg indicator value"
  legendavoet            =
  bron                   =
  vedettetekst           =
  figuurtekst            =
  extratekst             =
  kaartlaag              = "thematisch"
    laagtype             = grid
    geodataset           = f_all
    laagvolgnr           = 1
    kleurblok
      classificatie      = standaard
      item               = value
      dataset            =
      relate             =
      legvolgnr          = 1
      legendatype        = standaard
      legendakop         =
      legenda            = 700 0 "no data"
      legenda            = 808 4 "       F <= 4"
      legenda            = 806 5 " 4 < F <= 5"
      legenda            = 804 6 " 5 < F <= 6"
      legenda            = 932 7 " 6 < F <= 7"
      legenda            = 812 8 " 7 < F <= 8"
      legenda            = 852 12 " 8 < F"
      legenda            = 800 112 "urban area"
  eindekaartlaag
  kaartlaag              = "provinciegrenzen"
    laagtype             = polyalsline
    geodataset           = lannladovprv1_95 spacebase
    laagvolgnr           = 3
    kleurblok
      classificatie      = een
      item               =
      dataset            =
      relate             =
      legvolgnr          = 0
      legendatype        = standaard
      legendakop         =
      legenda            = 1 # "border layer 1"
  eindekaartlaag
eindekaart
===========================================================================
```

Appendix III.1 — Semivariogram features

Table A.5. Semivariogram (γ) features for each time period, Ellenberg indicator variable and stratum

Time period	Ellenberg	Stratum	γ type	γ range	γ partial sill	γ nugget
actual data set	F	sand poor	Gau	30531.1	0.505296	1.37895
		sand rich	Exp	10642.8	0.330412	1.08875
		sand calcareous	Gau	20967.6	0.604346	1.54323
		clay non-calcareous	Sph	23878.8	0.276299	0.893392
		clay calcareous	Sph	41467.3	0.433591	1.21801
		loam non-calcareous	Sph	20015.7	0.118284	0.381501
		peat	Gau	17506.9	0.429949	0.644828
		aquatic	Sph	66157.4	1.87726	0.0238393
	R	sand poor	Sph	55859.4	0.905925	1.63804
		sand rich	Sph	4430.62	0.313707	1.05615
		sand calcareous	Sph	2940.37	0.0723223	0.615876
		clay non-calcareous	Gau	38113.7	0.370697	0.483369
		clay calcareous	Sph	14413.1	0.0187672	0.134566
		loam non-calcareous	Sph	16448.4	3.01046	0.256574
		peat	Gau	13736	1.06678	0.651919
		aquatic	Gau	13157.9	0.581084	0.227116
	N	sand poor	Sph	67102.8	0.95717	1.62385
		sand rich	Sph	3890.06	0.259043	1.09133
		sand calcareous	Sph	14631.8	0.695128	1.33546
		clay non-calcareous	Gau	23718.4	0.455484	0.703885
		clay calcareous	Sph	5492.49	0.13957	0.486199
		loam non-calcareous	Sph	13466.7	1.97752	0.58679
		peat	Gau	16854.4	1.3182	1.05419
		aquatic	Sph	95110.7	1.12999	0.363035
	S	sand poor	Gau	133687	0.132272	0.0532109
		sand rich	Gau	20973.3	0.00666001	0.0419665
		sand calcareous	Sph	17274.6	0.0283415	0.055254
		clay non-calcareous	Gau	127128	0.276197	0.0417673
		clay calcareous	Sph	10792	0.118578	0.0214697
		loam non-calcareous	Sph	18146.8	0.0169199	0.00958046
		peat	Sph	80209.7	0.0320205	0.0389723
		aquatic	Lin	21890.6	0.0316262	0.433386
historical data set	F	sand poor	Sph	4691.45	1.01803	1.38764
		sand rich	Sph	9047.91	0.653341	1.06986
		sand calcareous	Exp	70575.7	2.46913	0.95082
		clay non-calcareous	Sph	4737.56	0.740844	0.652922
		clay calcareous	Sph	3511.83	1.2204	0.208011
		loam non-calcareous	Sph	17855.7	0.520102	0.712918
		peat	Sph	10419.2	1.23579	0.281114
		aquatic	Exp	23989.9	0.458421	0.421363
	R	sand poor	Sph	19288.5	1.74001	0.971388
		sand rich	Exp	4319.84	0.927468	1.16226
		sand calcareous	Sph	12619.2	0.550854	0.299173
		clay non-calcareous	Sph	3269.92	0.43142	0.315028
		clay calcareous	Sph	17239.2	0.166638	0.162992
		loam non-calcareous	Exp	588.3	1.13103	0
		peat	Sph	34943.7	0.315426	0.684402
		aquatic	Gau	46119.5	1.78523	0.305703
	N	sand poor	Exp	5282.88	2.02795	0.558702
		sand rich	Exp	2991.94	1.10212	1.13201
		sand calcareous	Sph	12697.8	0.805057	0.680817
		clay non-calcareous	Sph	5324.54	0.780387	0.47815
		clay calcareous	Sph	8357.11	0.659895	0.174614
		loam non-calcareous	Sph	6551.53	0.484775	1.0219
		peat	Exp	10417	0.41447	0.912549
		aquatic	Lin	207073	8.37842	0.448284
	S	sand poor	Sph	2429.47	0.447594	0
		sand rich	Gau	577063	9.6063	0.0631373
		sand calcareous	Gau	7900.21	1.07689	0.450988
		clay non-calcareous	Lin	slope =	1.74018e-5	0.242575
		clay calcareous	Exp	16625.3	3.77653	0.121759
		loam non-calcareous	Sph	27903.9	0.0179868	0.0101108
		peat	Sph	2823	0.032	0.06
		aquatic	Lin	149190	0.325372	0.0831723

Appendix III.2 — Interpolation results

Table A.6. Summary of interpolation for each time period, Ellenberg indicator variable and stratum

Time period	Ellenberg	Stratum	Minimum	Maximum	Mean	SD
actual data set	F	sand poor	3.9	8.4	6.3	0.7
		sand rich	4.3	8.3	6.6	0.6
		sand calcareous	3.8	7.7	5.5	0.6
		clay non-calcareous	4.8	8.4	6.7	0.6
		clay calcareous	4.9	8.2	6.1	0.6
		loam non-calcareous	5.0	6.9	5.6	0.4
		peat	4.2	8.8	7.5	0.7
		aquatic	8.0	12.0	10.3	0.7
	R	sand poor	2.3	6.9	4.0	0.8
		sand rich	2.5	6.9	5.0	0.7
		sand calcareous	2.0	6.9	5.9	0.5
		clay non-calcareous	3.9	7.0	6.1	0.4
		clay calcareous	6.0	7.2	6.7	0.2
		loam non-calcareous	1.9	7.2	5.9	0.9
		peat	1.9	7.1	5.3	1.0
		aquatic	2.8	8.8	6.9	0.8
	N	sand poor	2.0	6.4	4.1	0.8
		sand rich	2.7	6.8	5.0	0.6
		sand calcareous	2.5	7.0	5.4	0.8
		clay non-calcareous	3.9	7.3	5.9	0.5
		clay calcareous	5.0	7.4	6.2	0.5
		loam non-calcareous	2.4	7.0	5.7	0.8
		peat	1.5	7.1	4.9	0.9
		aquatic	2.2	7.5	6.1	0.6
	S	sand poor	0.0	1.3	0.3	0.1
		sand rich	0.0	0.9	0.2	0.1
		sand calcareous	0.0	3.7	0.4	0.4
		clay non-calcareous	0.0	1.5	0.5	0.2
		clay calcareous	0.0	6.9	0.4	0.4
		loam non-calcareous	0.0	0.4	0.1	0.1
		peat	0.0	0.8	0.3	0.1
		aquatic	0.1	2.6	0.7	0.5
historical data set	F	sand poor	3.9	8.5	6.3	0.6
		sand rich	4.3	8.1	6.4	0.5
		sand calcareous	3.7	7.6	5.6	0.9
		clay non-calcareous	4.1	8.7	6.3	0.6
		clay calcareous	3.3	9.3	6.3	0.6
		loam non-calcareous	4.5	7.0	5.6	0.5
		peat	3.7	9.4	6.9	0.7
		aquatic	8.7	11.4	10.5	0.4
	R	sand poor	1.5	7.2	4.4	0.9
		sand rich	1.9	6.9	4.8	0.7
		sand calcareous	2.8	7.6	5.9	0.6
		clay non-calcareous	2.9	7.3	5.8	0.4
		clay calcareous	4.3	7.6	6.4	0.6
		loam non-calcareous	2.4	7.1	5.9	0.6
		peat	2.5	6.2	5.1	0.6
		aquatic	3.6	8.0	6.8	0.6
	N	sand poor	1.5	7.1	4.0	0.9
		sand rich	1.9	6.4	4.4	0.6
		sand calcareous	2.9	6.9	4.7	0.7
		clay non-calcareous	2.6	7.4	5.4	0.4
		clay calcareous	2.6	7.7	5.7	0.5
		loam non-calcareous	2.3	6.0	5.0	0.6
		peat	2.3	5.9	4.7	0.6
		aquatic	2.0	7.4	6.2	0.7
	S	sand poor	-0.0	6.4	0.3	0.3
		sand rich	0.1	2.6	0.4	0.4
		sand calcareous	0.0	7.3	1.2	1.4
		clay non-calcareous	0.0	6.8	0.7	0.9
		clay calcareous	0.0	7.9	1.4	1.6
		loam non-calcareous	0.1	0.5	0.1	0.1
		peat	0.1	3.3	0.4	0.1
		aquatic	0.1	2.9	0.6	0.5

Appendix III.3 — List of digital records

Table A.7. *List of digital records of ARCINFO grids and GEOVIEW krt-files*

Time period	Ellenberg indicator value	GRID interpolation value	GRID interpolation variance	krt-file interpolation value	krt-file interpolation variance
actual	F	f_all	fse_all	f.krt	f_var.krt
	R	r_all	rse_all	r.krt	r_var.krt
	N	n_all	nse_all	n.krt	n_var.krt
	S	s_all	sse_all	s.krt	s_var.krt
historical	F	f_1000	fse_1000	f1000.krt	f1000_var.krt
	R	r_1000	rse_1000	r1000.krt	r1000_var.krt
	N	n_1000	nse_1000	n1000.krt	n1000_var.krt
	S	s_1000	sse_1000	s1000.krt	s1000_var.krt

Appendix IV.1 — Location of sites mentioned in the discussion

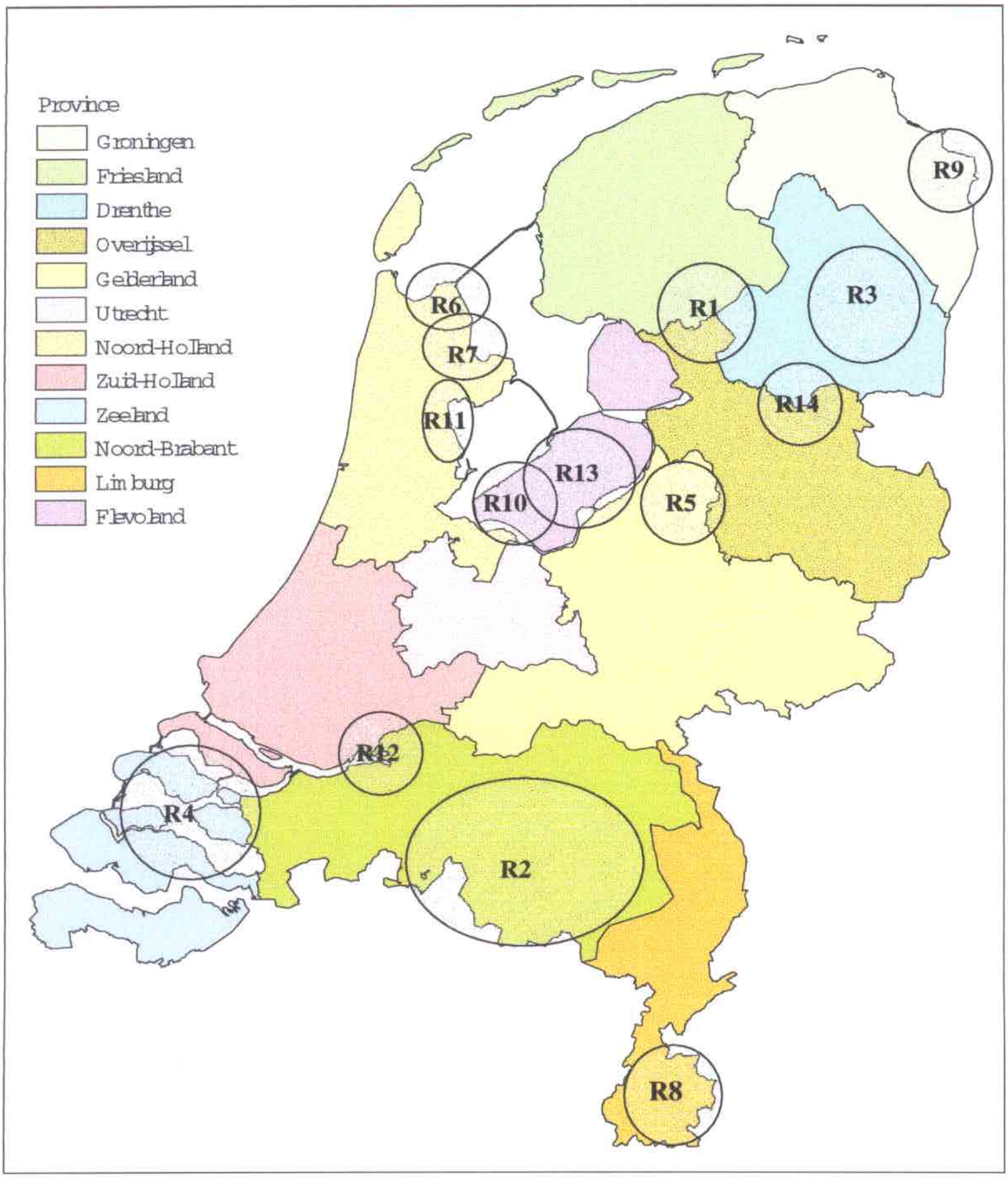

Figure A.12. *Map of the Dutch provinces showing problem areas.*

Appendix V Mailing list

1	Directeur-Generaal van het RIVM drs.ir. R. van Noort
2	Prof. dr. F. Berendse (vakgroep TON, LUW)
3	dr. C.J.F. ter Braak (CBW, DLO)
4	drs. J. Clausman (provincie Zuid-Holland)
5	dr. H.F. van Dobben (IBN, DLO)
6	drs. R. van Ek (DBW, RIZA)
7	dr. N.J.M. Gremmen
8	drs. W.B. Harms (SC, DLO)
9	drs. R.H. Kemmers (SC, DLO)
10	dr. J. Klijn (SC, DLO)
11	drs. J. Kros (SC, DLO)
12	drs. R. Meijers (IKC-N)
13	dr. P. Burrough (Fysische Geografie, RUU)
14	dr. E. Pebesma (Fysische Geografie, RUU)
15	drs. J. Runhaar (CML)
16	dr. J. Schaminée
17	dr. O. van Tongeren
18	ir. W. Wamelink (IBN, DLO)
19	dr. M. Wassen (Milieukunde, RUU)
20	ir. J.P.M. Witte (vakgroep Waterhuishouding, LUW)
21	dr. F.J. Zadelhoff (IKC, NBLF)
22	dr.ir. A Bregt (SC-DLO)
23	ir. F. Langeweg (Omgevingswetenschappen, LUW)
24	Depôt van Nederlandse Publicaties en Nederlandse Bibliografie
25	prof.ir. N.D. van Egmond
26	drs. ing. M. Bakkenes
27	ir. R. van den Berg
28	ing. G.P. Beugelink
29	ir. A.H.M. Bresser
30	drs. B.J.E. ten Brink
31	dr. A.L.M. Dekkers
32	dr.ir. J.J.M. van Grinsven
33	drs. A. van der Giessen
34	ir. M. de Heer
35	dr. A. van Hinsberg
36	drs. W. Lammers
37	dr. R. Leemans
38	drs. W. Ligtvoet
39	drs. J.G. Nienhuis
40	drs. R. Reiling
41	dr. H.J.P.A. Verkaar
42	ir. M. Vonk
43	drs. F.G. Wortelboer
44-47	auteurs
48	Hoofd Bureau Voorlichting en Public Relations
49	Bureau Rapportenregistratie
50	Bibliotheek RIVM
51	Bibliotheek LWD/ECO
52	Bibliotheek LBG
53	Bibliotheek IBN-DLO
54-68	Bureau Rapportenbeheer
69-80	Reserve