ELSEVIER

Journal of Hydrology 272 (2003) 250–263

Journal of Hydrology

www.elsevier.com/locate/jhydrol

# Predicting yield of barley across a landscape: a state-space modeling approach

Ole Wendroth[a,*], Hannes I. Reuter[a], K. Christian Kersebaum[b]

[a]*Department for Soil Landscape Research, Centre for Agricultural Landscape and Land Use Research, ZALF, Eberswalder Str. 84, 15374 Müncheberg, Germany*

[b]*Department of Landscape Systems Analysis, Centre for Agricultural Landscape and Land Use Research, ZALF, Eberswalder Str. 84, 15374 Müncheberg, Germany*

**Abstract**

Spatial crop yield prediction is an enigma that needs to be solved to avoid ecological and economical risks in agricultural crop production, that can result from local fertilizer surplus or deficiency. Current approaches for site-specific fertilizer distribution are based on patterns of soil properties and yield maps obtained from previous years. The aim of this study was to evaluate the quality of crop yield prediction in an arable field using two sets of variables in autoregressive (AR) state-space models. One set included detailed soil information (texture, organic carbon content) and yield data from the previous year at a high spatial resolution. In the other set, remotely sensed soil and crop information (vegetation index, crop nitrogen status, land surface elevation) was assembled, which is available under farm conditions without intensive soil sampling campaigns. Soil and remotely sensed variables were evaluated in bi- and multivariate autoregressive state-space analysis to predict spring barley grain yield. Remotely sensed variables showed to be better predictors for spatial grain yield estimation than soil variables. Transition coefficients determined from state-space analysis were applied in AR-equations with soil and remotely sensed information, but yet given only the initial value of the spatial yield series. Both sets of variables elicited similar prediction quality.



*Keywords:* Precision farming; Field-scale variability; State-space models; Land surface elevation; Autoregressive crop yield prediction

## 1. Introduction

Modern agricultural production is characterized by highly intensive and efficient production systems, and the average field size has been increased tremendously over past decades. Nowadays, large field units are managed homogeneously, although there exists a considerable inherent soil spatial variability causing spatially differing zones of fertility and physical properties. When such large field sites are fertilized homogeneously for example with nitrogen (N), economical and ecological disadvantages can be the consequence. The first is the case when the applied fertilizer dose is below the local optimum, the second when the applied fertilizer cannot entirely be used by the crop and may cause leaching of fertilizer nutrients. Farmers in the last decades have already intuitively met decisions with respect to the spatial variability pattern within their fields or have been varying fertilizer application rates locally, based on their experience and their expectations. This idea has been embedded in a technology, i.e. precision or site-specific farming, which are now possible due to

* Corresponding author. Fax: +49-33432-82280.
*E-mail address:* owendroth@zalf.de (O. Wendroth).

0022-1694/03/$ - see front matter 
PII: S0022-1694(02)00269-X

the availability of global positioning systems among others. Local records of crop yields and soil properties can now be taken into account in order to derive maps underlying locally varying fertilizer application rates (Wenkel et al., 2001). However, such a fertilizer distribution map requires a spatial expectation or prediction of crop yield, how its spatial variation is related to underlying soil properties, and how it can be spatially predicted even under spatially homogeneous management.

A lot of emphasis has been spent in using soil information and already existing yield maps from previous years as a basis for estimating crop yield distribution maps (Kersebaum et al., 2002). However, usually under farm conditions, the existing soil information is rather insufficient to draw a detailed map. Not to mention that such a map would neither allow to derive a spatial expectation of crop yield nor could a local fertilizer application rate be based on it.

Yield maps exhibit differing variability patterns for the same site in subsequent years (Stafford, 1999; Grenzdörfer and Gebbers, 2001). In addition, resources for intensive monitoring campaigns are limited and impractical under farm conditions. Moreover, the fact that weather conditions differ from year to year in the climatic zone of Western Europe with its strong maritime influence jeopardizes the use of soil data bases in their usual spatial resolution (50 by 50 m in Germany) for deriving a yield expectation map for a particular year. Low correlation coefficients at the field scale cause a considerable uncertainty in approaches in which crop yield estimations are based solely on soil properties such as textural grain size distribution and soil organic carbon content.

In addition to the available soil information, site information is nowadays used increasingly. Land surface images from either the bare soil surface or the crop can be obtained during different times of the growing season. If the pattern of such images is spatially related to a relevant soil state-variable, e.g. soil water content, or indicates the crop development distribution within a field site, this could be a useful alternative for predicting the spatial crop yield distribution. Moreover, even for conditions when the existing soil map does not exhibit a pronounced soil variation, considerable spatial yield variation can be observed. Shaw and Carter (2002) pointed out that although soil survey is efficient for grouping soil variability at the landscape scale, substantial spatial variability arises from near surface soil properties within the same mapping unit. Additional effects on the variation of grain yield (Ciha, 1982) may be related to landscape morphology (Pennock and Corre, 2001), i.e. surface relief and underlying soil development, including erosion.

Regardless of the information underlying a predicted map, another important issue has to be considered. The quality and evidence of a yield map obtained from a monitoring system has to be examined critically for at least two reasons. One is the uncertainty of such a map due to the measurement technique. The other is the fact that even one single intensive rain event before harvest can affect the crop yield, e.g. by causing lodging. Under such conditions, the yield obtained at a respective location does not necessarily reflect the previous vegetative development and biomass production relative to the local neighborhood.

During the last two decades, statistical tools based on autoregression that are appropriate for describing the process of a set of variables were adapted from applied statistical time series analysis to spatial series of soil and crop data. Using autoregressive state-space models, Morkoc et al. (1985) and Shumway et al. (1989) described the spatial association of soil water content and soil temperature. Nielsen and Alemi (1989) showed the coincidence between cotton yield and nematode infestation. Wendroth et al. (1992) identified processes underlying yield of a N-fixing legume and a non-fixing reference crop. Li et al. (2001) based cotton lint yield variability upon soil water content, nitrate-N and elevation. Due to the fact that state-space techniques concede measurement and model uncertainty that drive a stochastic updating and filtering step, and due to the autoregressive model equation, spatial relations between variables can be identified with greater accuracy than with ordinary regression techniques (Nielsen et al., 1999). Shumway (1985) addressed advantages of geostatistical interpolation techniques, e.g. kriging and cokriging, which allow spatial estimations based upon sparse collections of data points. On the other hand, time series techniques required equally spaced data. However, with time series techniques such as state-space analysis, problems can be approached which involve instationarity and missing data (Shumway, 1985).

The emphasis for applying state-space analysis here as in many current spatial and temporal data series is to identify underlying processes of one-dimensional series of observations (Nielsen et al., 1994, 1999) and to estimate model parameters based on maximum likelihood (Shumway, 1985).

The objective of this study was, to evaluate two sets (vectors) of different variables monitored at a site for predicting yield distribution. One group of variables consisted of grain yield, soil textural data, soil organic carbon content at 0–30 cm depth, and grain yield from the previous year and crop. The other group comprised grain yield, remotely sensed normalized differential vegetation index (NDVI98), crop N status (CNstat) at one time during the vegetative period, elevation, and slope. The first group of variables is rather soil-based and requires detailed information. However, except for the yield of the previous crop no annual or short-term influences upon crop development are integrated. The second set mainly consists of variables, which reflect current crop and soil status, hence integrate a lot of influences upon the recent crop development. Both sets should be examined with respect to their support of yield prediction. The validity and robustness of transition coefficients was examined in autoregressive predictions without any stochastic filtering and updating.

## 2. Material and methods

### 2.1. Site and experimental description

The field site of this investigation is located in Luettewitz on a farm of the Suedzucker company, southeastern Germany, in the federal state of Saxony (51°7′N, 13°14′W). The soil is derived from loess and is classified as a Stagnic Luvisol. The annual average air temperature is 8.0 °C, and the average annual precipitation is 662 mm. The size of the investigated field site is approximately 21 ha. Across this field, a regular grid of 15 × 15 points was laid out with 29 m distance between grid points in the north and east direction (Fig. 1). All measured and calculated information included in state-vectors for this study was aggregated to this 29 × 29 m grid. As state-space analysis is designed for observations taken in one dimension, observations were arrayed in the way indicated in Fig. 1, beginning in the upper left hand corner of the field and leading to the lower right hand corner of the field.

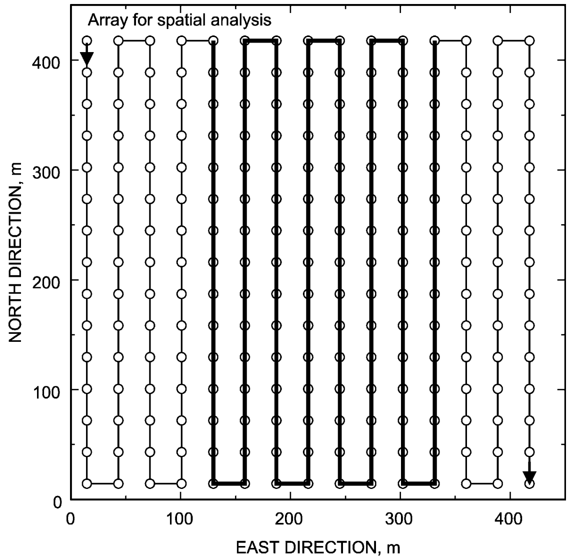


Fig. 1. Sampling grid and array of data in the North–South direction for spatial analysis. The bold part of the line refers to those measurement locations for which results are depicted in Figs. 2 and 3.

In spring 1998, spring barley (*Hordeum vulgare*, L.) was planted in the field. In the previous year 1997, the site had been grown to triticale (*Triticosecale wittmack*). Grain yield in both years 1997 (Yi97) and 1998 (Yi98) was determined during harvest with an automatic CLAAS yield monitoring system ‘CEBIS’.

On 30 May, 1998, an aerial infrared photograph was taken from an aircraft (Jürschik and Schmerler, 1997). This photograph was processed to obtain a map of NDVI (Baret, 1995). Soil textural grain size distribution was determined for three depths (0–30 cm, 30–60 cm, and 60–90 cm), and soil organic carbon content (SOC) at 0–30 cm depth. Soil texture and SOC were determined at every other location in both the north and south direction, hence at 64 points. In comparison to the other variables, this spatial resolution of textural data is coarse but this was the only affordable amount of measurements. However, for practical farm conditions, such a resolution would probably be unaffordable. For this study, soil texture was linearly interpolated for the unknown locations. From the variety of grain size fractions, silt (2–63 μm equivalent particle diameter) at 30–60 cm depth and sand content (63–2000 μm) at 60–90 cm depth were considered in the analysis. These grain size fractions were selected from the existing set of soil information, as they yielded largest correlation coefficients with spring barley grain yield among other fractions and other soil depths, respectively (Table 1). Moreover, ranges of spatial dependence determined from cross variograms between barley yield and respective variables indicated strong spatial associations between 61 and 198 m, approximately corresponding to 2 and 7 lags, respectively (Table 1).

Currently, crop sensors are becoming available that allow monitoring, e.g. the crop growth or nutrition status. At present such a sensor is being developed to monitor crop N status. However, this sensor was not available at the time of our experiments. Therefore, a corresponding crop N state-variable was generated from the crop growth and N dynamics model HERMES (Kersebaum, 1995). This model was not calibrated for the field site, and was applied at each of the 225 grid points, given the respective soil textural information and the soil mineral N content after harvest of the previous crop as the initial condition (Kersebaum et al., 2002). Simulated N in the above ground biomass was taken for 14 June, 1998 and included in the state-vector.

Table 1
Correlation coefficients and associated ranges of spatial dependence between spring barley grain yield (Yi98) and other variables

| | Correlation coefficient $r$ Yi98 | Range of spatial dependence, $m$ |
|---|---|---|
| NDVI98 | 0.541 (2) | 137.5 |
| Yi97 | 0.339 (3) | 99.8 |
| Crop-N-status on June 14, 98 (simulated) | 0.371 (4) | 61.2 |
| Elevation | 0.588 (1) | 197.6 |
| Slope | −0.098 | 151.4 |
| Silt 30–60 cm | 0.307 | 118.4 |
| Sand 60–90 cm | −0.280 | 143.6 |
| SOC | 0.269 | 144.3 |

Ranges of spatial dependence were derived from spherical models fitted to cross-variograms. Rank orders of $r$ are given in parenthesis( ).

For the field site, a digital elevation model was determined by laser altimetry. Slope was computed using the D8-method (steepest descent) on a $10 \times 10$ m basis. Both elevation and slope were averaged for the $29 \times 29$ m grid. For further details see Reuter et al. (2001).

### 2.2. Theory

State-space models in general consist of a model equation and an observation equation. For this study, an autoregressive equation was combined with the Kalman Filter (Kalman, 1960). In this case, the state-equation is

$$Z_i = \mathbf{\Phi} Z_{i-1} + \omega_i \tag{1}$$

which describes how the state-vector $Z_i$ at location $i$ which includes a set of $p$ variables is related to that at location $i-1$. This relation is manifested in the $p \times p$ transition matrix $\mathbf{\Phi}$ that consists of autoregression coefficients. Only one-dimensional experimental designs support autoregressive analysis (Shumway, 1985). Accordingly, the experimental data sampled across the field were arrayed in a spatial series that allows the application of time series analytical tools. The spatial direction of data series is certainly changed. However, spatial distances remain as if observations were taken along one line with regular sampling intervals. The model error term $\omega_i$ is a zero mean uncorrelated noise with a $q \times q$ covariance matrix $\mathbf{Q}$ the latter being variance per unit space or time, and depending on the interval between observations. This source of uncertainty is due to a systematic error implied in the model itself and in the underlying functions, in this case the magnitude and the general validity of the transition coefficients. This state-equation is solved simultaneously with the observation equation

$$Y_i = M_i Z_i + \upsilon_i \tag{2}$$

in which the true state $Z_i$ is reflected in the observation $Y_i$ by a measurement matrix $M_i$ and an uncorrelated zero mean measurement error with $\upsilon_i$ common covariance $R$. Hence, the observation does

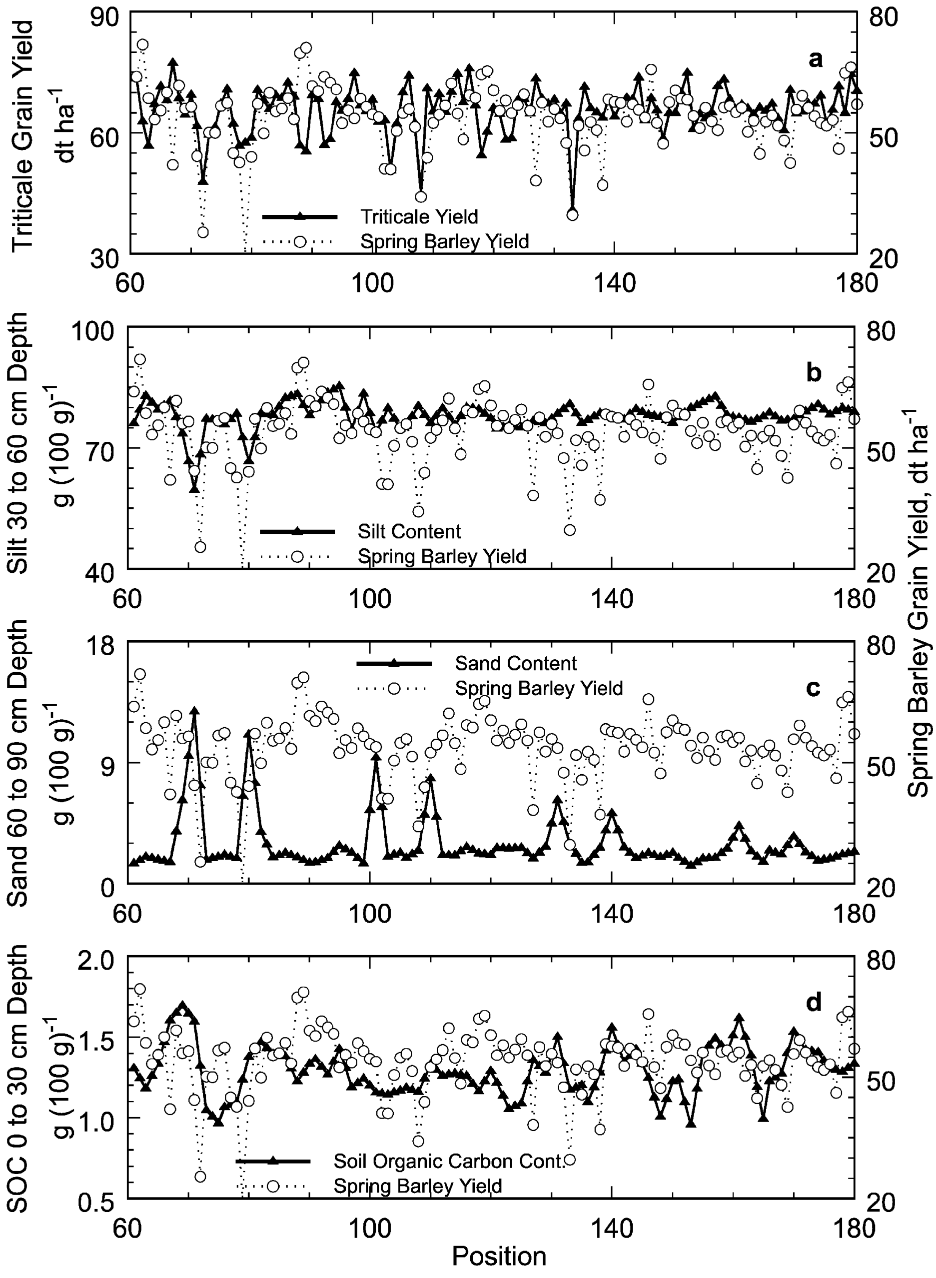


Fig. 2. Spatial process of spring barley grain yield, triticale grain yield (a), silt content in 30–60 cm depth (b), sand content in 60–90 cm depth (c), and soil organic carbon content (SOC) in 0–30 cm depth (d). For a better visualization of the spatial process of the series, data are presented for locations 61–180 (Fig. 1).

not have to be taken to be fully true but is only an indirect measure of the true system's state.

In this study, the system including the filtering, updating and smoothing steps was solved with the expectation maximization algorithm described by Shumway and Stoffer (1982). For further details, see Shumway (1988), Shumway and Stoffer (1982, 2000), Nielsen et al. (1994) and Nielsen and Wendroth

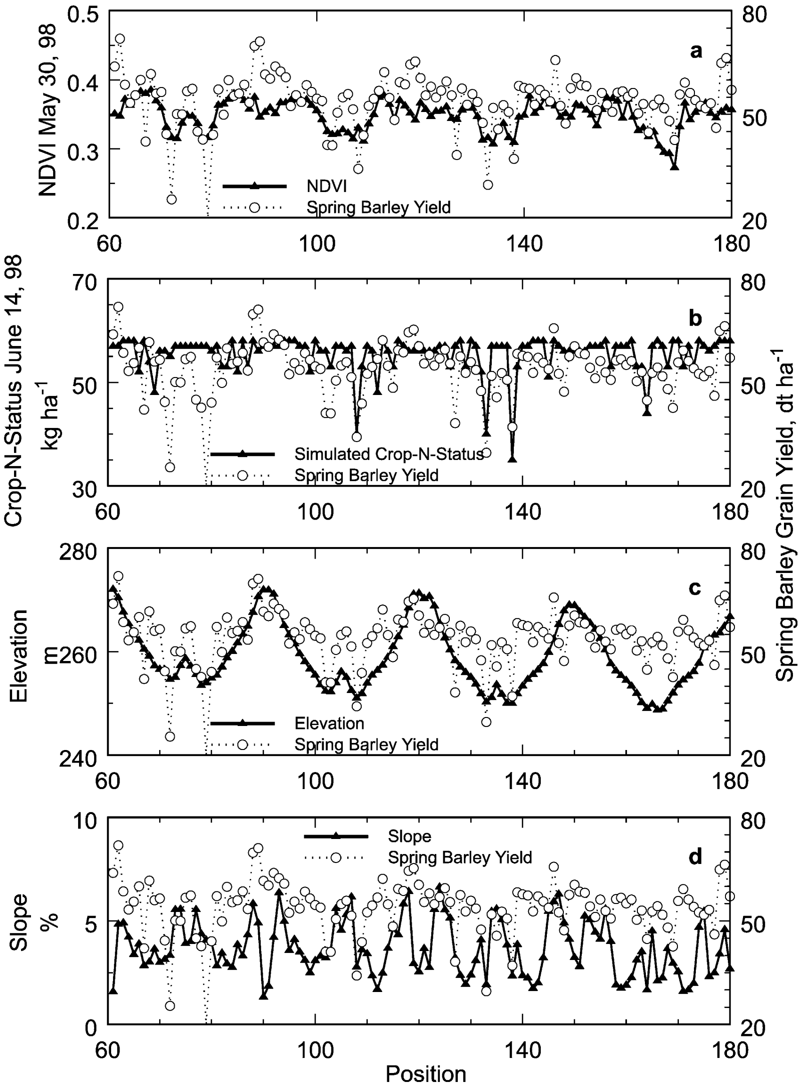


Fig. 3. Spatial process of spring barley grain yield, normalized differential vegetation index (NDVI) (a), crop N status on June 14, 1998 (b), elevation (c), and slope across the field (d). For a better visualization of the spatial process of the series, data are presented for locations 61–180 (Fig. 1).

(2002). One of the advantages of deriving autoregression coefficients in state-space models compared to ordinary autoregression models is the implication of the two error terms, the model and measurement noise and the mathematical separation of noise and signal. The autoregression coefficients are based on an update of the estimated state. However, for robustness of the solution the log likelihood is not affected by any updated or smoothed state but on the predicted state (Shumway, 1988 and Nielsen and Wendroth, 2002). Before applying state-space analysis, data were scaled by the following equation (Wendroth et al., 2001)

$$ysc_i = \frac{y_i - (\mu_y - 2\sigma_y)}{4\sigma_y}. \quad (3)$$

In this equation the scaled value $ysc_i$ of the original observation $y_i$ is calculated based on the mean $\mu_y$ and the standard deviation $\sigma_y$. This scaling procedure was applied for two reasons. One reason is to avoid numerical problems that can arise if variables differ by orders in their magnitude. Instead, if they are in the same order of magnitude, their transition coefficients reflect their relative contribution to the estimate. Data series have a mean of 0.5 and a standard deviation of 0.25 when normalized with Eq. (3). The second reason is the comparability of the log likelihood for combinations of variables. By this normalization procedure, the variables have the same probability density function, and only differ by their correlation structure (Nielsen and Wendroth, 2002). The log likelihood allows to compare how different state-vectors with the same number of variables and the same number of observations contribute to the estimation. The lower the log likelihood (value of $-2\ln L$), the better the prediction.

In order to evaluate the resulting transition coefficients from state-space analysis, these coefficients were applied in simple autoregressive predictions, where only the first yield value in the series is known, and all following values are calculated from the previous one and those from the underlying variables included in the respective state-vector. As a criterion for prediction quality, the average of squared deviations between measured spring barley yield $\mathrm{Yi98}^{\mathrm{meas}}$ and predicted spring barley yield $\mathrm{Yi98}^{\mathrm{pred}}$ was calculated with

$$\mathrm{SQD}_{\mathrm{avg}} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Yi98}^{\mathrm{meas}} - \mathrm{Yi98}^{\mathrm{pred}})^2. \tag{4}$$

## 3. Results and discussion

Spring barley grain yield across the 225 locations arrayed in one dimension is shown in comparison to triticale grain yield (Fig. 2a), silt content at 30–60 cm depth (Fig. 2b), sand content in the layer from 60 to 90 cm (Fig. 2c), and soil organic carbon content (SOC) at 0–30 cm depth (Fig. 2d). For better illustration of the spatial series, not all 225 data points but only those at positions 61 through 180 (Fig. 1) are shown. The three latter variables are correlated with $|r| < 0.31$ (Table 1), and are less associated to spring barley yield than grain yield of triticale, i.e. the yield of the previous crop. The spatial processes of the second group of variables is shown versus spring barley grain yield with NDVI (Fig. 3a), crop N status at a day in mid June (Fig. 3b), elevation (Fig. 3c), and slope (Fig. 3d). Except for Yi97, all variables in this group exhibit a closer relation with spring barley grain yield (Table 1). Correlation between yield and elevation was slightly lower than that observed by Li et al. (2001). Spatial ranges obtained from cross variograms between yield and elevation shown in Table 1 were longer than those reported by Li et al. (2001) owing to the fact that topographic differences in their field site were less pronounced and reached across shorter distances than those in our study. In their study on grain yield and underlying soil properties, Cassel et al. (2000) identified cross correlation lengths being shorter than the spatial ranges of dependence between yield and soil properties resulting from our study. Spatial ranges of variograms (results not shown) and cross variograms observed in our study, were similar to those observed by Sadler et al. (1988) for yield data. Based on their results, Sadler et al. (1988) recommended resolutions finer than 100 m for precision farming studies in Coastal Plain soils. This suggestion is confirmed by our results. The four highest correlation coefficients were obtained in the rank sequence of Elevation > NDVI > Yi97 > crop-N-status.

An elevation map of our field site is shown in Fig. 4 to convey an impression of topographical

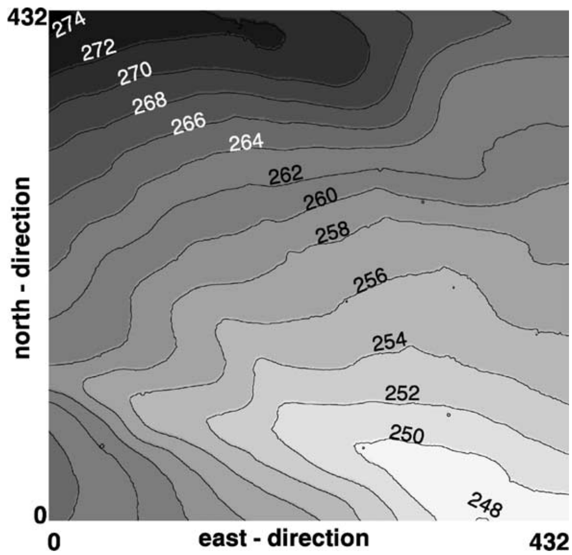


Fig. 4. Elevation map of the field investigated in this study (all units in meter).The field is located in Luettewitz, southeastern Germany.

differences. The large northeastern region of the field is a south-directed hillslope. A valley exists in the southeast corner of the field. When considered in the one-dimensional array in Fig. 3c, an association between elevation and spring barley grain yield is obvious. Rather than between crop yield and elevation, Timlin et al. (1998) found strong spatial coherence between corn yield and curvature. According to Timlin et al. (1998) including curvature in the analysis may explain differences in crop yield between different years when local water transmission differs.

All variables monitored were systematically compared in bivariate state-space analysis in two different scenarios, respectively. In one scenario, observations at each location were considered in the estimation. In this case, updating is possible at each location. In the second scenario, only one out of four spring barley grain yield measurements were taken into account, i.e. the updating step is possible only at those locations where an observation of spring barley grain yield is available.

For both groups of variables, an example is shown for the respective scenarios. Spatial prediction of spring barley grain yield based on silt content at 30–60 cm depth (Silt30) is shown in Fig. 5 with the respective state-equation. The values for the log likelihood resulted −965 and −1268 for the scenario considering all and only every fourth yield observation, respectively (see also Table 2). In both scenarios, the coefficients of the transition matrix remain relatively stable. The 95% confidence interval of the estimation increases for the case where less yield observations become available for an updating (Fig. 5). The closer the 95 % confidence interval, the better are local fluctuations of grain yield met by the model prediction.

As an example for the second set of variables, elevation is used in state-space analysis for estimating the spatial process of spring barley grain yield. The results (Fig. 6 and Table 2) indicate different weights on the previous yield observation for both scenarios. This may be due to the fact that, unlike in the original yield series (Fig. 6a), in the series of one out of four

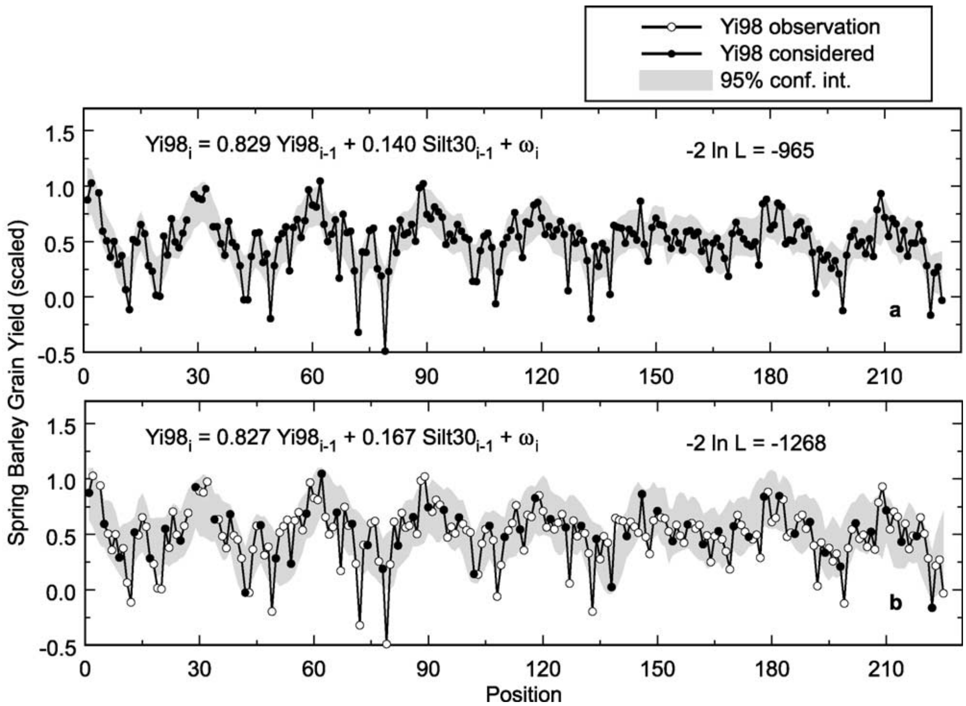


Fig. 5. Bivariate state-space analysis of spring barley grain yield (Yi98) and silt content at 30–60 cm soil depth (Silt30) considering all Yi98 observations (a) and only every fourth Yi98 observation (b), respectively. Variables were scaled using Eq. (4). Their respective state-equations and the log likelihood values ($-2 \ln L$) are given.

Table 2
Transition coefficients according to Eq. (1) for bivariate state-space analysis of spring barley grain yield [$\phi$ (Yi98)] and other covariables [$\phi$ (cvar)], as well as respective values for the log likelihood

| | State-space analysis (1st order bivariate) | | | | | |
|---|---|---|---|---|---|---|
| | All Yi98-observations considered | | | Every 4th Yi98-observation considered | | |
| | Transition coefficients | | Log likelihood | Transition coefficients | | Log likelihood |
| | $\phi$ (Yi98)$_{i-1}$ | $\phi$ (cvar)$_{i-1}$ | | $\phi$ (Yi98)$_{i-1}$ | $\phi$ (cvar)$_{i-1}$ | |
| NDVI98 | 0.818 | 0.141 | −1156 (2) | 0.901 | 0.076 | −1374 (2) |
| Yi97 | 0.684 | 0.289 | −876 | 0.517 | 0.504 | −1117 |
| Crop-N-status on June 14, 98 (simulated) | 0.385 | 0.600 | −851 | 0.589 | 0.430 | −1115 |
| Elevation | 0.638 | 0.322 | −1498 (1) | 0.811 | 0.127 | −1788 (1) |
| Slope | 0.854 | 0.133 | −887 | 0.829 | 0.165 | −1071 |
| Silt 30–60 cm | 0.829 | 0.140 | −965 | 0.827 | 0.167 | −1268 (4) |
| Sand 60–90 cm | 0.915 | 0.071 | −968 (4) | 0.920 | 0.072 | −1237 |
| SOC | 0.838 | 0.129 | −1062 (3) | 0.933 | 0.049 | −1355 (3) |

The numbers in parentheses represent the ranks of the respective four best estimation results.

considered Yi98$_i$ values (Fig. 6b) some extremely high and low yield values are not included. The magnitudes of the respective 95% confidence intervals are smaller than those for the combination of Yi98 and silt at 0–30 cm depth (Fig. 5). Values for the log likelihood (−1498 and −1788, respectively, Fig. 6, Table 2) indicate a stronger association of spring barley yield with elevation than with silt content.

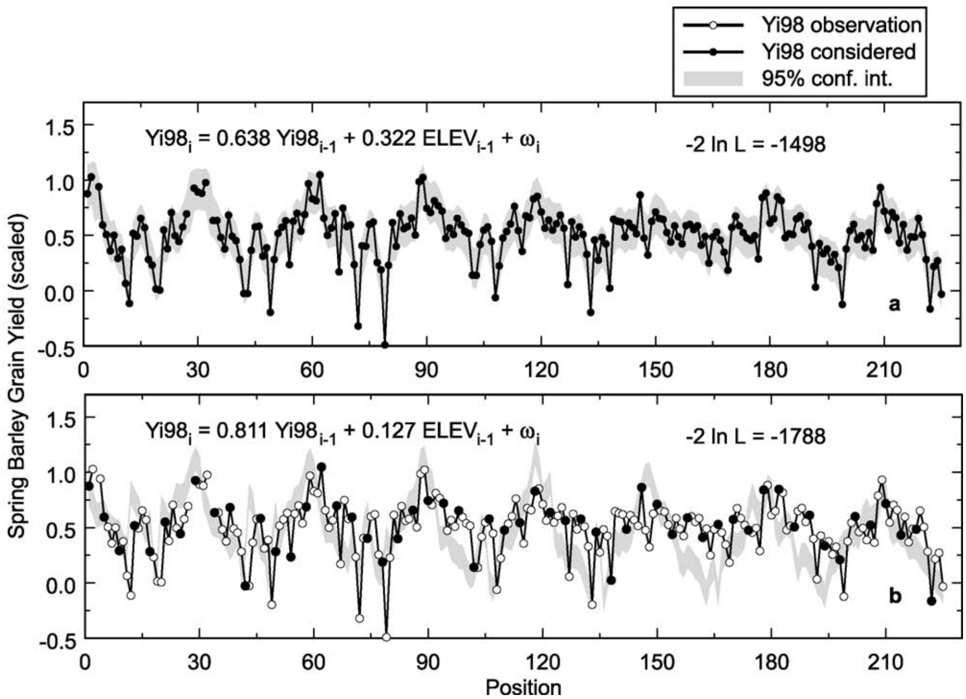


Fig. 6. Bivariate state-space analysis of spring barley grain yield (Yi98) and elevation considering all Yi98 observations (a) and only every fourth Yi98 observation (b), respectively. Variables were scaled using Eq. (4). Their respective state-equations and the log likelihood values ($-2 \ln L$) are given.

Unlike ordinary correlation analysis, the four strongest contributing variables that were identified in state-space analysis were in the rank Elevation NDVI98 > soil organic carbon (SOC) > sand, and silt content, respectively, for the different scenarios (Table 2).

From both groups of variables, state-vectors were composed including three variables besides spring barley grain yield itself, i.e. Silt30, SOC, and Yi97 in one set, and NDVI98, crop N status (CNstat) and elevation in the other set. The two different scenarios regarding yield resolution were the same as above. Results for the multivariate state-space analysis of the soil-based set of variables are shown in Fig. 7. The contributions of the three variables and the weight of the previous Yi98 observation change considerably compared to the bivariate state-space analysis (Fig. 5 and Table 2). Notice, that the contribution of Silt30 increases strongly from 0.19 to 0.68 when only one out of four Yi98 observation is taken into account in the estimation (Fig. 7). Again, the 95% confidence interval increases when only every fourth Yi98 observation is considered. Especially, at position 100–225, the spatial process of Yi98 is rather smooth in the second scenario (Fig. 7b), i.e. fluctuations cannot be conserved.

The second set of variables with rather sensor and short-term integrative variables besides Yi98 causes a better prediction of spring barley grain yield, as indicated by the log likelihood (−2580 versus −2058 in the first set, and −2814 versus −2346, in the scenarios, respectively). Unlike for the state-vector with soil-based variables, the fluctuations at positions 100–225 remain pronounced in the second scenario, and proceed less smoothed (Figs. 7b and 8b).

So far, the estimation of transition coefficients and the prediction of spring barley grain yield has been based in all scenarios and all different state-vectors on observations of the yield itself, though with varying observation densities.

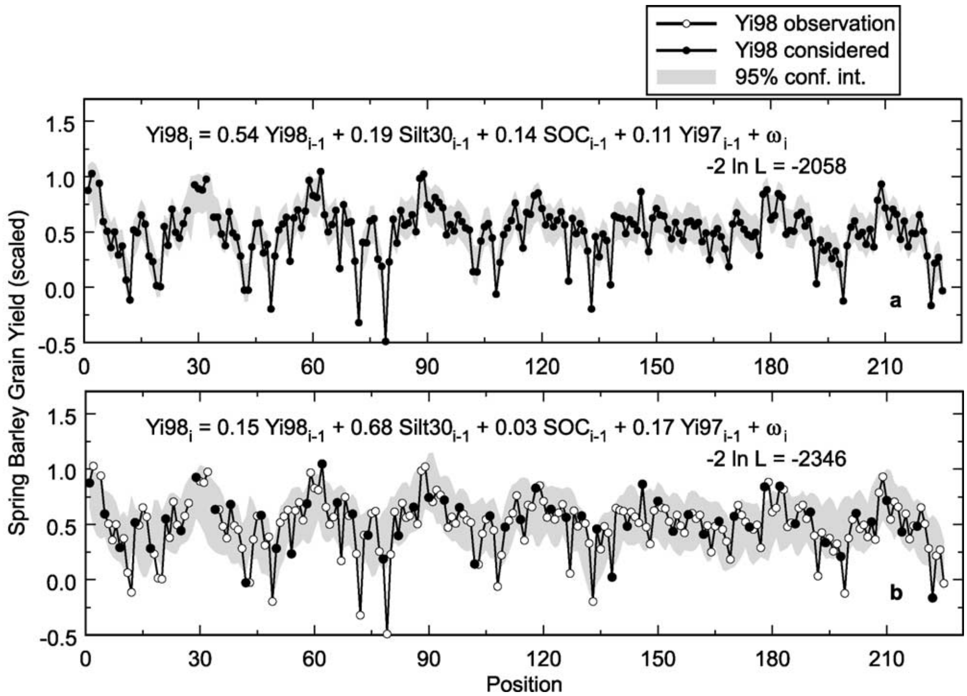


Fig. 7. Multivariate state-space analysis of spring barley grain yield (Yi98), silt content at 30–60 cm soil depth (Silt30), soil organic carbon content (SOC), and triticale grain yield from the previous year (Yi97) considering all Yi98 observations (a) and only every fourth Yi98 observation (b), respectively. Variables were scaled using Eq. (4). Their respective state-equations and the log likelihood values (−2 ln $L$) are given.

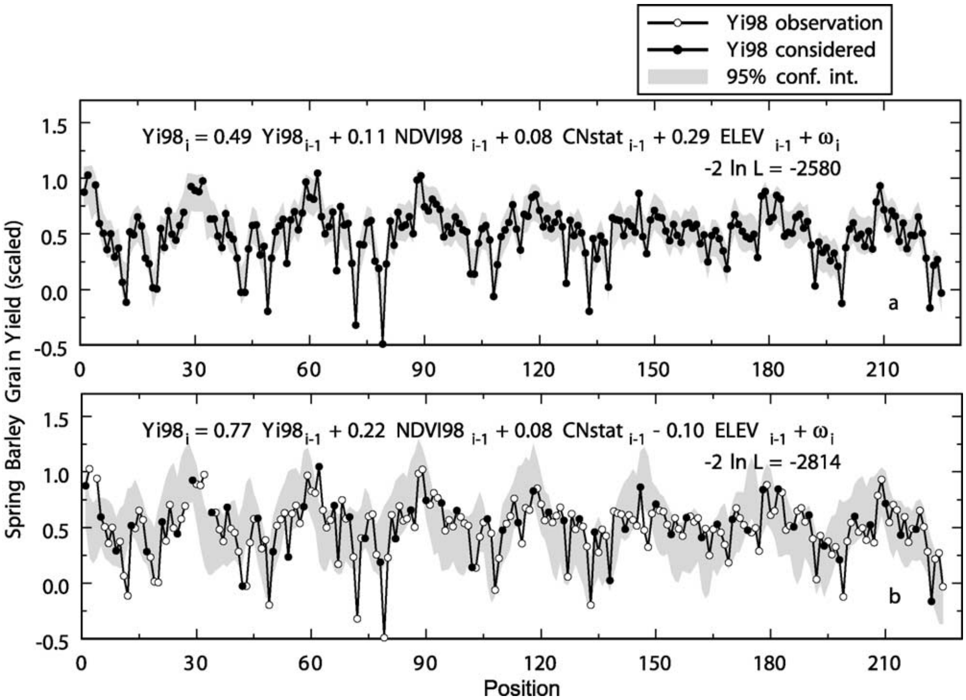


Fig. 8. Multivariate state-space analysis of spring barley grain yield (Yi98), normalized differential vegetation index (NDVI98), crop-N status calculated for June 14, 1998 (CNstat), and elevation considering all Yi98 observations (a) and only every fourth Yi98 observation (b), respectively. Variables were scaled using Eq. (4). Their respective state-equations and the log likelihood values ($-2 \ln L$) are given.

The question remains, how well the results obtained can contribute in a more predictive scenario, i.e. in a case where the main variable of interest, the spring barley grain yield is not known. Notice, that barley grain yield measurements had to be included in the state-space estimation procedure in order to determine underlying processes. However, under farm conditions, yield observations of the currently growing crop are not readily available. Therefore, transition coefficients from state-equations obtained above should be evaluated via application in ordinary autoregressive models, where only the initial yield value at position 1 is given. Hence, these equations could be used for the rather soil-based state vector after harvest of the previous crop, and at any time during the vegetation period when the information of the respective underlying state-variable becomes available.

Autoregressive spring barley yield predictions are given for the soil-based state-vector, based on autoregression or transition coefficients obtained in the respective scenarios above. The average squared deviation between prediction and observation is 0.0522 if transition coefficients were used from the scenario using all observations, and 0.0659 if coefficients are based on the scenario with only one out of four yield observations known, respectively (Fig. 9).

The second set of variables is applied in this prediction procedure, as well. The results exhibit similar prediction quality with $SQD_{avg}$ being 0.0526 for the set of transition coefficients obtained from the scenario with all observations (Fig. 10a). Prediction quality is similar to that obtained from the soil information ($SQD_{avg}$ being 0.0675 versus 0.0659) when transition coefficients are examined, that have been derived from the scenario with only one out of four yield observations.

The results in the last two figures are strongly depending on the initial value of Yi98 at location 1, here assumed to be known. However, at least

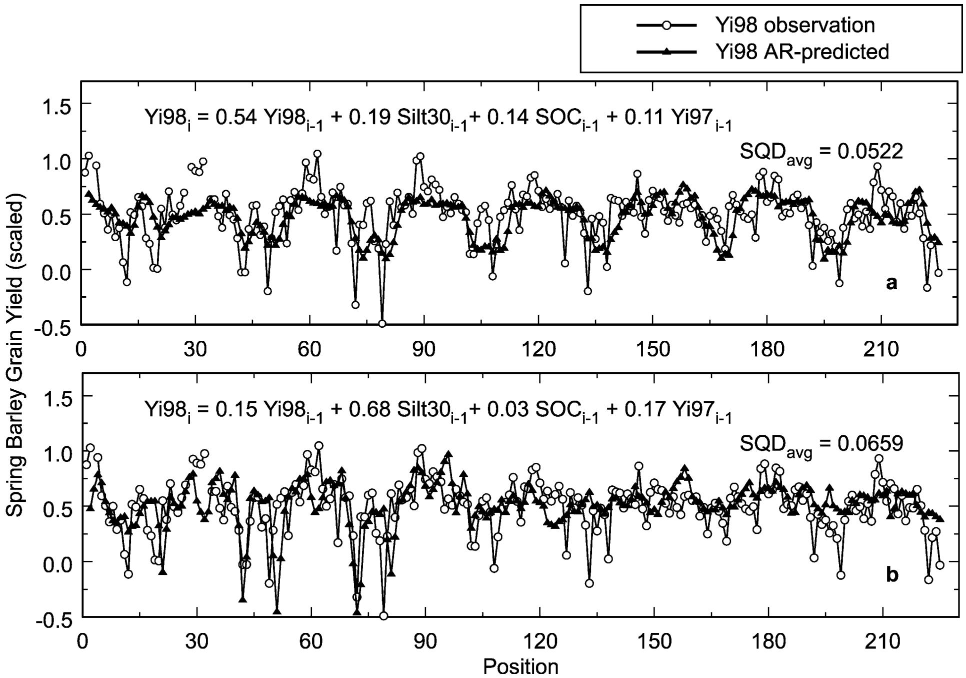


Fig. 9. Autoregressive prediction of spring barley grain yield (Yi98), based on Yi98, silt content at 30–60 cm soil depth (Silt30), soil organic carbon content (SOC), and triticale grain yield from the previous year (Yi97). Autoregression coefficients were obtained from state-space analysis (Fig. 7). Variables were scaled using Eq. (4). The average squared deviation $SQD_{avg}$ is given as a measure of prediction quality (see text).

the relative spatial process of yield is obtained from this autoregressive prediction. It remains open whether it is possible at all, to gain a precise prediction of yield magnitude, since a variety of short-term effects even a few days prior to harvest can change the level of magnitude of final yield. Here, it is assumed, that relative yield distribution across that field is the prediction goal. That aim has been reached especially with the second set of variables with encouraging accuracy. The method presented here is based on empirical analysis of spatial processes underlying crop yield variation. Therefore, variables were selected that integrate considerable deterministic information. Physiological interpretation cannot be expected from autoregressive equations. For the purpose of biogeochemical reactions, physical and bio-geochemical equations can be combined in state-space models (Nielsen et al., 1994). The analytical approach presented here was chosen as an alternative to deterministic description of field-scale crop yield variation.

## 4. Conclusion

Bivariate state-space analysis resulted in a different answer which variables are helpful estimating spring barley grain yield series, compared to ordinary correlation. Overall, soil-based variables proved to be less helpful in both bivariate and multivariate analyses than variables obtained from sensors and elevation. It remains open, how crop N status obtained from a sensor contributes to yield estimation and prediction. Results from this study obtained from model calculations of an uncalibrated model are encouraging. Future work has to show, how both rather soil-based and information obtained during the growing season can support yield prediction. Validity of coefficients has to be examined both, for neighboring fields and for other years. In this study, here, the field-scale was the focus of this investigation. The question remains open for future research how useful the variables considered here and others behave at different scales,

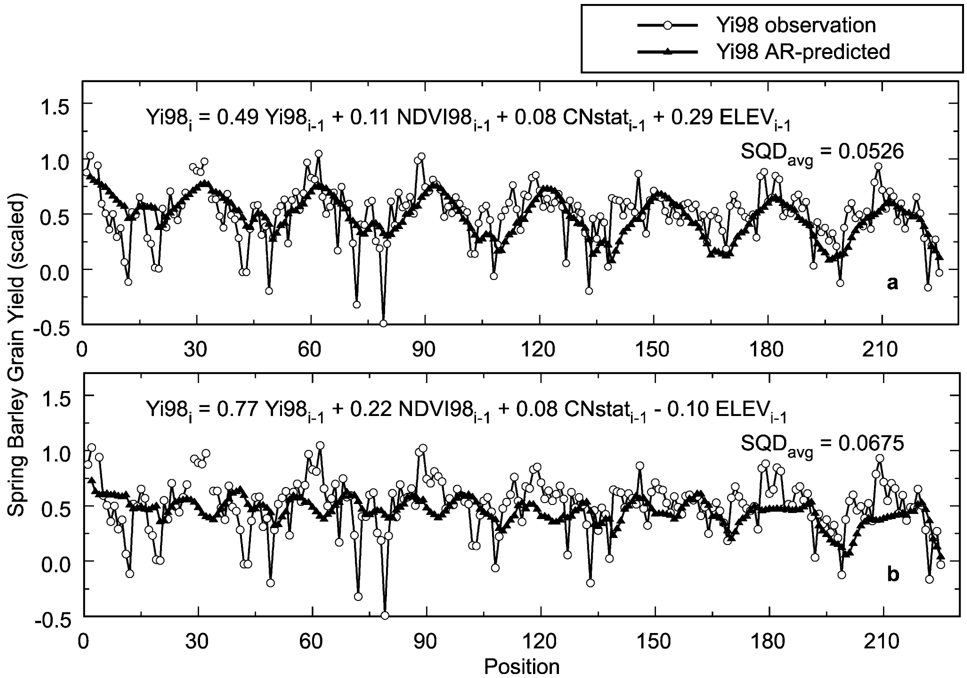


Fig. 10. Autoregressive prediction of spring barley grain yield (Yi98), based on Yi98, normalized differential vegetation index (NDVI98), crop-N status calculated for June 14, 1998 (CNstat), and average daily solar radiation (SRAD) in the month of May. Autoregression coefficients were obtained from state-space analysis (Fig. 8). Variables were scaled using Eq. (4). The average squared deviation $SQD_{avg}$ is given as a measure of prediction quality (see text).

e.g. for regional large scale predictions. A better support for on-farm prediction may result from remotely sensed information. However, for predictions at the scale where soil properties dominate the yield magnitude, the available soil information stays important. In this study, the soil within the investigated field was relatively homogeneous. Under conditions where soil variability is more pronounced, a stronger impact of soil properties on crop yield can be anticipated.

## Acknowledgements

The authors are indebted to the two reviewers for their helpful comments. The technical assistance of Norbert Wypler (ZALF), Antje Giebel and Michael Heisig (ATB Potsdam) is greatly appreciated. Special thanks to the farmer Eduard Steigerwald. Support of this study by German Research Foundation (DFG, Bonn), Suedzucker, Agrocom, Amazonen-Werke is acknowledged.

## References

Baret, F., 1995. Use of spectral reflectance variation to retrieve canopy biophysical characteristics. In: Danson, F.M., Plummer, S.E. (Eds.), Advances in Environmental Remote Sensing, Wiley, New York, pp. 33–51.

Cassel, D.K., Wendroth, O., Nielsen, D.R., 2000. Assessing spatial variability in an agricultural experiment station field: opportunities arising from spatial dependence. Agron. J. 92, 706–714.

Ciha, A.J., 1982. Slope position and grain yield of soft white winter wheat. Agron. J. 76, 193–196.

Grenzdörfer, G.J., Gebbers, R.I.B., 2001. Seven years of yield mapping—analysis and possibilities of multi year yield mapping data. In: Grenier, G., Blackmore, S. (Eds.), ECPA 2001. Proceedings of Third European Conference on Precision Agriculture, Montpellier, France, pp. 31–36.

Jürschik, P., Schmerler, J., 1997. Site specific plant production—acquisition of input data and first practical results of site specific fertilizing. In: Bork, H.R., (Ed.), Land Surface Processes, ZALF-Bericht No. 31, Müncheberg, Germany, pp. 47–57.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Engng 8, 35–45.

Kersebaum, K.C., 1995. Application of a simple management model to simulate water and nitrogen dynamics. Ecol. Modell. 81, 145–156.

Kersebaum, K.C., Reuter, H.I., Lorenz, K., Wendroth, O., 2002. Modeling crop growth and nitrogen dynamics for advisory purposes regarding spatial variability. In: Ahuja, L.R., Ma, L., Howell, T. (Eds.), Agricultural System Models in Field Research and Technology Transfer, Lewis Publishers, Boca Raton, FL, pp. 229–252.

Li, H., Lascano, R.J., Booker, J., Wilson, L.T., Bronson, K.F., 2001. Cotton lint yield variability in a heterogeneous soil at a landscape scale. Soil Till. Res. 58, 245–258.

Morkoc, F., Biggar, J.W., Nielsen, D.R., Rolston, D.E., 1985. Analysis of soil water content and temperature using state-space approach. Soil Sci. Soc. Am. J. 49, 798–803.

Nielsen, D.R., Alemi, M.H., 1989. Statistical opportunities for analyzing spatial and temporal heterogeneity of field soils. Plant Soil 115, 285–296.

Nielsen, D.R., Wendroth, O., 2002. Spatial and temporal statistics—sampling field soils and their vegetation, Geoecology text book, Catena, Reiskirchen (in press).

Nielsen, D.R., Katul, G.G., Wendroth, O., Folegatti, M.V., Parlange, M.B., 1994. State-Space Approaches to Estimate Soil Physical Properties from Field Measurements, Proceedings on 15th Conference on ISSS, vol. 2a., pp. 61–85.

Nielsen, D.R., Wendroth, O., Pierce, F.J., 1999. Emerging concepts for solving the enigma of precision farming research. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Proceedings of Fourth International Conference on Precision Agriculture, Minneapolis, MI, pp. 303–318.

Pennock, D.J., Corre, M.D., 2001. Development and application of landform segmentation procedures. Soil Till. Res. 58, 151–162.

Reuter, H.I., Wendroth, O., Kersebaum, K.C., Schwarz, J., 2001. Solar radiation modelling for precision farming—a feasible approach for better understanding variability of crop production. In: Grenier, G., Blackmore, S. (Eds.), ECPA 2001. Proceedings of Third European Conference on Precision Agriculture, Montpellier, France, pp. 845–850.

Sadler, E.J., Busscher, W.J., Bauer, P.J., Karlen, D.J., 1988. Spatial scale requirements for precision farming: a case study in the Southeastern USA. Agron. J. 90, 191–197.

Shaw, J.N., Carter, E.A., 2002. Timber harvesting effects on spatial variability of southeastern US Piedmont soil properties. Soil Sci. 167, 288–302.

Shumway, R.H., 1985. Time series in soil science: is there life after kriging. In: Nielsen, D.R., Bouma, J. (Eds.), Soil Spatial Variability. Proceedings Workshop ISSS/SSSA, Las Vegas, USA,.

Shumway, R.H., 1988. Applied Statistical Time Series Analysis, Prentice-Hall, Englewood Cliffs, NJ, 379p.

Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal. 3, 253–264.

Shumway, R.H., Stoffer, D., 2000. Time Series Analysis and its Applications, Springer, New York, 549p.

Shumway, R.H., Biggar, J.W., Morkoc, F., Bazza, M., Nielsen, D.R., 1989. Time- and frequency-domain analyses of field observations. Soil Sci. 147, 286–298.

Stafford, J.V., 1999. An investigation into the within-field spatial variability of grain quality. In: Stafford, J.V., (Ed.), Precision Agriculture, Proceedings of Second European Conference On Precision Agriculture, Odense, Denmark, Academic Press, Sheffield, UK, pp. 353–361.

Timlin, D.J., Pachepsky, Y., Snyder, V.A., Bryant, R.B., 1998. Spatial and temporal variability of corn grain yield on a hillslope. Soil Sci. Soc. Am. J. 62, 764–773.

Wendroth, O., Al-Omran, A.M., Kirda, C., Reichardt, K., Nielsen, D.R., 1992. State-space approach to spatial variability of crop yield. Soil Sci. Soc. Am. J. 56, 801–807.

Wendroth, O., Jürschik, P., Kersebaum, K.C., Reuter, H.I., Van Kessel, C., Nielsen, D.R., 2001. Identifying, understanding, and describing spatial processes in agricultural landscapes—four case studies. Soil Till. Res. 58, 113–128.

Wenkel, K.O., Brozio, S., Gebbers, R.I.B., Kersebaum, K.C., Lorenz, K., 2001. Development and evaluation of different methods for site-specific nitrogen fertilization of winter wheat. In: Grenier, G., Blackmore, S. (Eds.), ECPA 2001. Proceedings of Third European Conference on Precision Agriculture, Montpellier, France, pp. 743–748.