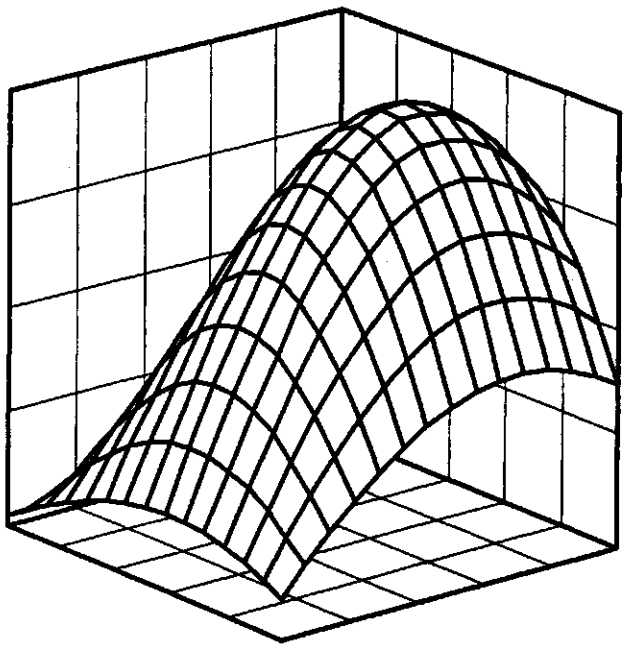
Models in action

Proceedings of a seminar series 1995/1996



PRODUCTION ECOLOGY

ab-dlo

Quantitative Approaches in Systems Analysis

Quantitative Approaches in Systems Analysis supports publications in reviewed journals on dynamic simulation models, optimization programs, Geographic Information Systems (GIS), expert systems, data bases, and utilities for the quantitative analysis of agricultural and environmental systems. The series provides staff members, students and visitors to AB-DLO and PE with an opportunity to publish updates of previously published models, extensive data sets used for validation, background material to refereed journal articles, and other documentation that is valuable to others. The inclusion of listings of programs in an Appendix is encouraged.

All manuscript are reviewed by an editorial board comprising one AB-DLO and one PE staff member: B.A.M. Bouman (AB-DLO) and M.K. van Ittersum (TPE-WAU). The editorial board may consult external reviewers. The review process includes assessing the following: relevance of the topic to the series, overall scientific soundness, clear structure and presentation, and completeness of the presented material(s). The editorial board evaluates manuscripts on language and lay-out matters in a general sense. However, the sole responsibility for the contents of the reports, the use of correct language and lay-out rests with the authors. Manuscripts or suggestions should be submitted to the editorial board. Reports of the series are available on request.

Quantitative Approaches in Systems Analysis are issued by the DLO Research Institute for Agrobiology and Soil Fertility (AB-DLO) and The C.T. de Wit Graduate School for Production Ecology (PE). AB-DLO, with locations in Wageningen and Haren, carries out research into plant physiology, soil science and agro-ecology with the aim of improving the quality of soils and agricultural produce and of furthering sustainable plant production systems.
The 'Production Ecology' Graduate School explores options for crop production systems associated with sustainable land use and natural resource management; its activities comprise research on crop production and protection, soil management, and cropping and farming systems.

Address for ordering copies of volumes in the series:
H.E. de Ruiter
TPE-WAU
Bornsesteeg 47
NL-6708 PD Wageningen
Phone: 31(0).317.475770
Fax: 31(0).317.484892
E-mail: hennie.deruiter@staff.tpe.wau.nl

Addresses of editorial board (for submitting manuscripts):

B.A.M. Bouman
AB-DLO
P.O. Box 14
NL-6700 AA Wageningen
Phone: 31(0).317.482382
Fax: 31(0).317.423110
E-mail: bouman@ab.dlo.nl

M.K. van Ittersum
TPE-WAU
Bornsesteeg 47
NL-6708 PD Wageningen
Phone: 31(0).317.475972
Fax: 31(0).317.484892
E-mail: martin.vanittersum@staff.tpe.wau.nl

Models in action

Proceedings of a seminar series 1995/1996

A. Stein, F.W.T. Penning de Vries & P.J. Schotman

Quantitative Approaches
in Systems Analysis No. 6
June 1996

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

C. Rappoldt & D.W.G. van Kraalingen

Models in action. Proceedings of a seminar series 1995/1996
/ A. Stein, F.W.T. Penning de Vries & P.J. Schotman -
Wageningen : DLO Research Institute for Agrobiology and Soil Fertility;
Wageningen : The C.T. de Wit Graduate School for Production Ecology. -
(Quantitative approaches in systems analysis ; no. 6)
ISBN 90-73384-45-1
NUGI 835
Subject headings: simulation models, models and statistics, models and scale, models and Goal oriented Information Systems, stages of life of models

Keywords
Statistics, modelling, information systems, GIS, scale, philosophy

Guidelines 'Quantitative Approaches in Systems Analysis'

Manuscripts or suggestions should be submitted to the editorial board (B.A.M. Bouman, AB-DLO, or M.K. van Ittersum, TPE-WAU). The final version of the manuscripts should be delivered to the editors camera-ready for reproduction. The submission letter should indicate the scope and aim of the manuscript (e.g. to support scientific publications in journals, program manual, educational purposes). The costs of printing and mailing are to be borne by the authors.

The English language is preferred. Authors are responsible for correct language and lay-out. Overall guidelines for the format of the texts, figures and graphs can be obtained from the publication editor at AB-DLO, or from the PE office:

J.M. Braber
AB-DLO

P.O. Box 14
NL-6700 AA Wageningen
Phone: 31(0).317.475723
Fax: 31(0).317.423110
E-mail: braber@ab.dlo.nl

Th.H. Jetten
Secretariat C.T. de Wit Graduate School
for Production Ecology
Lawickse Allee 13
NL-6701 AN Wageningen
Phone: 31(0).317.485116
Fax: 31(0).317.484855
E-mail: theo.jetten@beleid.spp.wau.nl

Preface

Models evolve. Specialized models are constructed, and are sometimes applied for practical purposes. When constructing models, important decisions are made. On the other hand, models age, but the scientific core improves and extends, challenging new applications. Models are constructed on the basis of existing knowledge with respect to plant, soil, environment and are put into action. Finally, models go out of date and are replaced by newer models.

Apart from scientific disciplines, therefore, general aspects on modelling are distinguished. During four seminars organized by the C.T. de Wit Research School for Production Ecology attention was given to these aspects for agricultural modelling. Important questions to be addressed were statistics, scale, information systems and the future of modelling as such.

During the four seminars, communications were given by distinguished national and international speakers. Each presentation was followed by a concise discussion. Each afternoon, the seminars were concluded by a general discussion. The speakers were invited to summarize their contribution as a paper, based upon previous scientific publications which are collected in this special issue of Quantitative Approaches in System Analysis. The papers are a review rather than an original contribution.

At this place we like to thank the C.T. de Wit Research School for Production Ecology for their financial and organizational support.

A. Stein	Dept. of Soil Science and Geology, Wageningen Agricultural University
F.W.T. Penning de Vries	DLO Research Institute for Agrobiology and Soil Fertility
P.J. Schotman	Dept. of Information Sciences, Wageningen Agricultural University

Table of Contents

1. Models and statistics

Given a practical problem, a model is constructed and applied. Many statistical question come into view. Statistics is applied primarily to quantify uncertainties. Also, the importance of some variables is stressed as compared to other variables. In this chapter some statistical questions occurring at the many stages of model development are highlighted. Since the realm of statistics is broad, attention is focused on those statistical methods that are closely related to modelling.

1.1 Spatial and Temporal Variations with Applications in Agriculture

Annette Kjaer Ersbøll

The Royal Veterinary and Agricultural University, Department of Mathematics and Physics, Thorvaidsensvej 40, 1871 Frederiksberg C, Denmark

Classical statistics assume that sampling units are independent and therefore contain no reference to their spatial distribution. Therefore, knowledge concerning the spatial relation between the observations are not included. Similarly, repeated sampling of measurements in the experiment introduce temporal correlations. Agricultural experiments often contain at least one of these two items, i.e. correlation in space or time. It seems therefore natural to incorporate and utilize these correlations in the analysis. In the spatial analysis of field trials the variation between measurements at a certain distance can be modeled using the semivariogram. The knowledge of the spatial structure can be utilized in e.g. the statistical analysis of the field trial, the design of new experiments and spatial estimation of the measurements by kriging. The spatial semivariogram is briefly mentioned and estimation using kriging is described. Suggestions to extensions 'of the spatial semivariogram are given in which repeated measurements are included in the spatio-temporal semivariogram models. Examples of applications in agriculture are given.

1. Spatial variation

1.1 Semivariogram

Let $Z(x)$ be a random function and let $Z(x_i)$ be the spatial (regionalized) variable, with position x_i in a d-dimensional space $D \subset R^d$,

$$Z(x) = \{Z(x_i),\ \forall_{xi} \in D\}. \tag{1}$$

Let the random function $Z(x)$ be given as

$$Z(x) = \mu(x) + \varepsilon(x) \tag{2}$$

where $\mu(x)$ is the mean and $\varepsilon(x)$ is a (zero mean) stochastic process at position x.
The random function $Z(x)$ is said to be intrinsic stationary when:
1. The expectation exists and is independent of the position x

$$E\{Z(x)\} = \mu\ \forall_x . \tag{3}$$

2. The variance between $Z(x)$ and $Z(x+h)$ is finite and does not depend on the position x. It is defined as

$$Var\{Z(x) - Z(x+h)\} = E\{[Z(x) - Z(x+h)]^2\} = 2\gamma(h), \ \forall_x \quad (4)$$

where $\gamma(h)$ is the semivariogram at distance h.

Furthermore, the random function $Z(x)$ is defined to be second-order stationary when (1) is satisfied and the covariogram (covariance function) exists. The covariance $C(h)$ is defined as

$$C(h) = E\{Z(x+h)Z(x)\} - \mu^2, \ \forall_x \quad (5)$$

where h is a vector of coordinates. The semivariogram $\gamma(h)$ is said to be isotropic if it is a function only of the distance and thereby directionally independent.

Given the observations $z(x_i)$, $i = 1, \ldots, n$ where n is the number of experimental units (e.g. plots) and xi is the position, then the semivariogram can be estimated. Probably the most frequently used estimator, originally proposed by Matheron (Cressie, 1991), is

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (z(x_i) - z(x_i+h))^2 \quad (6)$$

where $n(h)$ is the number of pairs of observations at a distance h from each other.

Cressie & Hawkins (1980) suggest other more robust semivariogram estimators as e.g.

$$\dot{\gamma}(h) = \frac{\left(\frac{1}{n(h)} \sum_{i=1}^{n(h)} \sqrt{|z(x_i) - z(x_i+h)|} \right)^4}{2\left(0.457 + \frac{0.494}{n(h)} \right)} \quad (7)$$

The semivariogram models are (in principle) characterized by the nugget variance (due to the phenomenon nugget effect), the sill and the range of influence. The nugget variance C_0 is a positive finite value to which $\gamma(h)$ approaches as h approaches 0. This nugget effect may be due to small-scale variations less than the smallest sampling distance, measurement errors etc.

In most cases $\gamma(h)$ increases with increasing distance h to a maximum (approximately the total variance of the data). This distance is the range of influence a. The semivariance $C_0 + C$ at distance a is called the sill. Models with a sill are called transition models (Journel & Huijbregts, 1978) and the sill value of a transition model is the a priori variance of $Z(x)$.

A semivariogram can be modeled by different functions (Journel & Huijbregts, 1978; Cressie,

1991). The two most commonly used semivariogram models are the spherical and exponential models.

Spherical model with nugget effect:

$$\gamma_S(h)=\begin{cases}0 & h=0\\ C_0+C\left[\frac{3}{2}\left(\frac{h}{a}\right)-\frac{1}{2}\left(\frac{h}{a}\right)^3\right] & 0<h<a\\ C_0+C & h \geq a\end{cases} \tag{7}$$

Exponential model with nugget effect:

$$\gamma(h) = \begin{cases}0 & h=0\\ C_0+C\left[1-\exp(-\frac{h}{a})\right] & h>0\end{cases} \tag{8}$$

1.2. Applications

As an example, the semivariogram for yield measurements in a Danish uniformity trial with spring barley at Jyndevad Experimental Station is given figure 1. The trial was laid out on sandy soil (clay content approximately 4 %) as 26 rows and 20 plots within each row, each plot with the dimension 1.5 × 8.5 m^2. A strong anisotropy has been estimated and data has been adjusted towards isotropy. The anisotropy ratio (β/α) and angle ψ have been estimated using the directional function given by Burgess et al. (1981)

$$\Omega(\theta)=\sqrt{\alpha^2\cos^2(\theta-\psi)+\beta^2\sin^2(\theta-\psi)} \tag{9}$$

The knowledge of the spatial structure in the uniformity trial at Jyndevad has been used in spatial experimental design of new field trials (Kristensen & Ersbøll, 1992). The consequence of changing plot and block size and shape can be estimated as the residual variance. Different layouts can in this way be compared and the most optimal one can be chosen.

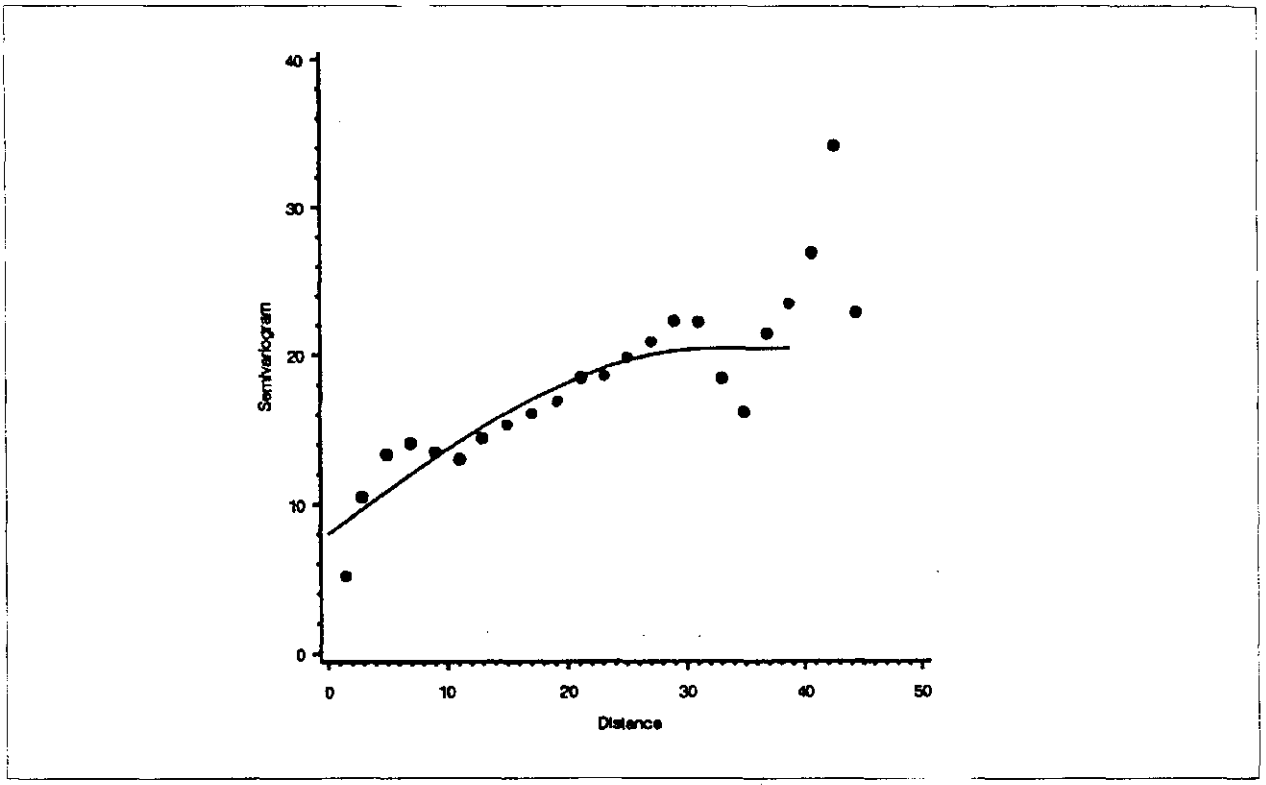


Figure 1: Empirical semivariogram and fitted spherical model for the Jyndevad uniformity trial. A strong anisotropy has been estimated and data has been adjusted towards isotropy.

2 Estimation using kriging

2.1 Kriging

Kriging is a local estimation method which provides the best linear unbiased estimator (BLUE) of the response variable $Z(x)$ based upon the observations $z(x_1)$, $z(x_2)$,. .., $z(x_n)$. Let $Z(x)$ be the random function in study defined on supports v centered in x_i and either intrinsic or second-order stationary. The estimate of Z over a volume V centered in x_0 is

$$z_V(x_0)=\sum_{i=1}^{n} \lambda_i z_V(x_i) \tag{10}$$

where λ_i, $i = 1, ., . ., n$ are weights assigned to the sampling point. An unbiased estimator with minimum estimation variance is obtained when

$$\begin{aligned} &\sum_{i=1}^{n} \lambda_i = 1 \\ &\sum_{i=1}^{n} \lambda_i \bar{\gamma}(v_i, v_j) + \psi = \bar{\gamma}(v_j, V), \quad \forall j = 1, \ldots, n \end{aligned} \tag{11}$$

where

ψ is a Lagrange multiplier introduced to ensure minimization
$\bar{\gamma}$ is the average semivariogram,

The estimation variance or kriging variance is then given by

$$\sigma_V^2 = \sum_{i=1}^{n} \lambda_i \bar{\gamma}(v_i, V) + \psi - \bar{\gamma}(V, V) \tag{12}$$

The optimal weight λ_i, $i = 1, \ldots, n$ concerning non-bias and minimum estimation variance is obtained by Lagrangian techniques. This procedure provides a system of $n+1$ linear equations in $n+1$ unknowns which is the "kriging system" (or "kriging equations"), given in equation (11). This estimation procedure is called ordinary kriging.

With point support and estimation at points the estimate, estimation variance and the kriging system is reduced to

$$\begin{aligned} z(x_0) &= \sum_{i=1}^{n} \lambda_i z(x_i) \\ \sigma_{x_0}^2 &= \sum_{i=1}^{n} \lambda_i \gamma(x_i, x_0) + \psi \end{aligned} \tag{13}$$

Formulation of the kriging system assumes stationarity of the random variable $Z(x)$.
However, in practice drifts can be seen when

$$E\{Z(x)\} = \mu(x) \tag{14}$$

where $\mu(x)$ is the drift e.g. linear or quadratic. This drift can be taken into account in the estimation by introducing the drift parameters in the kriging system (cf. Journel & Huijbregts (1978)). This estimation procedure is called universal kriging.

Hawkins & Cressie (1984) suggest an approach for robust kriging which is more robust against outliers or extreme observations compared to ordinary kriging. A robust semivariogram estimator is used and data are weighted (or edited) depending on the degree of contamination.

2.2 Applications

The grain yield measurements in the Jyndevad uniformity trial have been estimated using ordinary kriging. The original yield measurements are seen in Figure 2. The kriged (estimated) yield measurements are seen in Figure 3 and the corresponding kriging variances are given in Figure 4. Another area in field experimentation where kriging can be useful is in relation to characterization and description of weed plants in a field. An example is a field trial with reduced input of herbicides where the number of weed plants have been estimated using kriging (Ersbøll et al. 1994; Heisel et al. 1995).

3 Spatio-temporal semivariogram

Spatio-temporal models are models which in some way include both the spatial and temporal properties of the data. In relation to field experiments the combined spatio-temporal models may have different applications which can support or even improve the models and analysis obtained with more traditional methods.

As an example consider the variance-covariance structure of a field experiment with repeated measures. The variance-covariance matrix can be estimated- using a spatio-temporal covariogram (or semivariogram). A (presumably) improved analysis of variance of the repeated measurements can then

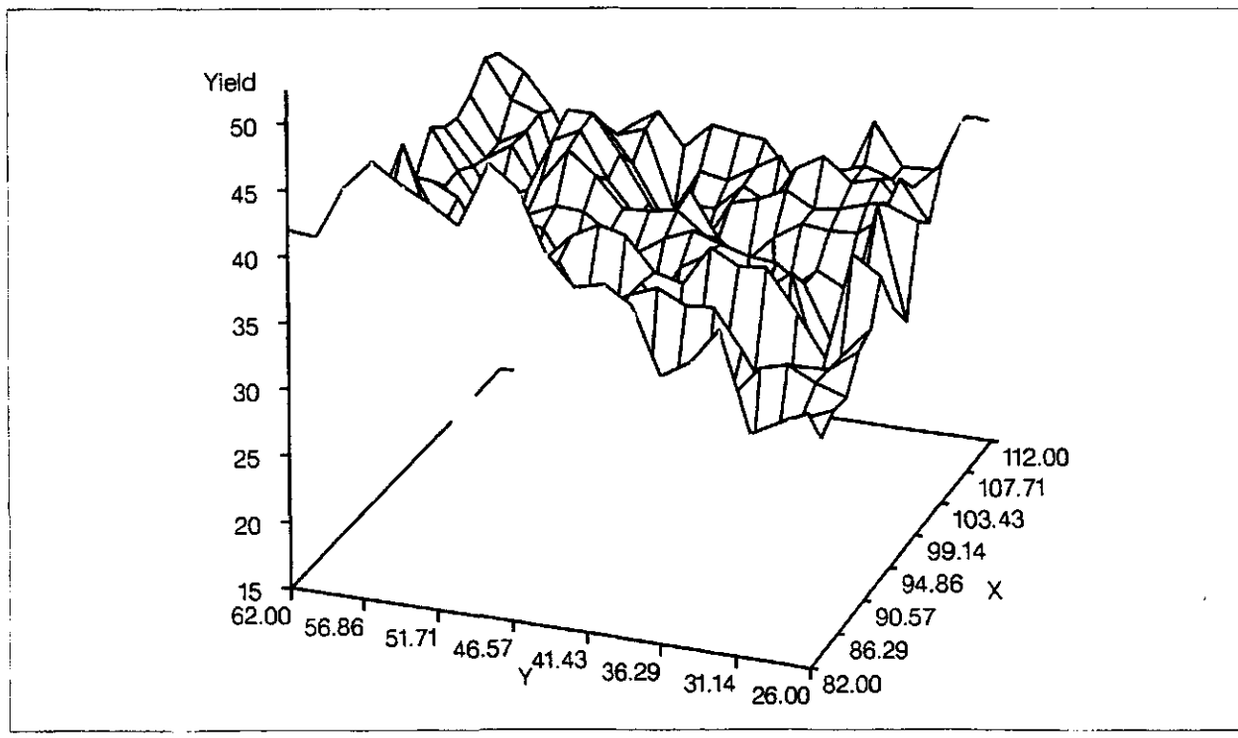


Figure 2. The grain yield measurements in the Jyndevad uniformity trial.

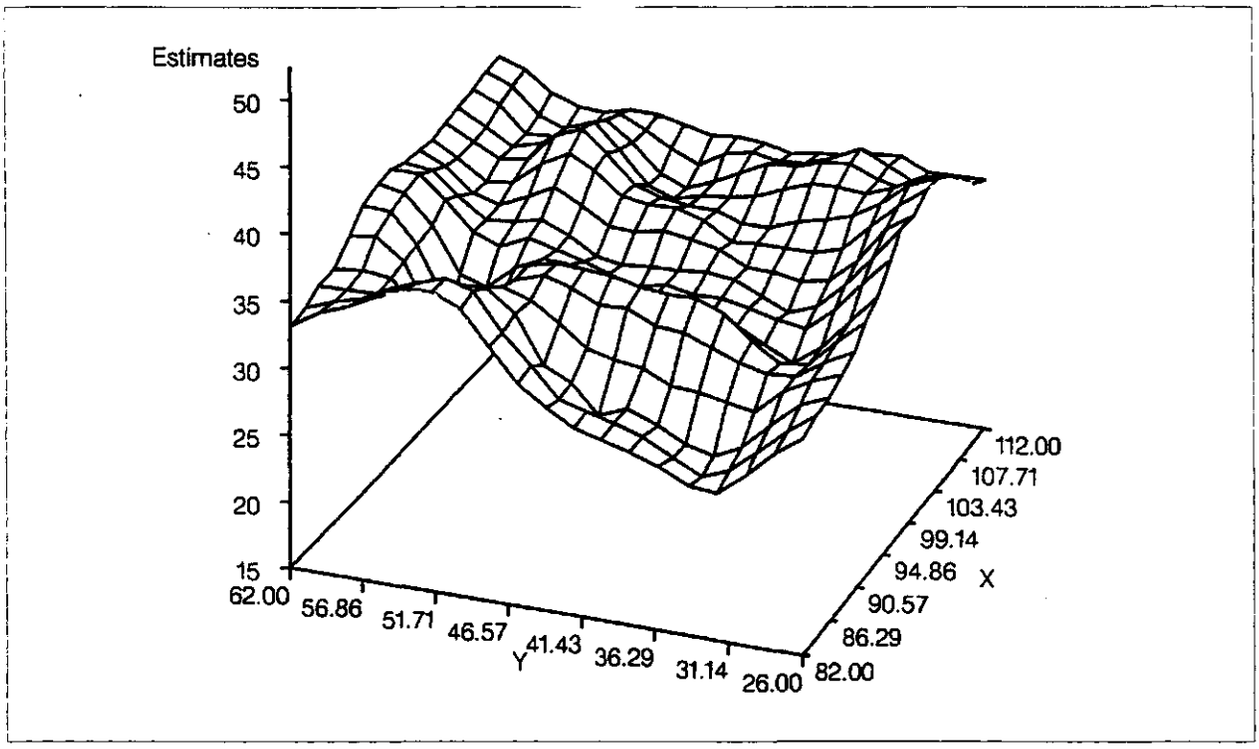


Figure 3. The kriged grain yield measurements in the Jyndevad uniformity trial using ordinary kriging.

be performed using the variance-covariance matrix by including both the spatial and temporal

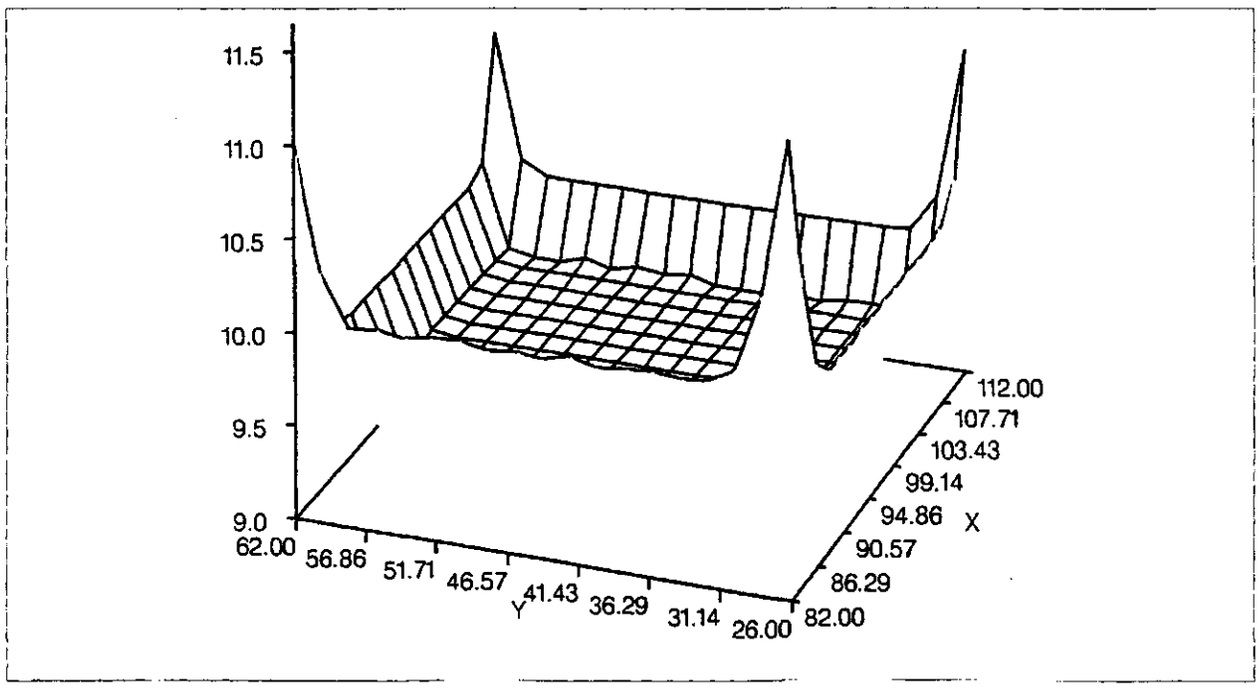


Figure 4. The kriging variance for the kriged grain yield measurements in the Jyndevad uniformity trial.

correlations.

It should also be mentioned that spatial estimation of repeated measurements using kriging can be done using spatio-temporal models. The kriging estimation is then based on a semivariogram including variations between measurements at a certain distance as well as measurements with a certain separation in time at the same or at different plots. Two approaches are given for modeling the spatio-temporal semivariogram. A simple procedure given by Posa (1993) is described, and a suggestion to extend the model is given by Ersbøll (1994).

3.1 Simple spatio-temporal semivariogram

Posa (1993) has suggested a simple spatio-temporal model for the combined semivariogram. By assuming a nugget effect and a range of influence independent of the time the spatio-temporal covariance is

$$C(h|t) = C_t\rho(h) \tag{15}$$

where C_t is the partial sill at time t and $\rho(h)$ is the correlation model of measurements with distance h. The semivariogram model is given by

$$\gamma(h|t)=\begin{matrix} 0 & h=0 \\ C_0+C_t\rho(h) & h>0 \end{matrix} \tag{16}$$

in which the correlation model $\rho(h)$ can be e.g. spherical or exponential.

3.2 Extended spatio-temporal semivariogram

Modeling the spatio-temporal semivariogram using the procedure outlined by Posa (1993) is based on strong assumptions about a time independent range of influence and a time independent nugget effect. An (extended) alternative to this procedure is given. The spatio-temporal semivariogram is estimated including both correlations in space and time, and combinations of these. An empirical semivariogram is estimated as

$$\gamma(h,\tau)=\frac{1}{2N(h,\tau)}\sum_{i,j,\tau}(z_i(t)-z_j(t+\tau))^2 \tag{17}$$

where

z_i and z_j are measurements at positions i and j

τ is the separation in time, $\tau \geq 0$

$N(h,\tau)$ is the number of pairs of observations with a distance h and separation in time τ.

A model can then be fitted to the empirical spatio-temporal semivariogram. The usual semivariogram models have to incorporate the time. A suggestion for an exponential model is

$$\gamma(h) = \begin{cases} 0 & h=0,\tau=0 \\ C_0+C\left[1-\exp(-\frac{\tau}{\alpha})\right] & h>0,\tau>0 \\ C_0+C\left[1-\exp(-\frac{h}{a})\right] & h>0,\tau=0 \\ C_0+C\left[1-\exp(-\frac{h}{a}-\frac{\tau}{\alpha})\right] & h>0,\tau>0 \end{cases} \qquad (18)$$

In this model a and α are the ranges of influence in space and time, respectively.

3.3 Applications

Spatio-temporal models have been estimated as combined semivariograms for a long-term field experiment with reduced soil tillage (Rasmussen, 1991). Treatment eliminated residuals have been used in the estimations. The treatment effect has been estimated and eliminated using the procedure of treatment (median) polish, semivariogram estimation and estimation of treatment effects using the covariogram (Ersbøll, 1994).

The simple spatio-temporal semivariogram has been estimated for the experiment. Empirical semivariograms have been estimated for each year of experimentation using the treatment eliminated residuals. The simple spatio-temporal model has then been fitted to the empirical semivariograms assuming a time independent range of influence and nugget effect. An exponential model without a nugget effect has been fitted

$$\gamma(h|t) = \begin{cases} 0 & h=0 \\ C_t\left(1-\exp(-\frac{h}{a})\right) & h>0 \end{cases} \qquad (19)$$

The common range of influence a is estimated at 8.67 m.. The sills C_t for each year of experimentation lie within the range. 3.1-26.9 $(kg/ha)^2$. In figure 5 the 14 fitted semivariograms are shown.

An exponential model with nugget effect has also been fitted to the empirical semivariograms. This has been performed with one restriction in order to obtain convergence. The semivariogram for the year 1974 has a constant semivariogram equal to the sill at all distances. The spatio-temporal semivariogram with nugget effect has a range of influence at 17.96 m, about the double of the range of influence when no nugget effect is included in the model. Although, introducing a nugget effect,

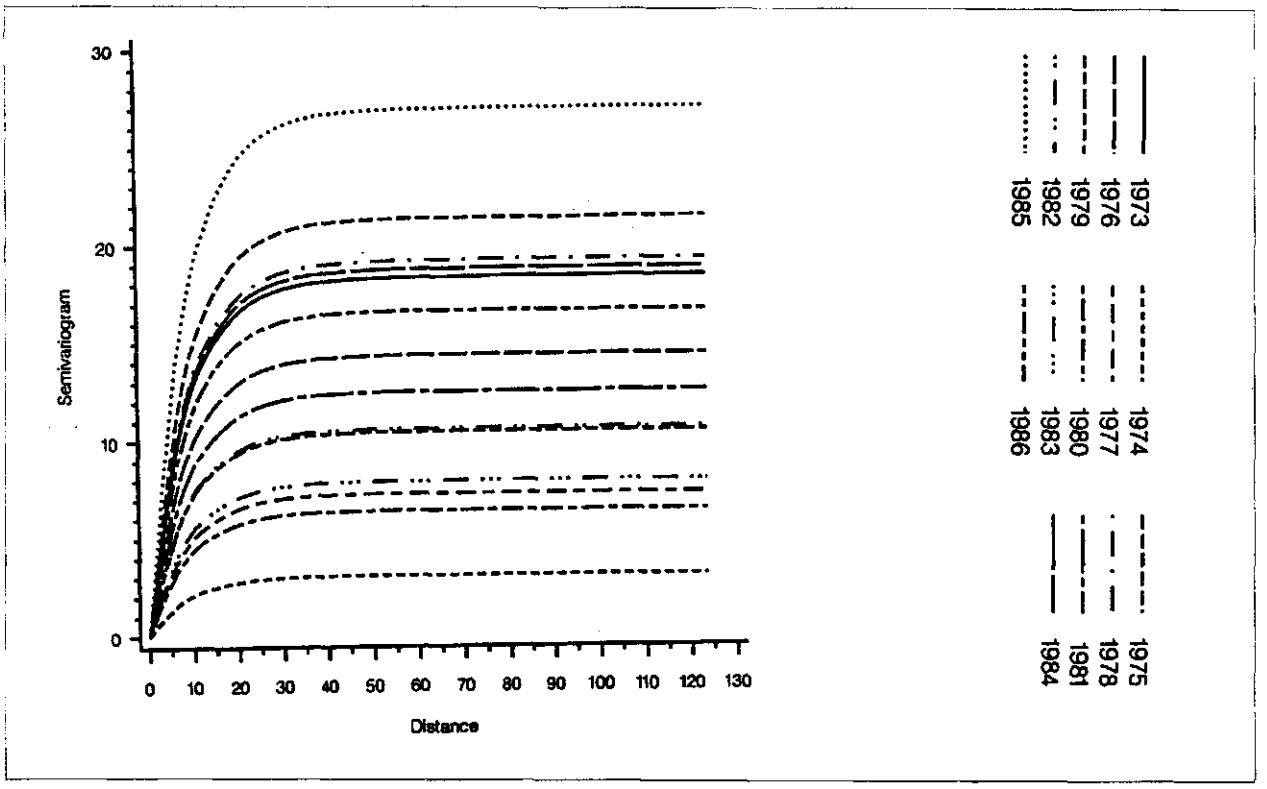


Figure 5: The fitted simple spatio-temporal semivariograms for the 14 years of experimentation. An exponential model without nugget effect has been used. The long-term field experiment with reduced soil tillage

the sills $C_0 + C_t$ are similar to the previous sills, estimated without a nugget effect.

3.4 Extended spatio-temporal semivariogram

The extended procedure to estimate the spatio-temporal semivariogram has also been used for the reduced soil tillage experiment. The empirical spatio-temporal semivariogram is shown in Figure 6. An exponential model with nugget effect as model (18) is fitted to the empirical semivariogram as

$$\gamma(h) = \begin{cases} 0 & h=0,\tau=0 \\ 6.746+8.452\left[1-\exp\left(-\frac{\tau}{0.440}\right)\right] & h>0,\tau>0 \\ 6.746+8.452\left[1-\exp\left(-\frac{h}{25.862}\right)\right] & h>0,\tau=0 \\ 6.746+8.452\left[1-\exp\left(-\frac{h}{25.862}-\frac{\tau}{0.440}\right)\right] & h>0,\tau>0 \end{cases} \quad (20)$$

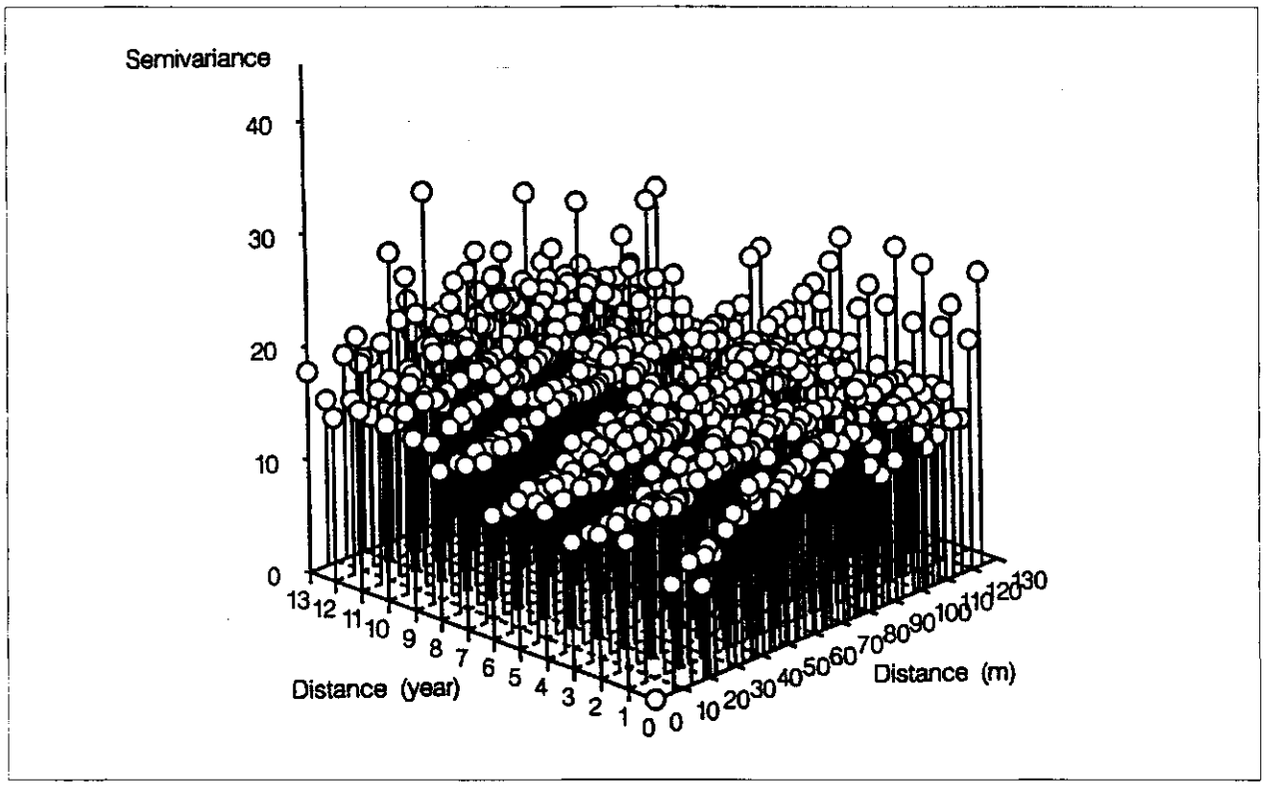


Figure 6. The empirical (extended) spatio-temporal semivariogram for the long-term field experiment with reduced soil tillage.

with effective ranges of influence estimated at 77.6 m and 1.3 years, respectively. Standard erros of the parameters are given in Table 1. The nugget effect and sill are 6.7 and 15.2 $(kg/ha)^2$, respectively. The fitted spatio-temporal semivariogram is shown in Figure 7.

The nugget effect and sill estimated for this extended spatio-temporal semivariogram are of similar size as the parameters estimated seen using the simple spatio-temporal semivariogram. The estimate of the range of influence in space with the extended model is greater than the corresponding estimate for the simple model. This might be due to a possible correlation between neighboring plots Table 1: Asymptotic standard errors of the parameters for extended spatio-temporal exponential model with nugget effect.

Parameter	Asymptotic standard error
a	1.747
α	0.120
C_0	0.316
C	0.301

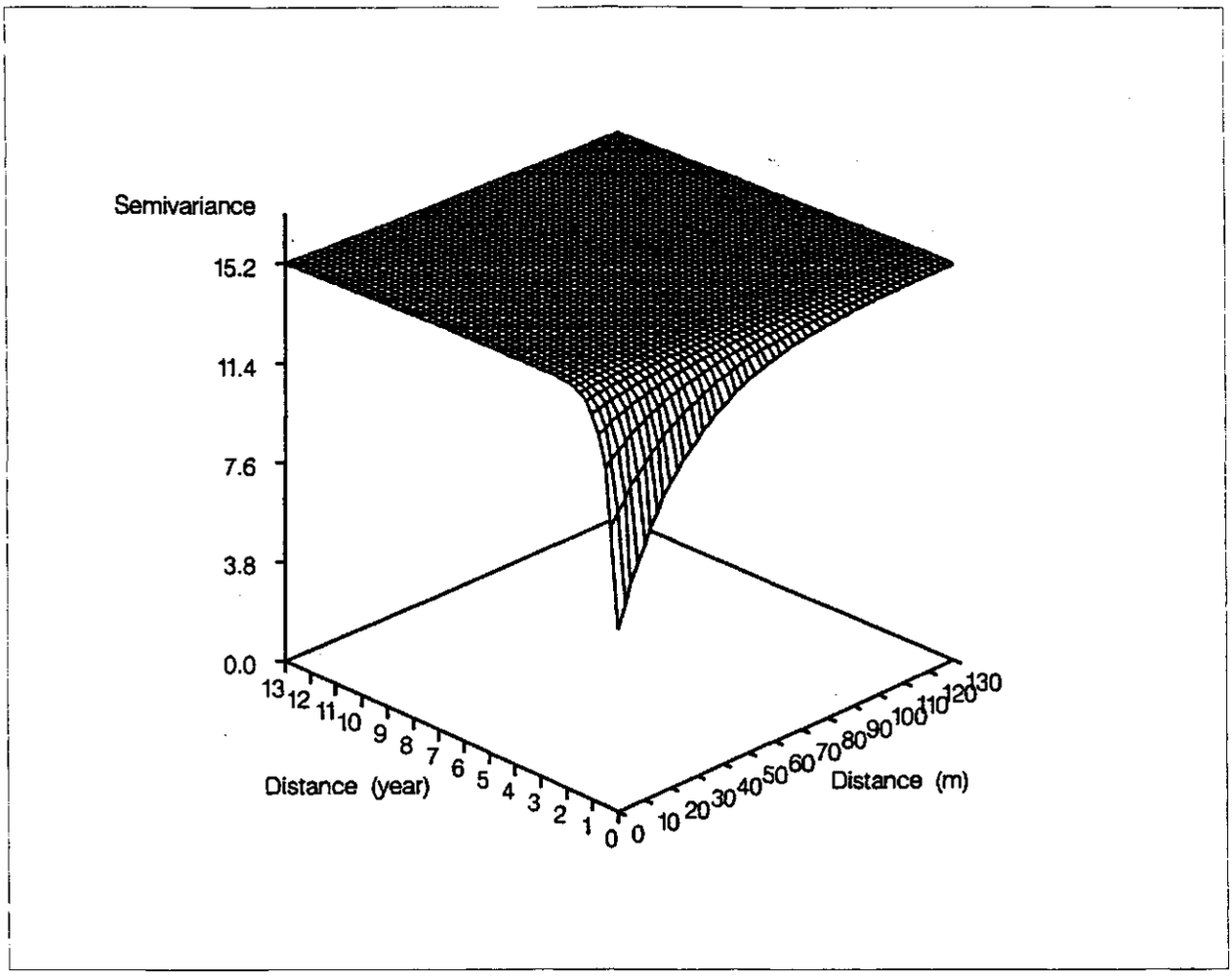


Figure 7: An exponential model with nugget effect fitted to the extended spatio-temporal semivariogram. - The long-term field experiment with reduced soil tillage.

at different years, which are included in the extended model, but not in the simple model. However, the correlation between years seems to be very small, only a few years.

4. References

Burgess, T. M., Webster, R., McBratney, A. B. 1981. Optimal interpolation and isarithmic mapping of soil properties. VI. Sampling strategy. Journal of Soil Science 321, 643-659.

Cressie, N. Hawkins, D. M. 1980. Robust estimation of the variogram: I. Mathematical Geology 12, 115-125.

Cressie, N. 1991. Statistics for Spatial Data. John Wiley and Sons, New York.

Ersbøll, A.K., Kristensen, K., Nordbo, E., Christensen, S. 1994. Estimation af ukrudtsmaengde ved kriging. In Boelskifte, S. (Ed.), Symposium i Anvendt Statistik, 19921 pp. 140-151 Aalborg. UNI-C.

Ersbøll, A. K. 1994. On the Spatial and Temporal Correlations in Experimentation with Agricultural

Applications. Ph.D. thesis, Institute of Mathematical Modelling, Technical University of Denmark. IMM-PHD-1994-7.

Hawkins, D. M., Cressie, N. 1984. Robust kriging - a proposal. Mathematical Geology 16, 3-18.

Heisel, T., Andreasen, C., Ersbøll, A. 1995. Weed densities can be mapped with kriging. Weed Research.

Journel, A. G., Huijbregts, C. J. 1978. Mining Geostatistics. Academic Press, London.

Kristensen, K., Ersbøll, A. K. 1992. The use of geostationary methods in planning variety trials. Biuletyn Oceny Odmian 24-25, 139-157.

Posa, D. 1993. A simple description of spatial-temporal processes. Computational Statistics and Data Analysis 15, 425-437.

Rasmussen, K.J. 1991. Reduced soil tillage and Italian ryegrass as catch crop. I. Growing conditions, yields, crop analyses and weeds. Tidsskrift for Planteavl 95, 119-138.

1.2. Model Building in Animal Breeding

R. Thompson[1] and R. E. Crump[2]

[1]IACR-Rothamsted, Harpenden, Herts. AL5 2JQ,United Kingdom
[2]Roslin Institute, Edinburgh 1, Roslin, Midlothian EH25 9PS, United Kingdom

This paper discusses the various stages of setting up a system to allow the genetic evaluation of pedigree beef cattle in the United Kingdom. These include developing models to take account of various environmental factors, including herd, season and contemporary group, and modelling genetic factors including maternal effects. Various approaches to validation are considered, including novel experimental designs.

1. Introduction

For about 30 years, the performance of pedigree beef cattle in the UK has been recorded as an aid to selection by pedigree breeders. Records of performance have been mostly collected on-farm although, until the mid-1980s, some were on central performance stations. The original schemes concentrated on measuring weights at approximately 100 day intervals, but now other traits including calving ease, ultrasonic fat depth, muscling score and ultrasonic muscle depth are measured on some animals.

The scientific question is how should beef cattle breeders choose animals as parents, both from their own and other herds. The aim of breeders should be to choose parents so as to improve the economic merit of their offspring. This economic worth depends on value of the carcase of the animal, which in turn depends on the weight, composition and conformation of the carcase and on the value of the food eaten by the animal. The process of selection is made more difficult by the fact that the carcase value is only available on dead animals and we need live cows to produce calves. Another difficulty is that we do not observe the genotypic effect of an animal alone, we observe a phenotypic value, a combination of genetic and environmental effects. A further difficulty is that all animals are not observed in the same environment but in different herds in different years. In Britain, the average beef herd size is low (5 to 12 cows for the major beef breeds; Simm et al. 1990); this leads to low accuracies of predicted breeding values.

2. Modelling environmental factors

The weight measurements are influenced by several environmental factors and there is a need to adjust the measurements for these effects in genetic analysis. Alternative linear models were investigated . The different weights were not taken exactly at 200 and 400 days, so linear and quadratic adjustments for age were used. Sex, male and female, effects were included in the model. In order to adjust for herd, season and management effects, animals were grouped into contemporary groups. The groups were small, of the order of 6-10 animals, and variable as

different calving patterns existed. Various definitions of contemporary groups were compared, including having fixed time periods and natural periods. Three parameters MXDIF1, MXDIF2 and MING were used to construct different schemes. The first two parameters were associated with the range of birth dates within groups and the third with the minimum size of the group. Natural periods involved grouping together animals according to their dates of birth provided the difference between the first and last animal born was less than MXDIF1 months. If the group size was less than MING, adjacent groups were merged together provided the dates of birth of the first and last born animals in the combined group was less than MXDF2 months. Schemes were compared by constructing the accuracy of mass selection, which depends on the residual variance and the average size of a contemporary group.

There is an optional choice for the number of groups. As the number of groups increases, the residual variance decreases, but the group size decreases and so the number of comparisons within groups decrease. Crump *et al.* (1996) found that MXDIF1 = 3 months, MXDIF2 = 6 months, and MING = 5 records, were reasonable choices of parameters in 5 beef breeds.

3. Modelling Genetic Factors

The major problem of not observing the animal's genetic value and only its phenotypic value is circumvented by essentially regressing the genotype on the phenotype. So that

$$\text{estimated genotype} = Reg_{GP} \times \text{phenotype.} \quad (1)$$

In the simplest case, Reg_{GP} depends on the genetic (σ^2_G) and residual (σ^2_e) variation, so that

$$Reg_{GP} = \sigma^2_G / (\sigma^2_G + \sigma^2_e). \quad (2)$$

When we have related animals, we can use

$$\text{estimated genotypes} = Reg_{GP} \times \text{phenotypes} \quad (3)$$

with

$$Reg_{GP} = G\,(G+E)^{-1} \quad (4)$$

where G describes genetic variation and covariation, and E residual variation and covariation.

Usually, we use Fisher's infinitesimal model assuming a large number of genes, each of small effect (Fisher, 1918). Then the variance of an individual is σ^2_G and the covariance between parent and offspring is $(1/2)\,\sigma^2_G$ and between parent and descendant t generations apart is $(1/2)^t\,\sigma^2_G$. This is a geometric decline with generations. The other important covariances are those between full-sibs $(1/2)\sigma^2_G$ and between half-sibs $(1/4)\sigma^2_G$. The environmental variation matrix E in this application is diagonal, but in other applications can be modified; for example, to take account of common environmental effects when litters of pigs are reared in different pens.

Of course, in applications, we need to adjust phenotypes for environmental effects using linear models discussed in the previous section. Equations can be set up similar to those for least squares, to simultaneously estimate environmental effects and predict breeding values. These equations, known in the animal breeding literature as Henderson's mixed model equations, take

account of the information that the variance of animal effects is G (Henderson, 1973). These equations give best linear unbiased predictions of genetic effects and give weighted least squares estimates of environmental effects, with weight proportional to $(G+E)^{-1}$. These equations are the kernel of estimating genetic parameters; for example, σ^2_G and σ^2_E, by maximum likelihood (Patterson & Thompson, 1971). These methods depend on weighted sums of squares of residual and terms that depend on the variance of observations, predictions and estimated environmental effects.

In summary, this maximum likelihood approach gives efficient estimates of environmental effects and efficient predictions of breeding values. It also gives efficient estimates of variance parameters. Alternative variance models can be compared by comparing maximum likelihoods under the different models.

One added complexity is that data is available on 27 beef breeds from 1970-1995. The number of records vary from breed to breed and from trait to trait. For example, for three breeds there were over 40,000 records on birth weight, over 30,000 records on 200-day weight, and over 15,000 records on 400-day weight. It is convenient to consider the modelling in two stages: (1) adjustment for environmental effects, and (2) investigation of different genetic models. These beef cattle suckle milk from their mothers until approximately 200 days, so that a calf's 200-day weight is influenced by its mother's milking ability.

In considering birth weight and 200-day weight, it is necessary to consider these possible maternal effects that might have genetic and environmental effects. Mohiuddin (1995) estimated maternal and direct genetic variation in British beef populations, and found that there was maternal genetic variation in birth weight and 200-day weight, but that it was smaller than the direct genetic variation. He also found a negative genetic correlation between direct and maternal genetic variation. These negative genetic correlations have been found in other beef cattle analyses and in analyses of other species but no convincing biological explanation has been given for this.

The development in this section is in terms of single traits but there is no conceptual difficulty in expanding to multivariate cases. Table 1 gives heritabilities ($\sigma^2_G / \sigma^2_G + \sigma^2_e$) and genetic and environmental correlations between traits. From these values G and $P = G + E$ can be calculated. In general, the environmental correlations are smaller than the genetic correlations. There are two reasons for carrying out a multivariate analysis. Firstly, birth weight is correlated with 200-day weight, and so can help in predicting 200-day weight, especially if 200-day weight is not recorded. Secondly, if animals are culled on the basis of recorded performance, the analysis takes account of the culling.

Table 1: Genetic parameters (x 100) used in genetic evaluations. Heritabilities are on the diagonal; genetic and phenotypic correlations are above and below the diagonal, respectively.

	BWT	W200	W400	FAT	MSC
BWT	41	49	53	0	0
W200	35	28	82	0	0
W400	38	72	41	21	10
FAT	0	24	36	29	0
MSC	0	10	30	10	25

4. Validation

There are two cases to consider. The first case is when no intervention is possible; then there is a natural time dimension. It is then natural to compare predictions for time period t with observations up to time period (t-1). This might be considered in a likelihood framework by allowing the variation parameters (σ^2_G and σ^2_E) to vary with time (σ^2_{GT} and σ^2_{ET}). One problem with this approach that tests differences of σ^2_{GT} will generally be of low power, especially in efficient breeding programmes which select parents intensely (Hill, 1990).

One can partly overcome this if intervention is allowed, by not just selecting the highest-ranking animals as parents but also the lowest-ranking animals as parents. Then comparisons between the high and low groups can give more powerful evidence if parameters are changing. Cameron & Thompson (1986) considered an interesting two-variate extension. In this case, one might be interested in genetic variation in X and Y and covariation between X and Y. The classical approach is to select high and low on X and observe responses in X and Y and select on high and low Y and observe responses in X and Y. It was found that selecting as parents animals as extreme as possible (as different from the mean animal as possible) gave substantially more information on the genetic parameters.

5. Conclusion

Work is ongoing to include other measures of traits in the procedure, including calving case ease and muscle depths. Other work is trying to relate more precisely measured traits to the traits of economic value, and it is hoped that trials will be set up to validate some of the predictive equations.

6. Acknowledgment

This work has been supported by the Meat and Livestock Commission, the BBSRC and the Ministry of Agriculture, Fisheries and Food. IACR receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

7. References

Cameron, N.D., Thompson, R. 1986. Design of multivariate selection experiments to estimate genetic parameters. Theor. Appl. Genet. 72, 466-476.

Crump, R.E., Wray, N.R., Thompson, R., Simm, G. 1996. Assigning pedigree beef performance records to contemporary groups taking account of within herd calving patterns. Animal Science (submitted).

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edin. 52, 399-433.

Henderson, C.R. 1973. Sire evaluation and genetic trends. In Proc. Anim. Breed. Genet. Symp., Blacksburg, Virginia, pp. 10-41. Am. Society of Anmial Sciences, Champaign, Illinois.

Hill, W.G. 1990. Considerations in the Design of Animal Breeding Experiments. In: Advances in Statistical Methods for Genetic Improvement of Livestock. Eds: Gianola, D. &

Hammond, K. Springer-Verlag, Berlin, Heidelberg.
Mohiuddin, G., 1995 Estimation of Genetic Parameters for Beef Cattle in UK Herds. PhD thesis, University of Edinburgh.
Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrica 58, 595-554.
Simm, G., Steane, D.E., Wray, N.R. 1990. Developments in beef cattle breeding programmes in Europe. In: Proceedings of the 4th World Congress on Genetics Applied to Livestock Production, Vol. XV, 231-243.

2. Models and scale

Implications of scale in relation to building models are sometimes overlooked, in particular with respect to the use of models for practical problems. Attention is given to the scale-specifics of models: what is it, how to deal with it, up- and down-scaling and elements of inter- and extrapolation.

2.1 Models and Scale: Up- and Down-Scaling

Jo Smith

IACR-Rothamsted, Harpenden, Herts. AL5 2JQ, United Kingdom

1. Introduction

1.1 Scope

The need to change the scale of a developed model is a problem that scientists in most areas of systems modelling must address at some time. In particular, with the further development of modelling tools to solve real problems posed by policy makers, advisers and farmers, modelling systems that can function at a number of different scales are increasingly needed. Such systems should be able to determine the optimum scale of operation, and to select the most appropriate model to use. The aim of this paper is to draw together, from the area of soil/crop systems dynamic modelling of nitrogen and carbon behaviour, some generally applicable methods for up- and down-scaling so that a model performs at the scale at which data are available, and results are needed.

1.2 Definitions

In systems modelling, changing scale is often only considered in terms of changes to the spatial area incorporated in the model: up-scaling is an increase and down-scaling is a decrease in the size of the spatial area units. A small scale refers to large spatial units, whereas a large scale refers to small spatial units. In the context of this paper, a more general definition is used. Changing the scale of a model is defined as a change to the resolution of *any* values output from the model. Changes in scale may involve changing the resolution of *any* of the model input variables. This paper discusses the underlying scientific and modelling developments required to maintain model integrity when changing the resolution of input and output values. Examples are changes in spatial units - crop type from crop/ha to crop/5km^2; temporal units - rate of mineralisation from (kg/ha)/day to (kg/ha)/week; and yield units - nitrogen requirement of the crop from (kg N) / (t Grain) to (g N) / (t Grain).

1.3 Why change scale?

Leffelaar (1990) discussed the level of organisation to which a problem should most appropriately be analysed and itemised the criteria that should be considered when determining the scale at which to perform simulations. These include objectives of the study, level of organisation of the scale or processes, scale of scientific understanding, possibility of clearly defining subsystems, possibility of validating the model of a system and possibility of returning to the field level of organization in the systems synthesis phase. Working within these criteria, two practical reasons for changing scale emerge: *transferability* of the model to novel environments and model *robustness*.

An up-scaled model is often more reliably transferred to novel environments than a model developed with lower resolution. If a model describes the system using accurate scientific

paradigms of the underlying processes, the effects on each process of changing the environment will be simulated, and so the accuracy of the model will be maintained. It is often difficult to determine the underlying mechanisms at the larger scale because there is no opportunity for evaluating the simulation of each process. Therefore it becomes necessary to establish a description of the process mechanisms within a smaller scale model and then adapt the model for use at the larger scale.

A down-scaled model, however, is often more robust than a model developed at the smaller scale. Our understanding of a system can be greatly enhanced by characterising its behaviour using statistical methods alone. General rules may be derived to apply at the larger scale. Although it may not be transferrable to novel environments, such a model is correct within the limits of its development. The statistically derived rules may then be adapted, for instance using scientifically meaningful multiplication constants, so that the model can be used at the smaller scale. The scaled-down model maintains overall accuracy at the large scale, because it is not subject to propagation of errors from the individually modelled sub-processes. As a result, the robustness of the larger scale model should confer robustness to the scaled-down model.

In practice, however, the scale of system operation is dictated by the questions being asked of the system and the data available. As an example, a modelling system incorporating the dynamic field-based nitrogen-turnover model, SUNDIAL (Bradbury et al., 1993; Smith et al., 1996a) and the static catchment nitrogen model, NEAP-N (Miles et al., 1995; Anthony et al., 1996) is asked the questions "What is the peak nitrate concentration of nitrate leached from arable land?" and "When and where will the peak occur?". The questions demand that the models simulate peak nitrate concentration on a daily, weekly, or at least monthly basis for the field. However, if the attached database holds information about the dominant soil and land cover type on only a 1 x 1 km grid-square (e.g. Howard et al., 1995), the available data requires either that the system should run using 1 x 1 km grid-squares or that the data should be interpolated. The SUNDIAL model simulates nitrogen turnover with a weekly time-step, but is developed at the field scale. To use SUNDIAL with the available data either the data must be interpolated to the field scale or the spatial units of the model must be scaled-up to a 1 x 1 km grid-square. The NEAP-N model, conversely, is well adapted to simulate the nitrogen status of the soil on a 1 x 1 km grid-square, but with an annual time-step. To obtain the required result using NEAP-N, the temporal units of the model must be scaled-down to provide results on at least a monthly basis. One way of achieving this is through an algorithm developed to express peak concentration in terms of model parameters. The scale of system operation is then dependent on the relative errors introduced by interpolation of data, spatial up-scaling or temporal down-scaling. The best methodology to use can only be determined by quantifying the likely errors and assessing the need for a model that is transferrable to a novel environment.

1.4 What problems are encountered?

Changes in scale may result in *changes in the scope* of the model. For example, gravitational leaching of water down the soil profile is not an important process in the soil micro aggregate and so may be omitted from an aggregate model (eg. Leffelaar, 1988). Scaling up from the aggregate to the plot requires the additional process of profile leaching to be included. The consequences of omitting or incorporating additional processes must be examined and methods devised to ensure model validity is maintained.

Changes in scale also result in changes in the *heterogeneity* of the values of input variables. Up-scaling reduces accuracy and increases heterogeneity. The increased heterogeneity may be

continuous or discontinuous. For example, the variogram for volumetric water content indicates an increase in heterogeneity over the range 2 to 600 m (Addiscott & Tuck, 1996). Accurate up-scaling from lysimeter to field requires some method of accommodating this increased spatial variability, either by interpolating the input data or by adapting the model. Statistical methods for interpolating continuous data are discussed elsewhere (e.g. Webster & Oliver, 1990). However, the methods for interpolating discontinuous data are not so clearly characterised. An example of the latter is the definition of crop type over the 5 km² data units of a catchment model. There cannot be a continuous variation between crop species, but some method of describing the different crops incorporated in the data unit must nevertheless be established.

Down-scaling, by contrast, increases the accuracy needed in the input variables and reduces heterogeneity. In order to down-scale, the heterogeneity of the larger scale input variables must be established and the influence of each variable on the large-scale model partitioned accordingly. Determining the influence of each variable and resolving the partitioning of results may require as much or more scientific understanding than would be needed for development from inception of the model with higher resolution. However, down-scaling can confer greater robustness to the model, and so both approaches are valid.

Data requirements are generally greater for a model developed to describe processes with high resolution than with low resolution. There is often insufficient data to run and evaluate a detailed model at a larger scale. Scaling-up demands either that high resolution data requirements are fulfilled or the model is adapted to reduce the data requirement (Smith, 1995a). Methods of estimating input variables from the available data must be developed if the detailed model is to be applied unchanged. The accuracy of simulations achieved by simplifying the model or interpolating data should be compared when deciding on the best strategy for changing scale.

It is clear that if a small scale model is used to simulate processes at a larger scale, the *efficiency* of the simulation must be maximized. Lord (pers.comm.) calculated that if a field-scale nitrogen model was applied in a national calculation on a field-by-field basis, the computer processor time required for the calculation would be prohibitively high. In addition, data entry would be extremely time-consuming. Methods must therefore be developed to reduce the number of unit calculations necessary and increase the efficiency of data entry.

In the following text, potential solutions to the problems associated with changes in scope, heterogeneity, data requirement and computing efficiency will be discussed in the context of ongoing work to develop a new system for UK nitrate policy support, required to function at scales ranging from field and farm through to catchment, regional and national level (Powlson, 1996).

2. Changes in Scope

2.1 Determining Changes in Scope

A simple, but comprehensive and systematic procedure can be used for determining all

changes in scope associated with changing scale (Fig.1). This procedure is centred around drawing up lists to describe the characteristics of the system. The first stage is to list all questions that are driving the change in scale. This allows the output values required to answer each question to be inferred. All processes included in the model are then listed, and the processes that provide each required value are marked. Any required output value that is provided by none of the processes represents a necessary increase in the scope of the model. The model is then adapted to incorporate the additional processes.

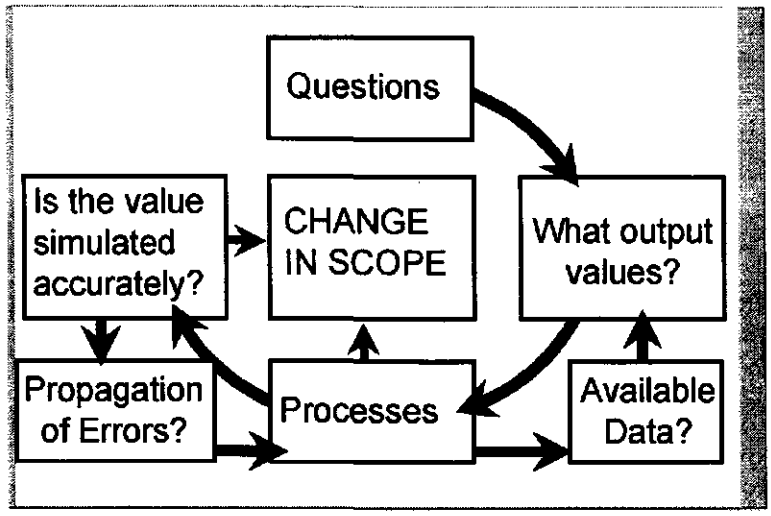


Figure 1 Systematic Determination of Changes in Scope

The data available to drive the simulation are listed and the input values required by each marked process itemized. Any inputs that are not included in the list of available data are added to the list output values that must be produced by some part of the model. Again, the processes that output each required value are marked, and the model is adapted to incorporate any further increases in scope necessary.

This procedure continues until all required input values are included in the list of available data. Any process that influences none of the required output values represents a decrease in the scope, and the model is adapted to omit the redundant process.

The next stage in determining the change in scope is to evaluate the accuracy with which each identified output value is simulated by the new model. Failure of the simulation may be due to propagation through the model of errors arising from the newly incorporated processes (Leenhardt, 1995). The error ($Err_{F(a)}$) arising from the newly incorporated process ($F(a)$) must be lower than the acceptable error (Err_{acc}) less the error ($Err_{G(F(a))}$) arising from the action of the rest of the model ($G(F(a))$) i.e.

$$G(F(a)+Err_{F(a)}) < G(F(a)) + Err_{acc} - Err_{G(F(a))} \quad (1)$$

Methods for determining the acceptable error are discussed by Smith et al., 1996b. If the standard errors of measurements are given, acceptable error may be defined statistically. Otherwise, the definition relies on arbitrary criteria.

The accuracy required in the simulation of each incorporated process to achieve an overall result within the acceptable error may be determined by a very simple sensitivity analysis in which the process module is substituted by an input variable, I.

The sensitivity analysis then allows the acceptable range of input values, $(I + Err_I)$, to be obtained from the acceptable error as before, i.e.

$$G(I+Err_I) < G(I) + Err_{acc} - Err_{G(I)} \quad (2)$$

A more comprehensive method of estimating the uncertainty contributions from several

independent multivariate sources is discussed by Jansen et al. (1994). The error associated with the incorporated process must be within the acceptable range determined by the sensitivity analysis, i.e.

$$Err_{F(a)} < Err_I \quad (3)$$

If it is established that the errors in the incorporated processes are within the acceptable error, the failure of the simulation must be due to errors in the scientific concepts of the processes involved, and the model should be modified.

Finally, the scope of the model is further refined by omitting any processes that the sensitivity analysis suggests are redundant at the larger scale.

2.2 Incorporating and Omitting Processes

The new system for UK nitrate policy support (Powlson, 1996) is designed to allow additional modules of processes to be easily incorporated as science develops. This requires that instead of describing the whole ecosystem using a complete model, as far as possible each process is characterised by an independent module. The ecosystem is then simulated by bringing appropriate modules together on a communicating framework.

Ensuring that each process is characterised by an independent module, requires individual evaluation of process modules. A complete model can often give a valid result because errors in the underlying processes compensate. This is especially likely to occur when a model has been constructed using optimisation procedures to obtain parameters that cannot easily be measured (Addiscott et al., 1996; Smith, 1995b). Effectively, a portion of the description of one process is subsumed into the description of another. If the process module is then omitted from the model, errors are introduced by the remaining portion. It may be difficult to obtain data to evaluate the individual processes, and so a model is often not comprehensively evaluated in its components. It is important, in this case, that the model should not be sub-divided any further than evaluation allows.

The communicating framework processes the questions entered into the system and calls the module that will provide the most accurate answer given the data and time available for the calculation. The framework will request the information demanded by the module from the most appropriate source: an attached database or another module. Mounting sub-divided models on such a framework, allows the strengths of one model to be borrowed by another. Work is ongoing to establish the precise nature of links between modules in the new system for UK nitrate policy support (e.g. Cropper et al., 1996).

2.3 Model Evaluation

Methods for evaluating the accuracy of a simulation and determining the acceptable error are discussed in detail elsewhere (Addiscott et al., 1996; Smith et al., 1996b). Each quantitative method described in this section provides information on a distinct aspect of the accuracy of the simulation. The method used to assess the goodness-of-fit between simulated and measured values depends on the type of measurements available (Fig.2).

In the following equations:
O_i are the observed (measured) values,
P_i are the predicted (simulated) values,

$\bar{O}$ is the mean of the observed (measured) data,

$\bar{P}$ is the mean of the predicted (simulated) data,

n is the number of paired values,

N is the number of experiments,

m_j is the number of replicates within each experiment,

P_j is the simulation for the jth experiment,

$\bar{O}_j$ is the mean of the measurements in the jth experiment,

$\bar{d}_j$ is the mean deviation = $\bar{O}_j - P_j$.

If the experiments have been replicated, the lack of fit statistic, *LOFIT* (Whitmore, 1991), should be calculated.

$$LOFIT = \sum_{j=1}^{N} m_j \bar{d}_j^2 = \sum_{j=1}^{N} m_j (\bar{O}_j - P_j)^2 \tag{4}$$

This statistic allows the experimental errors to be distinguished from the failure of the model. Where individual replicate measurements are available, the significance of *LOFIT* may be determined using an F-test, i.e.:

$$F = \frac{MSLOFIT}{MSE} = \frac{\sum_{j=1}^{N} (m_j - 1) \sum_{j=1}^{N} m_j (\bar{O}_j - P_j)^2}{N \sum_{j=1}^{n} \sum_{i=1}^{n} \left((O_{ij} - P_j) - (\bar{O}_j - P_j) \right)^2} \tag{5}$$

where *MSE* is the mean squared error. According to statistical convention, a value of F greater than the critical 5% F value may be taken to indicate that the total error in the simulated values is significantly greater than the error inherent in the measured values.

Where individual replicate values are not available, other tests must be used. After Loague & Green (1991), the total difference between the simulated and the measured values may be calculated as the root mean square error, *RMSE*, i.e.:

$$RMSE = \frac{100}{\bar{O}} \sqrt{\sum_{i=1}^{n} (P_i - O_i)^2 / n} \tag{6}$$

If standard errors of the measurements ($s_e(I)$) are available, the statistical significance of *RMSE* may be assessed by comparing to the value obtained assuming a deviation corresponding to the 95% confidence interval of the measurements (corrected from Smith et al., 1996b), i.e.

$$RMSE_{95\%} = \frac{100}{\bar{O}} \sqrt{\sum_{i=1}^{n} (t_{(n-2)95\%} \times S_e(i))^2 / n} \tag{7}$$

where $t_{(n-2)95\%}$ is the Student's-t distribution for n-2 degrees of freedom with a two-tailed P-value

of 0.05. An *RMSE* value less than $RMSE_{95\%}$ indicates that the simulated values fall within the 95% confidence interval of the measurements.

If no estimates of the standard errors are given, the accuracy of the simulation may be further assessed by calculating the modelling efficiency, *EF*, and the coefficient of determination, *CD* (Loague & Green, 1991). The modelling efficiency, *EF*, provides a comparison of the efficiency of the chosen model to the efficiency of describing the data as the mean of the observations, i.e.:

$$EF = \frac{(\sum_{i=1}^{n}(O_i-\bar{O})^2 - \sum_{i=1}^{n}(P_i-O_i)^2)}{\sum_{i=1}^{n}(O_i-\bar{O})^2} \tag{8}$$

Values for *EF* can be positive or negative. A positive value indicates that the simulated values describe the trend in the measured data as effectively as the mean of the observations. A negative value indicates that the simulated values described the data less well than a mean of the observations.

The coefficient of determination, *CD*, is a measure of the proportion of the total variance in the observed data that is explained by the predicted data.

$$CD = \sum_{i=1}^{n}(O_i-\bar{O})^2 \; / \; \sum_{i=1}^{n}(P_i-\bar{O})^2 \tag{9}$$

The lowest value of *CD* is 0. A value of 1 or above indicates that the deviation of the predictions from the mean of the measured values is less than that observed in the measurements, i.e. the model describes the measured data as well as the mean of the measurements. A *CD* value less than 1 indicates that the deviation of the predictions from the mean of the measured values is greater than that observed in the measurements, i.e. the mean of the measurements describes the data better than the model. Taken together, *EF* and *CD* allow *RMSE* to be further interpreted where standard error values of the measurements are unavailable.

The bias in the total difference between simulations and measurements may be determined by calculating the relative error, *E* (Addiscott & Whitmore, 1987).

$$E = \frac{100}{n}\sum_{i=1}^{n}(O_i-P_i)/O_i \tag{10}$$

If the standard errors of measurements are available, the significance of *E* may again be determined by comparing to the value obtained assuming a deviation corresponding to the 95% confidence interval of the measurements (corrected from Smith et al, 1996b):

$$E_{95\%} = \frac{100}{n}\sum_{i=1}^{n}(t_{(n-2)95\%} \times S_e(i))/O_i \tag{11}$$

An *E* value greater than $E_{95\%}$ indicates that the bias in the simulation is greater than the 95%

confidence interval of the measurements. The nature of the bias may be further examined using the mean difference, M (Addiscott & Whitmore, 1987):

$$M = \sum_{i=1}^{n} (O_i - P_i)/n \tag{12}$$

The mean difference between measured and simulated values gives an indication of the bias in the simulation but is less informative than E because errors are not proportioned to the size of measurement. However, it is a useful statistic when standard error values are not available to derive a value for $E_{95\%}$ since M can be related directly to t:

$$t = \frac{M}{s_d/\sqrt{n}} = \frac{\sum_{i=1}^{n} (O_i - P_i)/n}{s_d/\sqrt{n}} \tag{13}$$

where:

$$s_d^2 = \frac{\sum_{i=1}^{n} (d_i - M)^2}{n-1} = \frac{\sum_{i=1}^{n} ((O_i - P_i) - (\sum_{i=1}^{n} (O_i - P_i)/n))^2}{n-1} \tag{14}$$

d_i is the difference between measured and simulated values; and
M is the mean difference between measured and simulated values.

The t statistic (here after Chatfield, 1983) is used to show a significant difference between simulated and measured values. In line with statistical convention, a t value greater than the critical two-tailed 2.5% t value may be taken to indicate that the simulation shows a significant bias toward over- or under-estimation when compared to measured values.

To assess whether simulated values follow the same pattern as measured values, the sample correlation coefficient, r, may be calculated, i.e.:

$$r = \frac{\sum_{i=1}^{n} (O_i - \bar{O})(P_i - \bar{P})}{\left(\sum_{i=1}^{n} (O_i - \bar{O})^2\right)^{1/2} \left(\sum_{i=1}^{n} (P_i - \bar{P})^2\right)^{1/2}} \tag{15}$$

This statistic can be useful in assessing how well the shape of the simulation matches the shape of the measured data. However, if there is no clear trend in the measured data to give a spread of paired measured and simulated data values, the correlation coefficient is of only limited use in determining how well a model simulates the measured data.

The statistical procedures outlined in this section are summarized as a flow diagram in Figure 2.

3 Heterogeneity

3.1. Continuous Heterogeneity

Continuously heterogeneous variables are often most conveniently scaled-up by considering the integrated sub-processes to be characterised by the mean (Smith, 1995a). However, this assumption is only valid if the model response to the mean value of the inputs is the same as the mean of the model responses to all inputs individually, i.e.

$$\sum_N \frac{F(a_N)}{N} \equiv F\left(\frac{\sum_N a_N}{N}\right) \tag{16}$$

where
$F(\ )$ represents the model response;
a_N is the Nth input variable; and
N includes the whole population.

This is unlikely to be true, especially if the model is non-linear with respect to its parameters (Rao et al., 1977; Addiscott, 1993). Methods of obtaining a sample distribution of input values include the "bacon-slicer" (Addiscott & Wagenet, 1985), the Monte-Carlo (e.g. Warrick et al., 1977) and the Taylor series methods (e.g. Rao et al., 1977). In this context, the "bacon-slicer" method is used for normally distributed populations because it requires only the mean and the standard deviation as input values and maintains the minimum number of necessary calculations. The normally distributed population is divided into slices with equal areas, and the mid-point of each slice is taken as the model input. Running the model using this sample population, provides a sample distribution of output values, equivalent to the distribution of the whole population. The mean of model responses to all inputs is then approximately equivalent to the mean of the sample output values, i.e.

$$\sum_N \frac{F(a_N)}{N} \approx \sum_n \frac{F(a_n)}{n} \tag{17}$$

and

$$F\left(\frac{\sum_N a_N}{N}\right) \approx F\left(\frac{\sum_n a_n}{n}\right) \tag{18}$$

where n includes only the sample population.

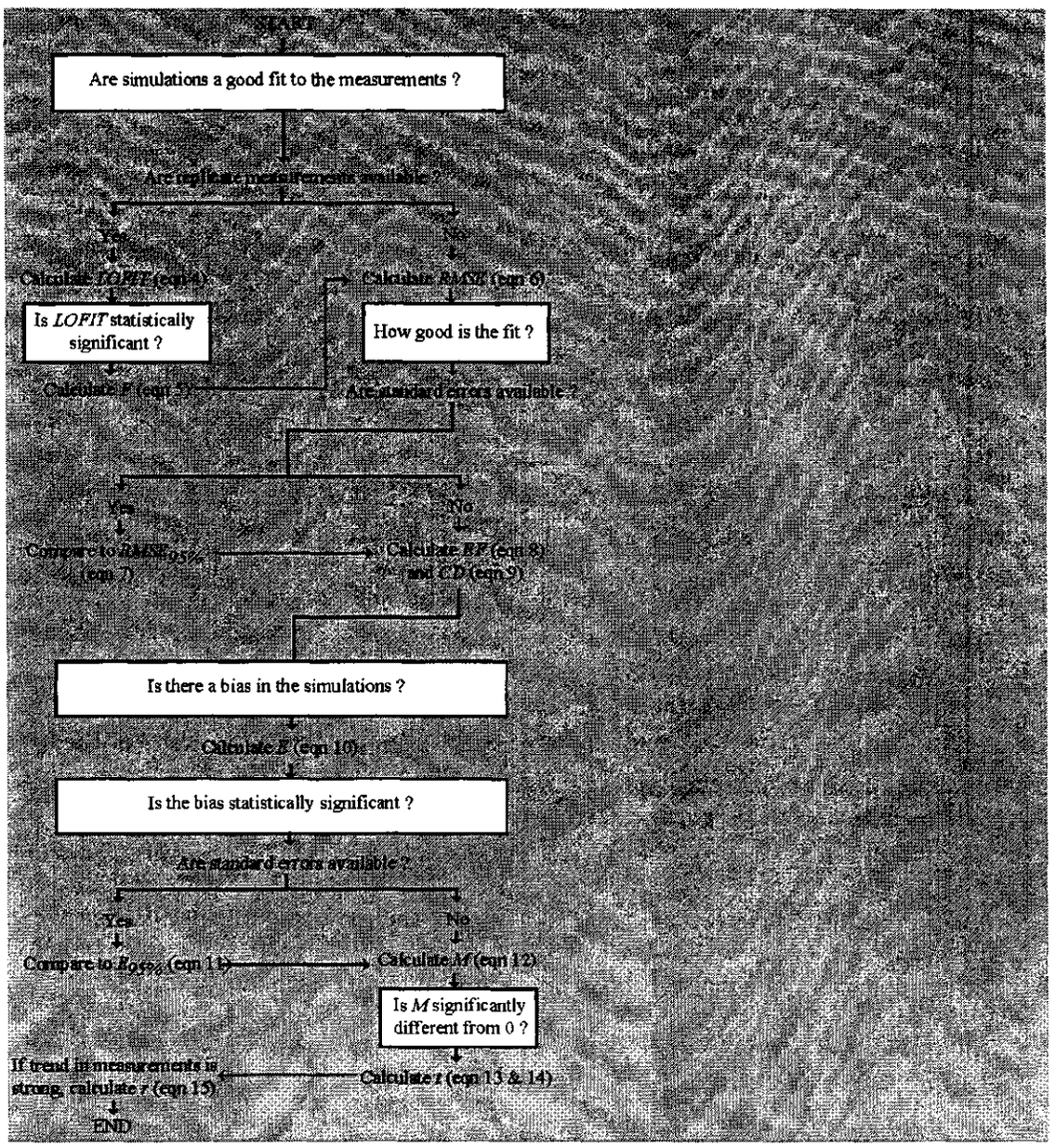


Figure 2 Statistical methods followed for the quantitative comparison of model predictions to measured values (reproduced from Smith et al., 1996c)

The assumption that the integrated sub-processes can be characterised by the mean is treated as valid only if the difference between the model response to the mean and the mean of model responses is less than the systematically defined acceptable error, Err_{acc} (Smith et al., 1996a). The range of input values within which the acceptable error will not be exceeded can be calculated from the mean, the standard deviation and the acceptable error, i.e.

$$\sum_n \frac{F(a_n)}{n} < F\left(\frac{\sum_n a_n}{n}\right) + |Err_{acc}| \tag{19}$$

This then allows the applicability of the mean input variable to be assessed, simply by inspecting the standard deviation associated with each input value, and comparing to the acceptable range calculated for each input value, held in look-up tables within the modelling system. The value of other inputs may, of course, affect the acceptable range, and a sensitivity analysis should be done to determine how the value of other inputs influence the acceptable range. If the observed range of inputs is outside the acceptable range, the input cannot be characterised by the mean, and the simulation must be run for the sample distribution of inputs.

This automated up-scaling procedure may be further extended by measuring the variance of the input values as the scale increases (e.g. Stein et al., 1989; Warrick et al., 1990). By obtaining the relationship between the variance and the size of sampling unit, the limits for reliable up-scaling may be determined. Addiscott & Tuck (1996) discuss the influence of possible shapes of such variograms on the up-scaled model. Relationships describing the variograms may then be incorporated into the modelling system to allow the system to select the most appropriate model for the required accuracy, scale and availability of data.

3.2. Discontinuous Heterogeneity

Discontinuously heterogeneous variables are most conveniently scaled-up by describing them as the majority value. For instance, crop type in a 1 x 1 km grid-square might be described as the majority crop type. This assumption is only valid if the model response to the majority value is statistically equivalent to the proportioned sum of the model responses to all possible inputs, i.e.:

$$\sum_n p_{a_n} F(a_n) < F(a_{max}) + |Err_{acc}| \tag{20}$$

where

p_{a_n} is the proportion of the population characterised as the discontinuous variable value a_n; and

a_{max} is the majority value.

By calculating the contribution of each discontinuous input value to the overall result, the acceptable proportion of each may be determined. Again, the value of other inputs will influence the acceptable proportion of the discontinuous inputs, and a sensitivity analysis should be performed to determine how continuous input variables affect the acceptable proportion. The influence of other discontinuous input variables on the acceptable proportion may be determined by running the model for all possible combinations. The probability of obtaining each calculated output can be assigned by multiplying together the proportions of each input, i.e.

$$\sum_m \sum_n p_{a_n} p_{b_m} F(a_n b_m) \quad < \quad \sum_m F(a_{\max} b_m) \quad + \left| Err_{acc} \right| \tag{21}$$

where

p_{b_m} is the proportion of the population characterised as a second discontinuous variable value b_m.

This calculation assumes that the discontinuous variables are independent of each other. However, discontinuous variables are often related to each other. For instance, good management practices result in a regional relationship between crop type and soil type, a variable that is also often treated discontinuously. Such relationships can be incorporated and used to reduce the number of calculations needed. A system to determine all possible scenarios using management practices and current legislation as inputs is currently under development (Smith & Glendining, 1996).

3.3 Partitioning the Influence of Heterogeneous Input variables

Because down-scaling increases the accuracy needed in the input variables and reduces heterogeneity, the heterogeneity of the larger scale input variables must be established and the influence of each variable on the large-scale model partitioned. Rodda et al. (1995) have down-scaled nitrate leaching calculated by the NCYCLE grassland model (Scholefield et al., 1991) from annual to monthly values. In this approach they establish a relationship to weather patterns for plant uptake of N, N mineralization and gaseous loss of N through denitrification according to a soil classification based on maximum soil moisture deficit. The fraction of the resulting leachable N left in the soil is related to the drainage volume and the peak and average nitrate concentrations are related to the total N leached according to soil type and drainage status. Finally, using the assumption that peak nitrate concentration is associated with the first 100 mm of drainage, the timing of nitrate peaks is estimated to the nearest month by a simple soil water balance.

Smith & Scholefield (1996) have used an alternative method in which unknown input values to a dynamic model are obtained from an annual model that is driven by more readily available input values. Annual crop nitrogen requirement, expected yield, and return of organic manure calculated by NCYCLE, are used as inputs to the dynamic arable model, SUNDIAL (Bradbury et al., 1993; Smith et al., 1996a). The standard SUNDIAL algorithms and fixed parameters specific to a grassland crop are then used to calculate the weekly turnover of nitrogen under the grassland crop. This effectively uses the SUNDIAL model to partition the robust NCYCLE values. The robustness of NCYCLE is conferred on SUNDIAL by down-scaling, whilst maintaining the transferability of the process-based SUNDIAL simulation.

4 Data Requirement

4.1 Adaptation of Models to Reduce Data Requirements

Many different methods are used to adapt a model to reduce the data requirements at the larger scale. The suitability of the different methods is dependent on the resources available and the precision of simulation required.

A very simple approach reduces the data requirements of a model by attaching a data-base

containing national, regional and sub-regional means (Cropper et al., 1996; Smith, 1995a). The data requirements of the SUNDIAL model have been effectively reduced by attaching a database containing the sowing date, harvest date, expected yield, fertilizer application rate and fertilizer application date. If no field specific input values are provided, the modelling system automatically selects the values from the sub-regional, regional, or national database, depending on the detail with which the geographical location is specified. The reliability of using data at each scale may be assessed against look-up tables of the acceptable data range as described in section 3.

If the look-up tables indicate that mean values will provide an unreliable simulation, the model can be run using the mean value and mean plus or minus the standard deviation of each input, or using the bacon-slicer method described earlier (Addiscott & Wagenet, 1985). These methods are more computationally expensive, but allow the error of the simulation to be quoted. The error increases as the scale increases, but the acceptable error also tends to increase with scale.

In an alternative method, the absence of direct measurements is overcome by calculating the input variable from some measured surrogate. Soil water properties are often estimated from easy-to-determine soil characteristics using pedotransfer functions (e.g. Salter et al., 1966; Gupta & Larson, 1979; Rawls et al., 1982; Pucket et al., 1985). Similar functions could be developed for estimating crop and organic manure properties (*phyto*transfer and *scato*transfer functions respectively?).

Anthony et al. (1996) have developed a simple minimum information requirement (MIR) model for application at the catchment scale. The MIR is a simplification of the results of more complex models (e.g. SLIM: Addiscott & Whitmore, 1991), attempting to retain the physical basis of simulations whilst reducing the dependence on input variables. The simplicity and comparative speed of the simulation should encourage a better understanding of the sensitivity of results to uncertainty in the inputs.

4.2 Estimation of Input Data

A wide range of methods have been used to interpolate the input data necessary to run models. Jacucci et al. (1993) developed a simple method for estimating soil characteristics from larger scale input values. Areal interpolation of land cover data for England and Wales, published annually at parish level, is described by Moxey (Moxey et al., 1995). In this method the variation in land characteristics across parishes is used to model variation in land cover, thereby allowing the spatial basis of the parish data to be transformed and improving the description of distribution of land cover. Weather data may be interpolated from regional meteorological offices by a method developed by Semenov & Barrow (1996), which uses spatial regression down-scaling together with a local stochastic weather generator.

5 Computational Logistics

If a modelling system is to be of any use to policy makers in decision support, the system should require a minimum amount of time for data loading and completion of simulations. The time required to run the system should be further reduced by attaching standard databases and using screen painting facilities to enter large bodies of data. The computational load represented by any model should be minimized by explicit design and efficient organisation of the operations performed by in the modelling system. Methods to maximize computation efficiency are currently being incorporated into the new system for nitrate policy support (Powlson, 1996).

6. Conclusions

In the above text, guidelines for choosing and changing the scale of a model have been described. When changing scale, the developer should itemise and demonstrate any costs and benefits, e.g. reduced accuracy, increased transferability, increased robustness. The scale of the model that is chosen should be driven by data availability and the questions asked. The general approach described above uses the following protocol:

On up-scaling:

1. Decide which input variables will be scaled up. This usually includes all parameters with one type of denominator. For instance, all variables changing with time or all variables changing over space.
2. Determine any changes in scope.
3. Use a systematic method to define the required accuracy of simulations.
4. Decide if each variable changes continuously or discontinuously.
5. Perform a sensitivity analysis to assess whether difference between the model response to the mean of continuous input variables and the mean of the model responses is within the required accuracy: if it is, the mean input variable can be used in the simulation; otherwise, the distribution of the input variables must be entered.
6. Deal with discontinuous heterogeneity.
7. Adapt the model wherever possible to reduce the data requirement.
8. Where not possible, develop methods of filling in the holes.
9. Organise the computational system to reduce the computational load and to assist data entry.

On down-scaling:

1. Decide which variables will be scaled down. This usually includes all parameters with one type of denominator. For instance, all variables changing with time or all variables changing over space.
2. Perform control analysis to determine changes in scope.
3. Use a systematic method to define the required accuracy of simulations.
4. Decide if each variable changes continuously or discontinuously.
5. Perform control analysis to decide which variables influence variable to be scaled-down.
6. Partition the variables either according to determined relationships or by linking to a model that functions at a lower scale.

In practice, however, data may be limiting and so less rigorous methods of changing scale will be adopted. In addition, most models are developed as complete packages that cannot easily be divided and validated as sub-processes. As a consequence, many modellers are committed to working with the model already developed, and use such methods merely as a means of justifying the application of the model at another scale. By building flexible systems, based on modules of sub-processes, into which new modules can easily be added, it is to be hoped that this understandable personal bias will eventually disappear. Improved solutions to problems will be more quickly achieved using new modules which slot into such modelling systems, rather than by developing yet another complete model of the whole eco-system.

7. References

Addiscott, T.M., 1993. Simulation modelling and soil behaviour. Geoderma, 60: 15-40.

Addiscott, T.M., Tuck, G., 1996. Sensitivity analysis for regional-scale solute transport

modelling. Application of GIS to the Modelling of Non-Point Source Pollutants in the Vadose Zone, ASA-CSSA-SSSA Bouyoucos Conference, Mission Inn, Riverside, CA. (In Press.)

Addiscott, T.M., Wagenet, R.J., 1985. A simple method for combining soil properties that show variability. Soil Science Society of America Journal, 49: 1365-1369.

Addiscott, T.M., Whitmore, A.P. (1987) Computer simulation of changes in soil mineral nitrogen and crop nitrogen during autumn, winter and spring. Journal of Agricultural Science, Cambridge 109: 141-157.

Addiscott, T.M., Whitmore, A.P., 1991. Simulation of solute leaching in soils of differing permeabilities. Soil Use and Management 7: 14-21.

Addiscott, T.M., Smith, J.U., Bradbury, N.J., 1996. Critical evaluation of models and their parameters. Journal of Environmental Quality, 24:803-807.

Anthony, S., Quinn, P., Lord, E., 1996. Catchment scale modelling of nitrate leaching. Aspects of Applied Biology, 46 (in Prep.)

Bradbury, N.J., Whitmore, A.P., Hart, P.B.S., Jenkinson, D.S., 1993. Modelling the fate of nitrogen in crop and soil in the years following application of ^{15}N-labelled fertilizer to winter wheat. Journal of Agricultural Science 121:363-379.

Chatfield, C. (1983) Statistics for Technology. Third Edition. Chapman & Hall, London.

Cropper, S.L., Smith, J.U., Scholefield, D., Brown, C.D., Hollis, J.M., 1996. An empirical approach to predicting nitrate leached to surface water from catchments of mixed land use. Journal of Agricultural Science, Cambridge. In Prep.

Gupta, S.C., Larson, W.E., 1979. Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density. Water Resource Research, 15: 1633-1635.

Howard, P.J.A., Loveland, P.J., Bradley, R.I., Dry, F.T., Howard, D.M., Howard, D.C., 1995. The carbon content of soil and its geographical distribution in Great Britain. Soil Use and Management, 11: 9-15.

Jansen, M.J.W., Rossing, W.A.H., Daamen, R.A., 1994. Monte-Carlo estimation of uncertainty contributions from several independent multivariate sources. In: Predictability and non-linear modelling in natural sciences and economics (Eds. J. Grasman & G. Van Straten), Kluwer, Dordrecht. pp.334-343.

Jacucci, G., Kabat, P., Pereira, L.S., Verrier, P., Bertanzon, G., Gianerini, G., Hoojier, A., Huygen, J., Simons, W., Steduto, P., Texeira, J., Toller, G., Tziallas, G., Uhrik, C., Vera-Montez, J., Youchev, P., 1993. HYDRA: A decision-support model for irrigation water management. Proceedings of the 2nd Workshop on Crop-Water Models, ICID 15th Congress on Irrigation and Drainage, The Hague, Netherlands, 17pp.

Leenhardt, D., 1995. Errors in the estimation of soil water properties and their propagation through a hydrological model. Soil Use and Management, 11:15-21.

Leffelaar, P.A., 1988. Dynamics of partial anaerobiosis, denitrification, and water in a soil aggregate: simulation. Soil Science, 146: 427-444.

Leffelaar, P.A., 1990. On scale problems in modelling: and example from soil ecology. In: Theoretical production ecology: reflections and prospects. (Eds. R. Rabbinge, J. Goudriaan, H. van Keulen, F.W.T. Penning de Vries, H.H. van Laar. Simulation Monographs, No. 34: pp.301. Pudoc Wageningen.

Loague, K., Green, R.E., 1991. Statistical and graphical methods for evaluating solute transport models: Overview and application. Journal of Contaminant Hydrology 7:51-73.

Miles, A.R., Turner, S.T.D., Lord, E.I., 1995. National agricultural pollution modelling in a GIS

environment. 7th Annual UK ESRI User Conference, Nottingham University.

Moxey, A., McClean, C., Allanson, P., 1995. Transforming the spatial basis of agricultural census cover data. Soil Use and Management, 11: 21-25.

Powlson, D.S. (Ed.) 1996. Application of predictive models for nitrate policy support. Annual Report to the Ministry of Agriculture Fisheries and Food, UK. Project Reference: NT1701. (In Prep.)

Puckett, W.E., Dane, J.H., Hajek, B.F., 1985. Physical and mineralogical data to determine soil hydraulic properties. Soil Science Society of America Journal, 49: 831-836.

Rao, P.S.C., Rao, P.V., Davidson, J.M., 1977. Estimation of the spatial variability of the soil-water flux. Soil Science Society of America Proceedings, 41: 1208-1209.

Rawls, W.J., Brakensiek, D.L., Saxtom, D.E., 1982. Estimation of soil water properties. Transactions of the American Society of Agricultural Engineers, 25: 1316-1320, 1328.

Rodda, H.J.E., Scholefield, D., Webb, B.W., Walling, D.E., 1995. Management model for predicting nitrate leaching from grassland catchments in the United Kingdom: 1. Model development. Hydrological Sciences, 40: 433-451.

Salter, P.J., Berry, G., Williams, J.B., 1966. The influence of texture on the moisture characteristics of soils. III. Quantitative relationships between particle size, composition and available water capacity. Journal of Soil Science 17: 93-98.

Scholefield, D., Lockyer, D.R., Whitehead, D.C., Tyson, K.C., 1991. A model to predict transformations and losses of nitrogen in UK pastures grazed by beef cattle. Plant and Soil, 132: 165-177.

Semenov, M.A., Barrow, E.B., 1996. Use of a stochastic weather generator in the development of climate change scenarios. In Prep.

Smith, J.U., 1995a. Nitrate Losses from Arable Land in Chichester Harbour: Initial Estimates using the SUNDIAL Model. In: Chichester Harbour Land Use and Nitrates Study. Report for National Rivers Authority (Sussex Area) and Chichester Harbour Conservancy. (M.J. Banham).

Smith, J.U., 1995b. Systematic determination model parameters. Application to field scale simulation of nitrogen dynamics in arable land. Proceedings of the British Society of Soil Science Annual Meeting, September, 11-14, 1995.

Smith, J.U., Glendining, M.J., 1996. Automatic design of arable rotations to minimize nitrate losses. AAB Conference Proceedings. (In Prep.).

Smith, J.U., Scholefield, D.S., 1996. Dynamic simulation of ley-arable rotations. (In Prep.)

Smith, J.U., Bradbury, N.J., Addiscott, T.M., 1996a. SUNDIAL: A PC-based Version of the Rothamsted Nitrogen Turnover Model. Agronomy Journal. (In Press.)

Smith, J.U., Smith, P., Addiscott, T.M., 1996b. Quantitative methods to evaluate and compare soil organic matter (SOM) models. In: Evaluation of Soil Organic Matter Models using Existing Long-Term Datasets (Eds. D.S. Powlson, P. Smith, J.U. Smith) NATO ASI Series I, 38:181-200.

Smith, P., Smith, J.U., Powlson, D.S., Arah, J.R.M., Chertov, O.G., Coleman, K., Franko, U., Frolking, S., Gunnewiek, H.K., Jenkinson, D.S., Jensen, L.S., Kelly, R.H., Li, C., Molina, J.A.E., Mueller, T., Parton, W.J., Thornley, J.H.M., Whitmore, A.P., 1996c. A Comparison of the Performance of Nine Soil Organic Matter Models Using Datasets from Seven Long-Term Experiments. In: Evaluation and Comparison of Soil Organic Matter Models Using Datasets from Seven Long-Term Experiments (Eds. P. Smith, D.S. Powlson, J.U. Smith, E.T. Elliot) Special Issue of Geoderma. (In Prep.)

Stein, A., Bouma, J., Mulders, M.A., Weterings, M.H.W., 1989. Using co-kriging in variability

studies to predict physical land qualities of a level river terrace. Soil Technology 2, 385-402.

Warrick, A.W., Mullen, G.J., Nielsen, D.M., 1977. Predictions of the soil water flux based upon field-measured soil-water properties. Soil Science Society of America Journal, 41: 14-19.

Warrick, A.W., Zhang, R., Moody, M.M., Myers, D.E., 1990. Kriging versus alternative interpolators: errors and sensitivity to model inputs. In: Field-scale solute and water transport through soil. (Eds. K. Roth et al.), Birkhäuser, Basel, Switzerland.

Webster, R., Oliver, M.A., 1990. Statistical methods in soil and land resource survey. Oxford University Press, New York. 316 pp.

Whitmore, A.P. (1991) A method for assessing the goodness of computer simulations of soil processes. Journal of Soil Science 42: 289-299.

Acknowledgements

Much of the work in this paper is based on the experiences of a project funded by the Ministry of Agriculture, Fisheries and Food, UK, to develop a system for nitrate policy support for use at the field, farm, catchment, regional and national level. D.S. Powlson (IACR, Rothamsted) is the project leader. T.M. Addiscott (IACR Rothamsted), N.A. Mirza (IACR Rothamsted) and D.E. Walling (Exeter University) are working on hydrological simulations. P. Verrier and B. Spencer (IACR, Rothamsted) are devising a user-friendly interface and modelling framework for the decision support system. M.A. Semenov (IACR Long Ashton) is developing a weather generator to interpolate climatic data. E.I. Lord, S. Anthony, P. Quinn and S. Turner (ADAS, Wolverhampton) are developing a land-use data system which combines census and satellite data systems and devising methods to scale-up and scale-down catchment agricultural, nitrate leaching and crop water use models. D. Scholefield, D. Lockyer and L. Cropper (IGER, North Wyke) are developing methods to scale-up and scale-down catchment and field grassland models. P. Loveland, C.D. Brown and J. Hollis (SSLRC) are devising interfaces between the LANDIS and HOST databases and the decision support system and developing catchment scale models utilising these databases. R. Fuller, P.J. Groom and A. Thompson (ITE, Monks Wood) are developing methods to use satellite images as inputs to the system. Finally, N.J. Bradbury, D.S. Jenkinson and P. Smith provide expertise in field scale nutrient cycling in arable systems. IACR receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

2.2. Spatial variability of data from uniformity trials as a function of scale

M.R. Hoosbeek, V. Epinat, A. Stein, T. J. Stomph, N. de Ridder and D. Hartkamp

Wageningen Agricultural University, C.T. de Wit School for Production Ecology, P.O. Box 37, 6700 AA Wageningen, The Netherlands.

1. Introduction

Input data to simulation models for use at a certain scale are usually based upon a certain support size. The support is defined as the size, shape, orientation and spatial arrangement of the spatial samples. For example, agronomic yield data used to validate crop growth or land use models are obtained from experimental trial plots. To test a hypothesis at field scale, a field is subdivided into plots to test, for example, differences between varieties or fertilizer applications. Based on the amount of variation expected within a plot, a minimum plot size is assumed to limit the variability necessary to test a hypothesis at field scale.

There is a relationship between support size of data and the distribution of their values (Isaaks & Srivastava, 1989). Averaging values over larger areas generally has the effect of reducing the variance of the data and of making their distribution more symmetric. The standard deviation, the coefficient of variation and the difference between the mean and the median all decrease as the support of the data increases. However, the mean is not affected by the support size.

In this study we will address the problem of spatial variability as a function of scale and associated support area. To do so, we have taken an example of a study focussing on the improvement of land use in central Côte d'Ivoire. A simulation model is being developed to test several land use scenarios (De Ridder et al., 1996). The model relates land use, e.g. crop cover and crop type, with catchment hydrology. The model simulates the effects of land use scenarios on water and nutrient budgets of a catchment area. To obtain information on the spatial variability of crop yield data, uniformity trial data from three different plot sizes, i.e. three different support sizes, were collected.

2. Methods

The experimental field is located at the research station of the West Africa Rice Development Association (WARDA), 15 km northwest of Bouaké in central Côte d'Ivoire (Hakkeling et al.,1989). The landscape is part of the interior plains of West Africa, which are slightly dissected peneplains with scattered inselbergs. Soils are underlain by a basement complex consisting of Precambrian granites and associated metamorphic rocks. The absence of tectonic activity allowed the soils to undergo several weathering cycles resulting in a mineralogy dominated by quartz and minor amounts of kaolinite and sesquioxides. Soils at the experimental site were classified as well drained Orthi-luvic Arenosols (Soil Survey Staff, 1992) with coarse

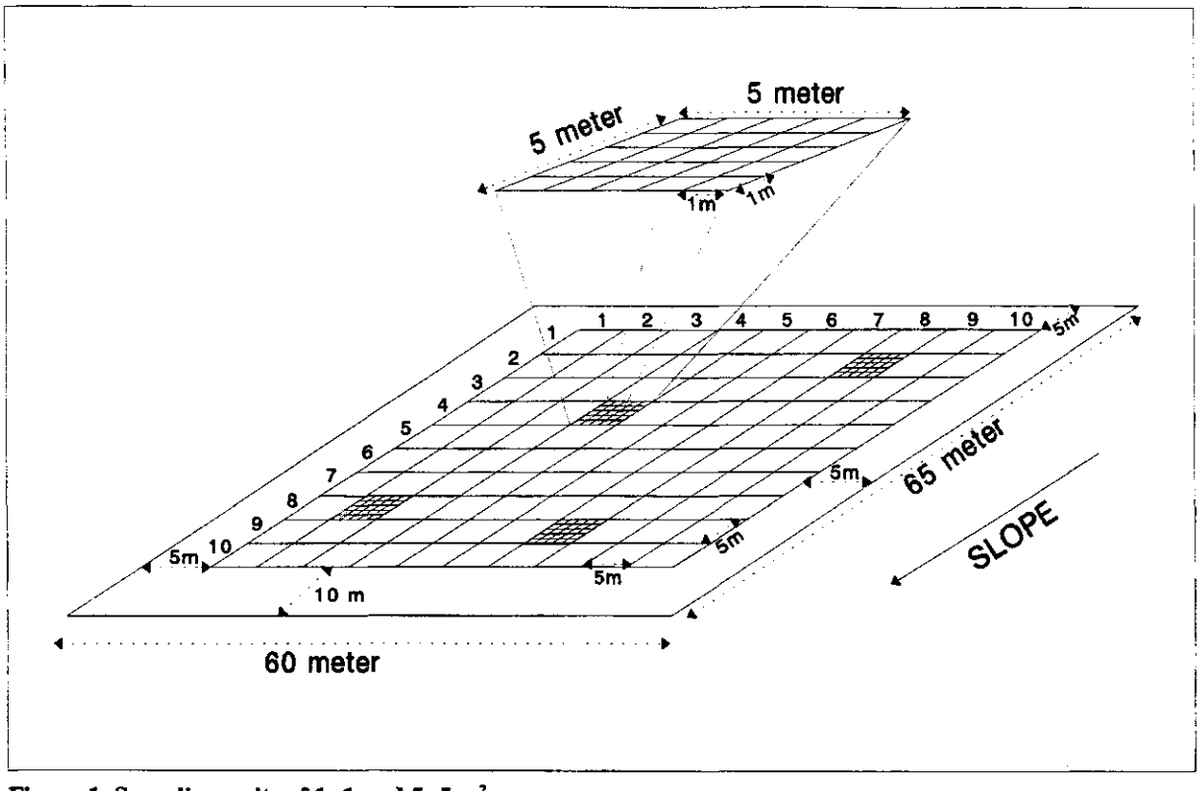


Figure 1. Sampling units of 1×1 and 5×5 m².

to medium sand in the upper part of the profile and a weak profile development.

The climate is characterized by a fairly stable average monthly temperature of 25°C with somewhat higher temperatures between January and April. Yearly precipitation and evapotranspiration are 1146 and 1478 mm. Most of the rainfall occurs during two periods, between April and June and between September and October. Maximum evapotranspiration occurs during the warmer period of January through May.

The 50×50 m^2 experimental field was subdivided into 100 plots of 5×5 m^2 plots. Four of these plots were subdivided, randomly assigned, into 25 1×1 m^2 subplots (Figure 1). A single variety of rice (WAB 56-50) was hand-sown in lines at a rate of 100 kg ha^{-1} between the 11th and 13th of June. Harvest took place between October 11th and 21st. The sequence of harvesting was based on differences in the development stage of the rice between plots.

The relationship between the support size of data and the distribution of their values, e.g. the yield, may be expressed by the dispersion variance. Isaaks & Srivastava (1989) defined the dispersion variance as an averaged squared difference that has the support of the individual values and the support of the mean explicitly stated:

$$\sigma^2(a, b) = \frac{1}{n} \sum_{i=1}^{n} (v_i - m_i)^2 \qquad (1)$$

where v_i are values with support size a and m_i is the mean calculated over the entire set of v_i's with support size b. In this study three different support sizes will be compared: $a_1 = 1$, $a_2 = 25$ and $a_3 = 100$ m^2. The third support size was obtained by averaging yields from 5×5 m^2 plots at 25 10×10

m^2 plots. Finally, the overall mean has a support size of $b = 2500$ m^2, i.e. the entire experimental field.

Yields obtained at plots close to each other are more likely to be similar than observations at a larger distance from each other. This spatial correlation is usually modeled with a variogram. The variogram is determined on the basis of the collected observations. For several distances h_1, h_2, ..., all pairs of points that have approximately such a separation distance are collected. The variogram for distance h_1 is estimated by taking half the average of the squares of the differences of all the pairs that belong to the distance h_1. For h_2 a similar procedure is applied. This gives variogram values for all the distances h_1, h_2, ... If these values are plotted as a function of the distance h, the variogram $\gamma(h)$ is obtained. It is estimated as

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Y(x_i) - Y(x_i + h) \right]^2 \qquad (2)$$

where $Y(x_i)$ and $Y(x_i+h)$ are a pair of observations with separation distance equal to h. The total number of such pairs is equal to $N(h)$. Half of the average value is determined, by dividing the sum of all the squared differences by twice the number of pairs.

A typical variogram may have a nugget variance indicating the variance at distance zero. This nugget variance represents the measurement error. With increasing lag distance the variance will increase up to the sill variance. The distance at which the sill is reached is called the range. Beyond this range the observations are not spatially correlated.

3. Results

The average yields for the three plot sizes are almost equal (table 1). However the standard deviation obtained on the 5×5 m^2 plots is about half of the standard deviation obtained on the 1×1 m^2 plots. Also on the 1×1 m^2 plots, the standard deviations are rather large compared to the mean yields. As compared to the overall mean, the dispersion variance is relatively high for the 1×1 m^2 plots, sharply decreases with an increase of support size to 5×5 m^2, and further decreases with an increases of support size to 10×10 m^2.

Variograms were constructed for the three support areas of respectively 1, 25 and 100 m^2 (Figure 2). The variance is plotted along the vertical axis. For all three variograms a power model was fitted (Deutsch & Journel, 1992):

$$\gamma(h) = c * h^s \qquad (3)$$

Table 1. Descriptives of rice yields.

Plot size (m^2)	Mean (g m^{-2})	Standard Deviation	n	Dispersion Variance
1 × 1	228.1	128.8	100	16586.7
5 × 5	226.6	57.8	98	3344.9
10 × 10	227.0	45.7	25	2091.9

Table 2. Variogram parameters (see figure 2).

Plot size (m^2)	Nugget variance	sill variance	range (m)
1×1	1670	10300	3
5×5	701	4341	34
10×10	17	3417	45

where $\gamma(h)$ is the variance at lag distance h, s is a power between 0 and 2 and c is a positive slope. The variogram parameters are presented in Table 2. The variances are relatively high for the smallest support area of 1 m^2. Increasing the support area from 1 to 25 m^2 steeply decreases the variance, followed by a moderate decrease in variance going from an area of 25 to 100 m^2.

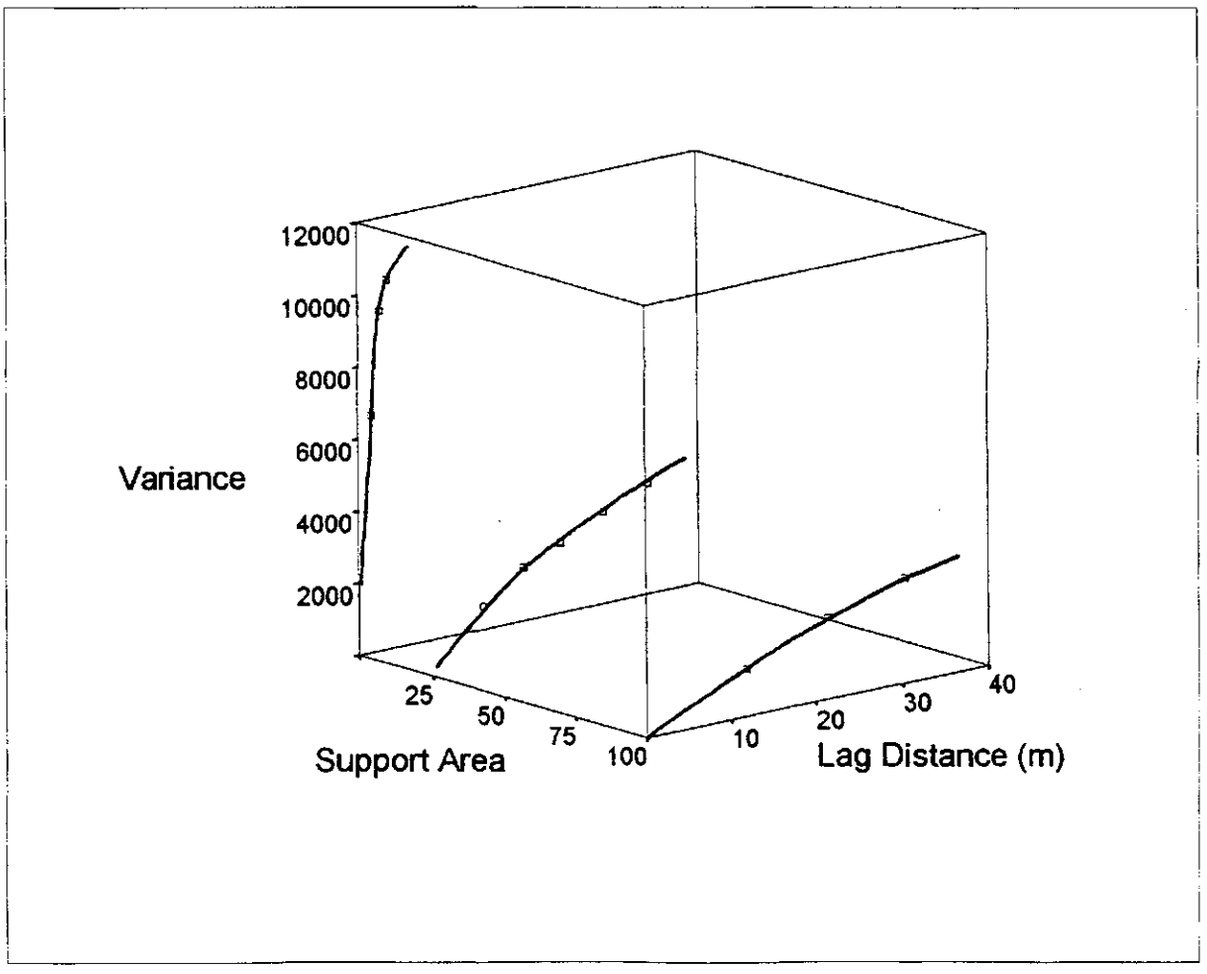


Figure 2. Variograms as a function of support area.

Fairfield Smith (1938) described the relation between the variance of yield data and the plot size with a power function. Both relations between variance and lag distance and variance and support area can be described by power functions. A surface describing variance as a function of lag distance and support size was modeled with the following equation:

$$\gamma(h, A) = \gamma(1, 1) * h^{0.35} * \frac{1}{\sqrt{A}} \qquad r^2 = 0.97 \tag{4}$$

with

$$\gamma(1, 1) = 6617 \tag{5}$$

where h is the lag distance (m), A is the support area, and $\gamma(1, 1)$ is the variance at a lag distance of 1 m and a support area of 1 m^2. The variance was found to be inversely related to the squared root of the support area. The power, $s = 0.35$, is a measure for the spatial variability primarily caused by soil heterogeneity.

4. Discussion

The obtained variance model, relating the variance as a function of lag distance and support area, has only two parameters that depend on local conditions. The variance $\gamma(1, 1)$ describes the crop and soil variability at a lag distance of 1 m and a support area of 1 m^2. The power, s, describes primarily the soil variability of the experimental field, i.e. soil variability between the 1 m^2 plots. Soil types with a relatively high degree of soil heterogeneity will yield a high s value. More homogeneous soil types, with low degrees of soil variability, will yield low s values.

More uniformity trial data, collected on different soil types, are needed to further test the variance model. If the relation between variance and support area proofs to be of a general nature ($\gamma(h) \sim A^{-1}$), then the model can be parameterized with data from 1×1 m plots only. The parameters $\gamma(1, 1)$ and s can then be calculated for each crop and soil type from 1×1 m^2 plot data. Having to use only, for instance, hundred 1×1 m^2 plots, instead of an equal number of 1×1, 5×5, and 10×10 m^2 plots, significantly reduces the cost of an experiment.

The land use models, mentioned in the introduction, will be applied at scales that are smaller (coarser) than the experimental plots. The here presented variance model facilitates an estimate of the variance at the scale at which the land use model will be applied.

5. References

De Ridder, N., Stomph, T.J., Fresco, L.O. 1996. Effects of land use changes on water and nitrogen flows at the scale of West African inland valleys: and explorative model. (submitted).

Deutsch, C.V., Journel, A.G. 1992. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York.

Fairfield Smith, H. 1938. An empirical law describing heterogeneity in the yields of agricultural crops. J. Agric. Sci. 26, 1.

Hakkeling, R.T.A., Smaling, E.M.A., Diatta, S. 1989. Detailed soil survey and land evaluation of the WARDA experimental area, Bouaké, Côte d'Ivoire. Winand Staring Centre, Report 22.

Isaaks, E.H., Srivastava, R.M. 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.

Soil Survey Staff, 1992. Keys to soil Taxonomy - 5th edition. SMSS technical monograph nr. 19. Pocahontas Press, Blacksburg.

2.3. Issues of scale

F.W.T. Penning de Vries

DLO Research Institute for Agrobiology and Soil Fertility, P.O. Box 14, 6700 AA Wageningen, The Netherlands.

1. Introduction

Linking and integrating research on growing crops is the specialty of the C.T. De Wit Graduate School Production Ecology (PE). Linking and integration occurs 'horizontally' in practical, complex problems that involve different disciplines, and 'vertically' at different scales in space and time. One of PE-trademark figures 'scales in time and space' reflects this (Fig. 1).

Integration of sciences is much in demand nowadays, because the combined knowledge from different disciplines is often required for solving practical problems. For instance, the newly defined research programs of the Netherlands Ministry of Agriculture, for the Agricultural Research Department, the 'DLO-programs', are much more directed towards 'useful' outputs than they used to be. One of the key words of its research is the so-called chain approach: consider all aspects of a product, i.e. from sowing to sale to consumers. Such developments indicate a strong demand for integration.

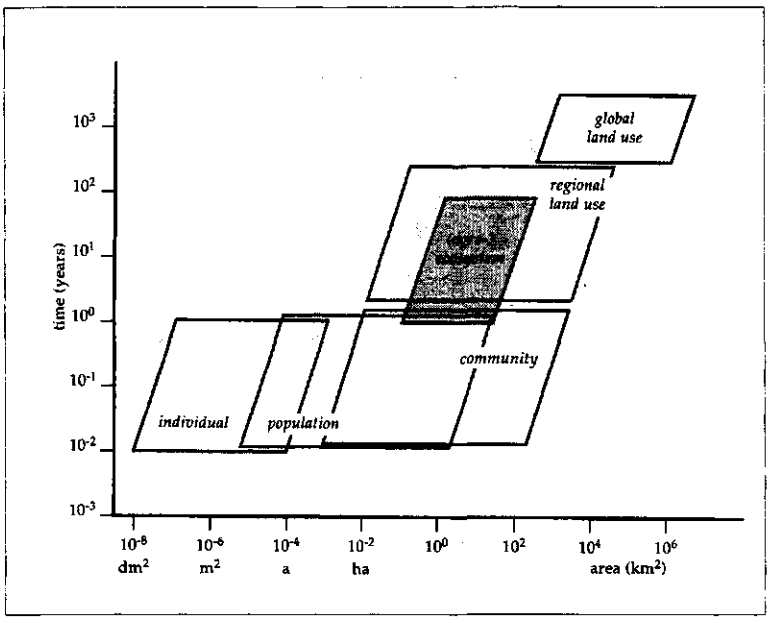


Figure 1. Scales in Space and Time (PE, 1995).

The development in quality and quantity of computers, personal computers in particular, and in a huge array of software has been astounding in the past decades. This provides better opportunities than ever for quantitative integration and linking of sciences.

The contents and shape of a simulation model should reflect the objective of the work in which it is used (De Wit, 1968; Penning de Vries & Rabbinge, 1995). Also the scale in time and

space at which a study is performed, determines the choice of the most appropriate model and the way in which processes are modelled.

The term 'scale' refers to three dimensions in crop models: time, space and complexity (i.e. involvement of more disciplines). Although these dimensions are independent, like the x, y and z axes of a 3D graph, complexity tends to increase with time and space, so that the scales of time and space increase often simultaneously. Modelling at different scales means that we are looking for results that refer to longer time spans, wider areas and to systems with more complex interactions. We will discuss changing of scales, or 'upscaling', for one of the common crop models using in PE: SUCROS (Spitters et al., 1989).

Upscaling by itself has no effect on a process. A first sight, therefore, there is no issue of scale. 'Scale' is reflected in the total system that is investigated. However, when the system is a field during a growing season, crop respiration is one of the many processes that play a role, whereas for a system consisting of a germinating seed, plant respiration is a major process (Table 1). Hence, 'scale' reflects on how we should approach the process, not on the process itself.

The term 'upscaling' can also be used to refer to making another model for an entire geographic region. Interactions at the physical and biological level are particularly common if our interest focuses on a watershed with crops instead of a single crops. Water containing nutrients and pesticides move downstream from fields higher up in watersheds; then pest epidemics can break out locally, and spread. Because of the physical scale, measurement at the regional level is difficult, and sometimes standardized methods not available. Also interactions in time occur: crops carry over water, nutrients, pests, to successive crops. Particularly in 'ecological agriculture' this is important.

Table 1. Glucose required during formation of organic components (1); plus related transport (2), expressed in g per g of the product formed (1+2). Values for N reduction (3) applying to leguminous crops includes the cost of transport (from Penning de Vries et al., 1989).

	Carbo-hydrates	Proteins	Fats	Lignins	Organic acids	Minerals
Biosynthesis (1)	1.211	1.793	3.030	2.119	0.906	0.000
Transport (2)	0.064	0.094	0.159	0.112	0.048	0.120
Growth (1+2)	1.275	1.887	3.189	2.231	0.954	0.120
reduction (3)		0.897				

2. SUCROS at different scales in time

SUCROS and SUCROS-type models are meant for simulation of growth and of C-balance processes of annual crops and these are used extensively for that purpose (Spitters et al., 1989, Stol et al., 1993; Matthews et al., 1995; Habekotté, 1996). The models are used to explore the functioning of different crops and their responses to the environment and environmental changes. Their time interval of integration is one day, basically in relation to the data of daily radiation; it has been used for shorter time intervals such as for quarter days, modelled by the L1Q-version (Penning de Vries et al., 1989) or even minutes, modelled by the PHOTON-version (De Wit 1978).

SUCROS-type models are also commonly used to explore stability of crop yield as a function of variable climate at one location. One example is agro-ecological zonation of wheat in India with yield probabilities at different locations (Aggarwal & Kalra, 1994). These results are informative in a semi arid environment where rainfall is highly variable and sowing date variable can be adjusted. Another example is APSRU's work in Australia to optimizing cropping systems for finding an ideotype (Meinke, 1996).

Both types of applications work fine, in principle. In practice, though, there are two problems:

- *with data.* For a fair analysis of average yield and yield variability, a series of 10-25 years is needed. In a highly variable environment, a longer series would be welcome. However, there are rarely available time series of weather data of sufficient lengths. In fact, series of reasonable quality of 3-10 years is often all there is. Obviously, without such data, the analysis cannot be performed. An elegant intermediate solution is called 'weather generation'. A specific program (e.g. Geng et al., 1986) can be used to generate weather data for any number of years with a statistical model for which the parameters are derived from real data. It needs a series of 3-5 years of data (and assuming absence of any long term trend). It is a very useful technique, when the limitations are carefully recognized.
- *with the model.* Crop models do a fair job in simulation and related growth processes under normal conditions. Extremes, however, are always poorly dealt with. For instance, the effects of a few days of scorching heat that leads to sterility in rice, of a night frost that kills blossom, or the dislodging effects of very strong winds are seldomly taken into account. Extremes are difficult to model because the physiological and morphological effects are not yet well quantified, and because the extent to which an extreme event occurs is difficult to predict or measure. When models are used for long time sequences, effects of extremes are generally neglected. We need more study of this phenomenon (Yin, 1996).

3. SUCROS at different scales in space

A very interesting application of crop modelling with upscaling in space is in 'yield prediction'. Governmental organizations need crop yield data as accurately and as early as possible, to arrange trade and prices and reserve funds for purchases, subsidies or special measures. Up till now, such data are often available only from official statistics. Not only are those data available only 1-2 years after the harvest (!), they are also inaccurate, because visual estimates are used and because different countries employ different methods. Private organizations interested in trade are keen to receive early warnings about possible bumper crops or failures of specific commodities, and to anticipate eventual large changes in quality of the produce.

In the recent MARS program (Bastiaans, 1993), a Crop Growth Monitoring System (CGMS) was produced, a prototype of which is now being used for Europe-wide yield predictions of several crops, such as wheat. The CGMS uses a crop model with generic species (or variety) specific coefficients, as well as a soil water balance model. The model is run to simulate water limited potential yields for each 50 x 50 km grid cell in the European Union. Real time weather data are collected and used to drive the CGMS. Grid cell data are aggregated to administrative units and countries. From June onwards, bi-weekly forecasts are given. The results help forecasters of crop yields of the European Bureau of Statistics to make decisions. A country specific calibration factor, reflecting sub-optimal management, is applied to reduce potential yield to the actual level.

Assuming that the important issues related to heterogeneity in field conditions are included in the model, there is still the important data issue: data should be available to reflect soil conditions, including spatial variability. They should be collected with a sufficiently dense network to allow reliable interpolation. Also information on the initial soil and crop situation is often needed and lacking. Quality of the predictions is obviously linked to availability of data. As an example we may consider treating of sugar beet fields all across the Flevopolder as being uniform. The CGMS-model simulates then a yield of 60 t ha^{-1} of sugar beets. Model performance increases significantly if for ten sites the dates of sowing are provided as well, (Table 2, from Bouman 1995). Results can be further improved when the state variable 'Leaf Area' in the model is updated during the season with remote sensing data (Table 1). This example shows how extra information can help. This will stimulate the use of IT technologies for improving forecasting.

However, for many areas and countries, in particular in the developing world, data will be scarce in the next decades. A research issue therefore is to develop methods to measure situation specific data, and to further develop remote sensing to upgrade information for crop models.

4. SUCROS at different levels of complexity

SUCROS simulates growth of crops under conditions where weather or soil water is limiting, but where growth limiting factors are absent and growth reducing factors are unimportant. For such circumstances, processes related to nutrients and pests are not required in the model. This situation is representative of conditions in some experimental fields. In farmers fields, however, nutrients may be in short supply and pests and diseases may be present. Then, the problem is more complex, and more processes must be included in the model (Bastiaans, 1994; Drenth et al., 1994). With the addition of more processes to the C-balance processes, the latter becomes relatively less important, and less detail in the model may in fact be advantageous.

Table 2. Actual and simulated tuber yield of sugar beet of 10 farmers in Flevoland, The Netherlands, 1991 (from Bouman, 1993).

Farm	Yreal	Ysim1	Ysim2	Ysim3	N
1	87	60	69.4	80.0	9
2	79	60	70.6	76.7	9
3	75	60	70.3	74.3	9
4	77	60	70.6	71.5	9
	81	60	74.6	----	----
5	70	60	63.5	70.5	8
6	68	60	63.3	65.3	11
7	70	60	65.6	69.9	12
8	61	60	58.2	68.1	11
9	70	60	63.6	72.2	15
10	77	60	72.8	75.7	12
	78	60	74.8	----	----
Average	74.3	60	68.1	72.2	11

Yreal = actual farmers yield (t ha^{-1})

Ysim1 = simulated farmers yield, using average sowing date of the region

Ysim2 = simulated farmers yield, using actual sowing date for each field

Ysim3 = simulated farmers yield, calibrated on optical remote sensing

N = number of remote sensing observations

Another issue is that of crop rotations. Replacing one crop by another while continuing the soil water and nutrient balances, to simulate carry over of water and nutrients, may seem basically a programming issue. However, in between crops there is soil management, there is carry over of soil diseases, seeds of weeds, and the latter become more important when less chemical inputs are used.

5. Data bases

Well defined data which characterize crops, soils and weather could be shared among many organizations and users of different models. IBSNAT (Uehara et al., 1993) was the first to operationalize this for soils data. Preparation of data sets for sharing implies standardization, and accessibility in data bases. Activities are almost completed to define standards for crop data (ICASA; Van Kraalingen & Hunt, *pers. comm.*). There is a great need for this. Of course, collection of data must be done in effective and efficient manner. Geostatistical procedures are useful here.

6. Validation

In any modelling study that aims at using a model at a different scale, validation of the model is essential. Yet, this rapidly becomes more difficult when models are applied for entire farms, for regions, or even for continents. It is already difficult for situations where a reference value exists, such as for yield prediction in individual countries, because the area on which it is based is not well defined. But it is even impossible for scenario studies focusing on climate change, or on the benefits of new crop types. We will have to be very thorough with various types of sensitivity analysis and checks of model quality. Calibration is necessary, but it also provides a way to control the output. When the subject is basically simple, such as a field crop, validation could be performed on many locations. When the system modelled is large such as a farm or a watershed, validation is not possible. Guidelines for validation have been identified (CAMASE, 1995).

Interactions caused by human interventions are common at the higher hierarchical levels. From a planning perspective, it requires elements of risk avoidance. For this purpose crop models and large sets of generated weather data have allowed quantification of probabilities of crop yields in specific environments (Roetter, 1993). Alternative sowing strategies and different crop types have been compared in this way in the search for optimum yields (Muchow & Carberry, 1993). When yields are highly variable due to changing weather conditions which occur regularly in semi-arid environments, these techniques seem to be most appropriate to be used. However, since farmers can respond by growing a second crop in case of a crop failure, the crop-level results should not be translated into farm-level results. Also, farmers in the Middle East who find that wheat grows unsatisfactory due to little rain, use it to feed their sheep (Seligman, *pers. comm.*).

Indeed, Bakker (1990) found in India that semi arid farmers did not accept insurance against low rainfall because of batter opportunities at the farm level to compensate.

7. Conclusions

At levels of farms and regions, scale matters a great deal when we try to apply our knowledge into practice. We have not yet looked into these very much. For PE, it would be good to give more attention to the six issues identified above: situation specific data (identify where), crop and soil data (contribute to building data bases), validation (develop techniques), physical and biological interactions (elaborate more examples) and human interactions (cooperate with socio-economists), and extremes (explore how to deal with them).

8. References

Aggarwal, P.K., Kalra, N. 1994. Analyzing the limitations set by climatic factors, genotype and water and nitrogen availability on the productivity of wheat. II Climatically potential yields and management strategies. Field Crops Research, 38, 93-103.

Bastiaans, 1993. Understanding yield reduction in rice due to leaf blast. PhD Thesis Agricultural University, Wageningen.

Bakker, E.J., 1993. Rainfall and Risk in India's agriculture. Wolters Noordhof, Groningen.

Bouman, B.A.M. 1994. Q-rier 1, 2. Newsletter Q-ray Agricmatica, Wageningen.

Bouman, B.A.M., 1995. Crop modelling and remote sensing for yield prediction. Neth. J. Agricultural Science, 93, 143-161.

CAMASE, Newsletter, issue 7b, 1995. AB-DLO, P.O. Box 14, 6700 AA, Wageningen.

De Wit, C.T., 1968. Theorie en Model. Inaugurele Rede, Landbouwhogeschool, Wageningen.

De Wit, C.T. 1978. Simulation of assimilation, respiration and transpiration of crops. Wiley, New York.

Drenth, H., Ten Berge, H.F.M., Riethoven, J.J.M. 1994. ORYZA simulation modules for potential and nitrogen limited rice production. SARP Research proceedings, AB-DLO and TPE-WAU, Wageningen and IRRI, Los Baños.

Geng, S., Penning de Vries, F.W.T., Supit, I.1986. A simple method for generating daily rainfall data. Agricultural and Forest Meteorology 36, 363-376.

Habekotté, B., 1996. Winter Oilseed rape, analysis of yield formation and crop type design for higher yield potential. PhD thesis, Agricultural University, Wageningen.

Matthews, R.B, Kropff, M.J., Bachelet, D., Van Laar, H.H. (eds), 1995. Modeling the impact of climate change on rice production in Asia. CABinternational, Wallingford, UK, and IRRI, Los Baños, Philippines.

Meinke, H., 1996. Optimizing Australian cropping systems with APSIM. PhD thesis, Agricultural University, Wageningen.

Muchow, R.C., Carberry, P.S. 1993. Designing improved plant types fro semiarid tropics: agronomists' viewpoints. Pp 37-62 in: 'Systems Approaches for Agricultural Development. F.W.T. Penning de Vries, Teng, P.S., Metselaar, K. (eds). Kluwer Academic Publishers, Dordrecht.

PE, 1995. C.T. de Wit Research School of Production Ecology, annual report 1994. Wageningen, The Netherlands.

Penning de Vries, F.W.T., Jansen, D.M., Ten Berge, H.F.M., Bakema, A.H. 1989. Simulation of Ecophysiological Processes of Growth of Several Annual Crops. Simulation Monograph Series 29, PUDOC, Wageningen, Netherlands and International Rice Research Institute, Los Banos, Philippines.

Penning de Vries, F.W.T., Rabbinge, R. 1995. Models in research and education, planning and practice. Pp 1-18 In: Ecology and modelling of potato crops under conditions limiting growth. Haverkort, A.J. & D.K.L. McKerron (eds).. Kluwer Academic Publishers, Dordrecht, Netherlands.

Roetter, R., 1993. Simulation of the biophysical limitation to maize production under rainfed conditions in Kenya: evaluation and application of the model WOFOST. PhD thesis University Trier.

Spitters, C.J.T., Van Keulen, H. Van Kraalingen, D.W.G. 1989. A simple and universal crop growth simulator: SUCROS87. Pp 147-181 In 'Simulation and systems management In crop protection. R. Rabbinge, S.A. Ward and H.H. van Laar (eds). Simulation Monograph, PUDOC, Wageningen.

Stol, W., Van Keulen, H., Van Kraalingen, D.W.G., 1993. The FORTRAN version of the Van Keulen-Seligman Spring wheat model. CABO-TT Simulation Reports 30. CAB, Wageningen.

Uehara, G., Tsuji, G.Y. 1993. The IBSNAT project. pp 505-514 In 'Systems Approaches for Agricultural Development. F.W.T. Penning de Vries, P.S. Teng and K. Metselaar, eds. Kluwer Academic Publishers, Dordrecht.

Yin, 1996. Quantifying the effects of temperature and photoperiod on phenological development of flowering in rice. PhD thesis Wageningen Agricultural University.

3. Models and goal-oriented information systems

Models are often applied in a more general setting, such as Geographical Information Systems, Knowledge based systems and Decision Support Systems. In this chapter attention is paid towards information science and its contribution to modelling. As in the previous chapter, the use of models will be the central focus point.

3.1 Environmental Modelling with Geographic Information Systems.

P.A. Burrough

Netherlands Institute for Geoecology, Faculty of Geographical Sciences, Utrecht University, P.O. Box 80.115, 3508 TC Utrecht, The Netherlands.

GIS-based modelling of natural environmental phenomena involves linking three different kinds of activities, namely *model building* (to describe the physical and chemical processes), *data collection* (to provide values for model parameters and attributes, and *GIS* (to provide an organized data structure for the automated handling of the input data to the computations and for the display of results). *The development of mathematical model building* for spatial and temporal data extends from rule-based logic through empirical (regression) models to deterministic and stochastic numerical descriptions of physical and chemical processes: spatial and temporal extents have grown from one-dimensional to four-dimensional, and increasing attention is paid to hierarchies of space and time. These conceptual developments are juxtaposed with a) *the collection of data* and b) *the design of current GIS*. The basic assumptions and thought structures in present-day GIS, including objects and continuous fields, overlay structures and relational databases and their suitability for environmental modelling are discussed. The basic principles, methods, advantages and problems of linking models and GIS are presented and evaluated. Special attention is paid to questions of uncertainty, spatial and temporal variation and upscaling, model calibration, validation and error propagation.

1. Introduction: environmental modelling with GIS.

Geographic information systems (GIS) are increasingly being used for inventory, analysis, understanding, modelling and management of the natural environment (Bouma & Bregt 1989; Burrough 1986; Goodchild et al 1993, 1996; Maidment 1993, 1995). Environmental modelling has at least two distinct aims (Moore et al 1993):

a) to help understand the physical world
b) to provide a predictive tool for management.

Scientists use models to understand the natural world better and their models may be complex. Simple models may enable managers to make useful predictions but may be less detailed than scientists would like. In recent years many environmental agencies have seen GIS as a way to link numerical models to spatial databases to provide both understanding and prediction in the form of attractive, easy-to-read graphs, maps, and multi-media demonstrations (Figure 1).

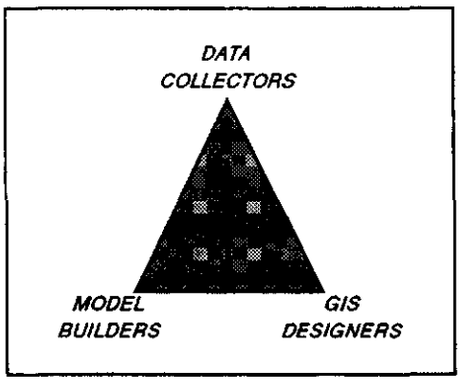


Figure 1. Data collectors, model builders and GIS designers often have different views of the world.

However, many people do not realize that modelling, data collection and GIS are separate activities, each with its own conventions, procedures and limitations so that linking them together at a merely technical level guarantees neither understanding nor useful prediction. Because *modellers*, *data collectors*, and *GIS designers* have very different training, conceptual views of the world, jargon and approaches to their separate disciplines, there is no *a priori* reason why the three should be mutually compatible. It is likely, therefore, that there will be mis-matches between the sub-parts of any combined GIS-data-modelling system, which of course, must affect the quality, costs and benefits of the modelling results. The purpose of this paper is to review briefly the various aspects of modelling, data collection and GIS which may affect the success of environmental modelling with spatial information systems.

2. Modelling as a scientific activity.

Irrespective of the application domain (e.g. climatology, ground water quality, crop yield forecasting, etc.) most environmental modelling is based on the assumption that any given process can be expressed in a formal mathematical statement or set of statements. Models are approximations of how the world works. The simpler the process, the easier it is to formulate an algorithmic compression in mathematical terms.

> *Without the development of algorithmic compressions of data all science would be replaced by mindless stamp collecting - the indiscriminant accumulation of every available fact. Science is predicated upon the belief that the Universe is algorithmically compressible... a belief that there is an abbreviated representation of the logic behind the Universe's properties that can be written down in finite form by human beings.*
> J.D. Barrow. (Theories of Everything, Vintage, 1991, p.11)

The advantages of modelling include the ability to provide a conceptual framework for understanding how a given process operates, for making quantitative and qualitative predictions, for summarizing knowledge in a succinct form, for guiding experimentation and research and for presenting ideas about complex phenomena in straightforward ways.

2.1 Model use and model development.

Models should be *parsimonious* (not more complex than necessary), *modest* (not too ambitious), *accurate* (unbiassed) and *testable* - Moore et al (1993). Creating a computer model involves the following steps (Jakeman et al 1991):

a) definition of aims
b) specification of the system of interest, the data required and any prior information
c) selection of the type of model - rule-based, empirical, or physical
d) identification of model structure

e) algorithmic implementation (programming)
f) estimation of model parameters and calibration
g) verification and checking
h) sensitivity analysis and error propagation
I) validation using other data sets from the same area or in other areas.

2.2 Types of model

I distinguish four classes of algorithmic compression or "model". These are: *rule based* (logical models); *empirical or black box* (regression models); *physical-deterministic or white box* (process-based - in principle everything about the process is known); *physical-stochastic* (the process is only approximated by the model but probabilities are known).

Rule-based models are based on the basic axioms of logic and straightforward set theory operations on discrete binary, ternary or similar data. For example, the hazards of diffuse pollution of ground water by agricultural fertilizers could be modelled by the following rule:

```
IF FERTILIZER SURPLUS ≥ 20%
     AND SOIL TEXTURE = SAND
          AND GROUND WATER LEVEL ≤ 50CM
               THEN NITRATE POLLUTION HAZARD IS SEVERE (1)
```

New developments in the field of Fuzzy Logic are supplementing the discrete, crisp logical models by an idea of *continuous membership*, with values for class membership ranging continuously between 0 and 1 (Klir & Folger 1988, Burrough 1989, Burrough & Frank 1996), permitting statements of uncertainty and inexactness not possible in the usual crisp logical model (Heuvelink & Burrough 1993).

Empirical regression models are data driven: they are often called *response function models* or *transfer models* (Bouma & Bregt 1989) and may be based on multivariate regression. The models do not guarantee causality (many spurious correlations have been observed), but through long and widespread use they may achieve a wide acceptance, being used outside the areas for which they were initially developed (e.g. the Universal Soil Loss Equation (Burrough, 1986)). They have the following general form:

$$U = \beta_0 + \beta_1 A^p + \beta_2 B^p + \beta_3 C^p + ... + \epsilon \qquad (2)$$

where U is the dependent variable to be estimated, the β_i's are regression coefficients, the A, B, C, .. are independent data raised to power p and ϵ is a normally distributed error term with zero mean and variance σ^2. Cross terms in AB, ABC, etc. may also be included. The "goodness of fit" or coefficient of multiple correlation squared (R^2) measures of how well the model fits the data.

Deterministic physical models attempt to "explain" the process or phenomenon in terms of basic physical and chemical "laws". For example, the movement of water through a porous medium is described by the physics of a driving force acting in a given direction through a medium with a

given conductivity. In an idealized medium the physics of fluid transfer in both saturated and unsaturated flow can be given by Darcy's law and the continuity equation.

The major assumptions of deterministic models are I) that the main physical driving forces behind the process are known completely and ii) the system can be regarded as essentially closed within given boundary conditions so all that is required is to collect the appropriate data. Uncertainties in the output of a deterministic model are considered to be a result of uncertainties in the *model parameters* or in the *data*. For a given data set or problem, uncertainties are reduced as much as possible by calibration or inverse modelling, that is by using some objective criterion to optimize the parameter values for the area in question. Where possible models can be validated by comparing model predictions with field measurements but models that forecast future events or situations can only be validated by using past data sets, or by waiting patiently.

Some scientists believe that the uncertainties in data mean that it is futile to continue to expand the sets of parameters used in a deterministic model:

> *Unfortunately very few earth science processes are understood well enough to permit the application of deterministic models. Though we know the physics and chemistry of many fundamental processes, the variables of interest ... are the end result of a vast number of processes... which we cannot describe quantitatively.*
> Isaaks & Srivastava 1989.

while others doubt that model validation with independent data has any use whatsoever:

> *Verification and validation of numerical models of natural systems is impossible ... because [they] are never closed and because model results are always non-unique. The primary value of models is heuristic.*
> Oreskes et al 1994.

The logical consequence of this point of view is to replace deterministic models and "exact data" by a probabilistic approach and to examine ways of understanding model sensitivity and the propagation of errors (cf. Heuvelink 1993).

Stochastic physical models also describe a natural process in terms of physical or chemical driving forces, but at least one of the model parameters or input variables is described by a *probability distribution* instead of a single number. The result of the stochastic model is not a single number but also a probability distribution. In many cases multivariate, probabilistic approaches are used (e.g Gomez-Hernandez & Journel, 1992). A properly calibrated deterministic model should give a result that is equivalent to the mean value of the output of the equivalent stochastic model, i.e. the deterministic model should be unbiased.

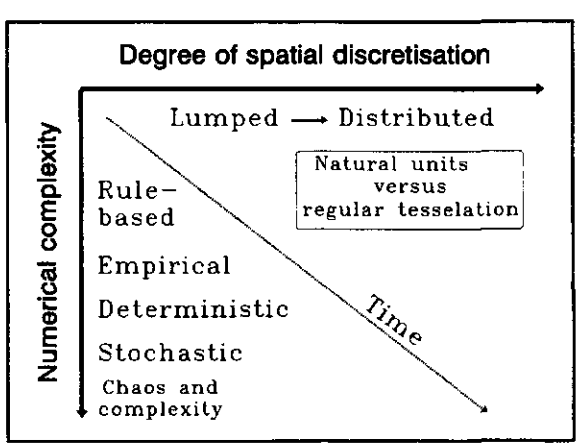


Figure 2. Trends in space-time modelling reflect computer power and data availability.

2.3. Discretization of space and time in models.

Most numerical models use some form of discretization of the space-time continuum because parameter values and variables are not known everywhere. *Finite element models (FEM)* use spatial or temporal units that are assumed to be internally homogeneous with respect to the parameters and variables of interest: they may be irregular or regular in form and can be derived using external information on landform, soil type or lithology. *Finite difference models (FDM)* divide the space-time continuum into sets of regular tiles which are a discretization of a continuous field -no prior information about the shape and size of "natural units" is used. FDM are being increasingly used in modelling because of their ease of handling in computers. Figure 2 shows how increasing computing power and data availability has led to more complex numerical approaches and to enhanced space/time resolution.

3. The collection of data

In most environmental sciences there are two main strategies for collecting data: systematic inventory and ad hoc, project-based data collection. Systematic surveys (including satellite remote sensing) are usually made by a national or regional agency according to accepted guidelines, standards and levels of spatial and temporal resolution or aggregation to support widely accepted, broadly defined uses. The data collected are often classified and made available as reports, maps and electronic databases, but there is rarely much information about intra-unit spatial variability or the uncertainty to be associated with attributes. Ad hoc surveys may cover similar attributes but are usually made for single purpose surveys or special studies: the spatial and temporal resolution may be specific for certain studies only, the data may not be generally available and may have little value outside the context of the study for which they were collected.

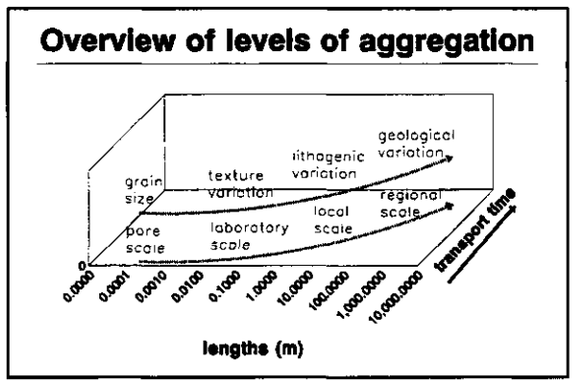


Figure 3. Schematic diagram of the variation of transport time for ground water as a function of spatial resolution.

Independent surveys (particularly in soil science - Beckett & Webster 1971, Burrough 1993a) demonstrate that environmental data are frequently more variable in space than most users would like; the same may be true of temporal data. Also, the numbers recorded must be linked to the levels of resolution and scales of aggregation, otherwise they will be used out of context (Figure 3).

4. The representation of spatial and temporal data in GIS

Current GIS use two main geographic data models (ie a formalization of geographic entities) for representing spatial phenomena: the entity model and the continuous field model. Ideal geographic entities have crisply defined spatial boundaries and a well defined set of attributes, such as a land parcel with accurately surveyed boundaries and attributes of area, ownership, land use, tax value, and so on, apply uniformly to the whole entity. However, other geographical phenomena are more often thought of as continuous fields -eg air pressure, elevation as represented by the hypsometric surface, hydraulic heads or pollution plumes. These are usually represented by smooth mathematical surfaces (often polynomial functions) that vary continuously and smoothly over space-time.

In the simple entity model of geographical phenomena, natural entities are represented by crisply delineated "points", "lines", "areas" (and in three dimensions) "volumes" in a defined and absolute reference system. Lines link a series of exactly known coordinates (points), areas are bounded by exactly defined lines (which are called "boundaries") and volumes are bounded by smooth surfaces. Lines are linked by a defined topology to form networks which, if open, can represent rivers or blood veins, or if closed, the abstract or defined boundaries of polygons that in turn represent land parcels, soil units or administrative areas. The properties of the space at the points, along the lines or within the polygons or volumes are described by attributes, whose value is assumed to be constant over the total extent of the entity (Burrough 1986). This is the *choropleth* (areas of equal value) model. Cartographic convention has reinforced these abstractions by insisting that mapped boundaries should only be represented by lines of a given style and thickness. Recent developments in computer programming have led to the development of *object oriented* systems in which complex database "objects" can be built from sets of interacting, linked point, line and area entities. The inclusion of inherent topology in the database improves the ability to model the transport of materials, as in hydrology.

In the simplest form of the continuous field model there are no boundaries. Instead, each attribute is assumed to vary continuously and smoothly over space: its variation can be described efficiently by a smooth mathematical function and it can be visualized by lines of equal value (*isopleths* - contours). In practice, these fields are often discretized to a regular grid at a given level of resolution, though variable density grids in the form of *quadtrees* can be used (Burrough 1986).

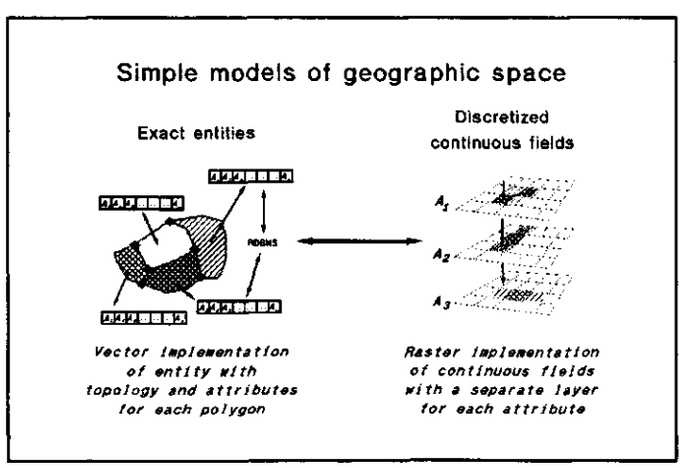


Figure 4 Exact entities or discretized continuous fields as models of geographical space.

The conventional entity model, object-orientation and the continuous field model are all abstractions of reality attractive for their logical consistency and their ease of handling using conventional reasoning and mathematics: all have been implemented in geographic information systems. The entity model in its simplest form has been implemented using a relational database structure for the attributes and a network data structure for the topology: the continuous field can be handled in a purely relational database

structure once it has been discretized to a regular grid, or alternatively approximated by a network structure such as a TIN (triangular irregular network). Both entity and field models are so accepted as fundamental aspects of geographical information systems that few persons question their general validity.

These spatial data models have been implemented using the graphic models of *vector* and *raster* structures. The vector structure enhances even further the abstraction to exact entities because in the computer digitized lines are by definition infinitely thin (Figure 4, left), while the raster structure introduces approximations in shapes and form because of discretization on a regular grid (Figure 4, right). Directed pointers between entities can easily be handled in vector systems but network interactions can only be handled in raster systems by considering cell-to-cell neighbour interactions.

5. Linking data and models.

In principle, each spatial unit can be treated as a local system "object" in which the state of the system is determined by its attributes and transfer of material can occur laterally from neighbour to neighbour or over defined topological links. This is relatively simple to achieve in raster systems because data on regular grids can be used to create a wide variety of derived attributes such as within neighbourhood indices, buffer zones with and without friction, first and second order derivatives (slope, aspect, profile curvature, viewsheds, insolation) and topological linkages (Figure 5). Temporal processes can be modelled by linking state attributes and transfer operations to time series (van Deursen 1995, Wesseling et al,(1996).

Although the "system object" approach is providing interesting new results data do not always come at the level of aggregation that the modeller wants, nor are they always stored in the most appropriate data model. We can distinguish five levels of aggregation:

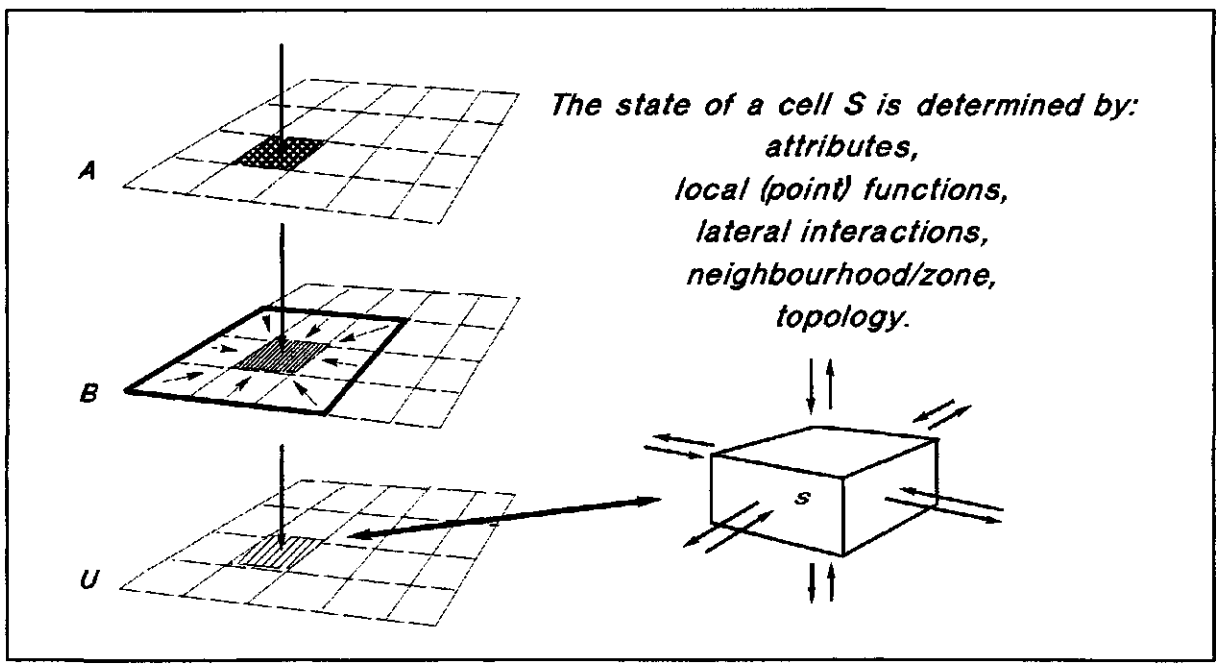


Figure 5 Dynamic raster models treat each cell as an object in a set of linked systems.

I) the measurement scale (the support). This can vary from a few milligrams of soil or a few square millimetres in area to satellite pixels of up to 5 x 5km.

ii) the level of discretization in the GIS - for example the grid cell size to which data are interpolated from measurements at points.

iii) the modelling scale, namely the level of spatial and temporal aggregation that is built in to the numerical model. A model that is designed to operate on long time periods (monthly averages, for example) will not be able to handle continuously recorded data without serious modifications. The shortest critical spatial or temporal scale in the model may strongly affect the way the model is built and optimized and the kinds of data it requires.

iv) the "natural" scale of phenomena in space and time that data collectors like to associate with "geographic objects" or "spatial entities". Though proponents of fractals would have us believe that landscapes are intrinsically statistically self-similar, fractal devotees are beginning to realize that the field scientist's concepts of characteristic geological, geomorphologic and hydrologic structures also have more than a grain of truth in them (Burrough 1993b, Lavallée et al 1993, Rodriguez-Iturbe et al 1994). A combination of determinism and chance in working with "natural structures" seems most appropriate.

v) the scale of the application. For the scientist, this may be the same as in iv); for the manager, it could imply a piece of land of any size or volume from a wagon load of polluted soil to a large river catchment. The application scale need not match any of the others.

The effects of data resolution on the quality and propagation of errors to the results of numerical models must not be forgotten when data are in a GIS. If the spatial discretization in the GIS matches the basic conceptual units of the numerical model then a simple 1:1 approach may be valid, which is a trivial operation in GIS. At both model and regional scales, however, there are usually the problems that the model blocks are larger than observations; there are more model blocks than observations; the size of the model blocks is much larger than the size of the observations; it is too expensive to measure all the attributes and parameters needed directly so these values must be derived from other data. Model calibration (sometimes known as *inverse modelling*) can suffer from problems of non-identifiability, non-uniqueness and instability.

If the measurements are considered to be truly representative of "natural units" which are stored in a GIS, then means, modes and standard deviations can be computed and assigned to model blocks which are smaller than the natural units. Alternatively, geostatistical methods can be used to interpolate or simulate values for model units within the natural units. Computing simple statistics is easy in most GIS but advanced geostatistical methods such as conditional simulation (Deutsch & Journel 1992) are usually more demanding and require special software. If the measurements are qualitative then methods like sequential indicator simulation may be used to convert field measurements of qualitative data into quantitative estimates at the model scale (Bierkens 1994, Bierkens & Burrough 1993ab).

6. Technical aspects of linking models and GIS

There are three ways to link a model with the GIS. In *loose coupling* the GIS (and any geostatistical software) is used to retrieve and pre-process the spatial data into the form required by the numerical model. The data are written to files which are then used as input to the numerical model, which may reside on another computer. The model computes the results and returns them as files of point data or areal data which are then displayed (or interpolated and displayed) by the GIS.

Loose coupling is appropriate when a standard numerical model is being linked to GIS as an experiment or as part of an exploratory process, or when there are particular computational requirements such as parallel processing that are not provided by the GIS hardware platform. Loose coupling may involve considerable work in changing data formats and data structures, particularly if the model has been obtained from another source.

Tight coupling also involves export of data from GIS to the numerical model with the return of results for display, but model configuration is done directly using the interactive tools of the GIS (setting up parameter values via menus, etc) and the exchange of data is fully automatic. This is appropriate if a given model is used as a standard for a large number of different applications (e.g. the incorporation of MODFLOW in the Intergraph MGE system or the integration of ILWIS and MICRO-FEM (Biesheuvel & Hemke 1993). Tight coupling requires considerable investment in programming and data management.

Embedded coupling is either a) the numerical model is written using the analytical engine of the GIS, or b) a simple GIS added to a complex modelling system to display results and provide interactive control (Fedra 1993). Embedded coupling implies a generic mathematical modelling language linked to a single, integrated database on a single hardware platform.

Current standard GIS analytical capabilities permit the user to carry out many kinds of logical and mathematical modelling such as variability within windows of a given size, slope and aspect, buffer zones, or fastest path over potential surfaces, which can be of great value for assessing the impacts of roads, computing shortest paths for water movement or pollutant transfer. However, most commercial GIS currently do not provide an embedded mathematical programming language; an example of a prototype system using regular grids is PC-RASTER (Wesseling & Heuvelink 1991; Wesseling et al 1996); Raper & Livingstone (1995) present an object-orientation approach.

The advantages and disadvantages of writing numerical models in a special GIS programming language are that the models can be easily developed or changed, encouraging modelling as a means of communication and exploration; the GIS-model provides a powerful decision support tool. Knowledge bases and error checking and error propagation studies can be incorporated (Burrough 1992). The disadvantages are that unskilled users may be uncritical of the results and the models may still be too simplistic. There is the danger that if modelling is too easy then field work for calibration, validation and investigation may be increasingly neglected.

To sum up: carrying out responsible environmental numerical modelling with GIS implies:

a) special training or trained personnel
b) special hardware and software
c) an organized database
d) large amounts of spatially referenced data.
e) possible increased computation times
f) need to learn/translate to and from a new computer language

However, the potential benefits that may accrue will include:

g) the model can be used for 2D, 3D or even 4D situations

h) standard data formats speed up data entry and display of results

I) results are immediately visible as graphs, maps, 3D displays and even videos so that managers can easily appraise the results.

j) sensitivity analyses and error propagation studies can be carried out interactively and the results can be seen in terms of their spatial context.

k) if appropriate interfaces are available, model building can be part of the GIS data analysis and so open to all users. This could lead to a better understanding of the modelling problem.

7. References

Barrow, J.D., 1991. Theories of Everything, Vintage, p.11.

Beckett, P.H.T., Webster, R. 1971. Soil variability - a review. Soils & Fertilizers 34: 1-15.

Bierkens, M.F.P. 1994. Complex confining layers: a stochastic analysis of hydraulic properties at various scales. Netherlands Geographical Studies No. 184.

Bierkens, M.F.P., Burrough, P.A. 1993a. The Indicator Approach to categorical soil data I. Theory. J. Soil Science 44: 361-368

Bierkens, M.F.P., Burrough, P.A. 1993b. The Indicator Approach to categorical soil data II. Application to mapping and landuse suitability analysis. J. Soil Science 44: 369-381.

Biesheuvel, A., Hemker, C.J. 1993. Ground water modelling and GIS: integrating MICRO-FEM and ILWIS. Proc. HydroGIS: Applications of GIS in Hydrology and Water Resources. IAHS Publ. 211, 289-296.

Bouma, J., Bregt, A.K. 1989. (Eds). Land Qualities in Space and Time, PUDOC, Wageningen.

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resource Assessment. Oxford.

Burrough, P.A. 1989. Fuzzy mathematical methods for soil survey and land evaluation. J. Soil Science 40: 477-492, 1989.

Burrough, P.A. 1992. The Development of Intelligent Geographical Information Systems. Int. J. Geographical Information Systems 6: 1-12.

Burrough, P.A. 1993a. Soil variability revisited. Soils and Fertilizers May 1993, 529-562.

Burrough, P.A. 1993b. Fractals and geostatistical methods in Landscape Studies, Chapter 5 in: Fractals in Geography, edited by Nina Lam and Lee da Cola. Prentice Hall pp. 85-121.

Burrough, P.A., Frank, A.U. 1996. Geographic Objects with Indeterminate Boundaries. Taylor and Francis.

Deutsch, C.V., Journel, A.G. 1992. GSLIB - Geostatistical Software library and user's Guide. Oxford University Press.

Fedra, K. 1993. Distributed models and Embedded GIS: Strategies and case studies of integration. Proc. Second International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modelling, Breckenridge, September 1993.

Goméz-Hernández, J.J., Journel, A.G. 1992. Joint sequential simulation of multigaussian fields. In: A. Soares (ed), Proc. Fourth Geostatistics Congress, Troia, Portugal. Quantitative Geology and Geostatistics (5) Kluwer Academic Publishers 85-94.

Goodchild, M.F., Parks, B.O., Steyaert, K.T. 1993. Environmental Modeling with GIS. Oxford University Press.

Goodchild, M.F., Parks, B.O., Steyaert, L.T., Glendinning, S. 1996. Environmental Modeling and GIS, Proceedings 2nd Conference on Integrating Geographic Information Systems and Environmental Modeling, GIS-World Publications 1996.

Heuvelink, G.B.M. 1993. Error Propagation in Quantitative Spatial Modelling. KNAG, University of Utrecht Pub. 163.

Heuvelink, G.B.M., Burrough, P.A. 1993. Error propagation in Cartographic Modelling using Boolean logic and continuous classification. Int. J. Geographical Information Systems 7:231-246.

Isaaks, E.H., Srivastava, R.M. 1989. An Introduction to Applied Geostatistics. Oxford, New York.

Jakeman, A.J., Littlewood, I.G., Whitehead, P.G. 1991. Catchment- scale rainfall-runoff event modelling and dynamic hydrograph separation using time series analysis techniques. In: D.G. Farmer & M.J. Rycroft (eds). Computer Modelling in the Environmental Sciences. Oxford.

Klir, G.J, Folger, T.A. 1988. Fuzzy sets, uncertainty and information. Prentice Hall, New Jersey.

Lavalée, D., Lovejoy, S., Schertzer D., Ladoy, P. 1993. Nonlinear variability of Landscape Topography: Multifractal Analysis and Simulation. Chapter 8 in: Fractals in Geography, edited by Nina Lam and Lee da Cola. Prentice Hall (1993). 158-192.

Maidment, D.R. 1993. Developing a spatially distributed unit hydrograph by using GIS. In: K.Kovar & H.P. Nachnebel (eds.). Application of Geographic Information Systems in Hydrology and Water Resources Management HydroGIS 1993. IAHS Publication No. 211. 181-192.

Maidment, D.R. 1995. GIS/Hydrologic Models of non-point source pollutants in the vadose zone. Proc. Bouyoucos Conference, May 1-3, 1995, Riverside, Ca.

Moore, I.D., Turner, A.K., Wilson, J.P., Jenson, S.K., Band, L. 1993. GIS and Land-surface-subsurface Process Modeling. in: Goodchild, M.F., Parks, B.O. & Steyaert, L.T., (eds). Environmental Modeling with GIS. Oxford University Press, New York, 196-230

Oreskes, N., Schrader-Frechette, K., Belitz, K. Verification, Validation and Confirmation of Numerical Models in the Earth Sciences. Science 263: 641-646.

Raper, J. Livingstone, D. 1995. Development of a geomorphological spatial model using object-oriented design. Int. J. Geographical Information Systems, 9: 359-383.

Rodriguez-Iturbe, I., Marani, M., Rigon, R., Rinaldo, R. 1994. Self- organized river basin landscapes: Fractal and multifractal characteristics. Water Resources Research 30:3531-3539.

Van Deursen, W.P.A., 1995. Geographical Information Systems and Dynamic Models. PhD thesis, Utrecht University, NGS Publication 190.

Wesseling, C.G., Heuvelink, G.B.M. 1991. Semi-automatic evaluation of error propagation in GIS operations. In Proceedings EGIS '91, edited by J. Harts et al. (Utrecht: EGIS Foundation), pp. 1228-1237.

Wesseling, C.G., Van Deursen, W.P.A., Karssenberg, D.J., Burrough, P.A. 1996. Integrating dynamic environmental models in GIS: the development of a Dynamic Modelling language. Transactions in GIS 1: (in press).

3.2 Requirements for model integration

D.W.G. van Kraalingen

DLO Research Institute for Agrobiology and Soil Fertility, P.O. Box 14, 6700 AA Wageningen, The Netherlands

1. Why integration of models

Summary and detailed models describing chemical, physical and biological processes are being developed in many areas of scientific research. It is a natural development that, after these models have more or less proven to be able to describe the target system, these models are pulled out of the research environment and are integrated as part of larger structures. Some of the common types of integration of crop models into larger structures are:

- coupling with Geographical Information Systems (GIS) and large scale yield monitoring systems (requires integration of crop models with mapping/database software, such as in CGMS, Hooijer & Van der Wal, 1994)
- coupling with General Circulation Models (GCM) (requires integration of crop models with atmospheric models predicting future climate, Matthews et al. 1995)
- coupling with systems for optimization of on-farm management (requires integration of crop models with risk analysis and/or sensing devices such as in Precision Agriculture, Bouma et al., 1995)
- coupling with atmospheric models in estimation of air pollution damage
- coupling with weed models in estimation of yield loss due to weed infestation (Kropff & Van Laar, 1995)
- coupling with disease and pest dynamics models in estimation of yield loss due to diseases and pests (Boote et al., 1993)
- coupling with educational software (requires integration with user-interface software)

Often the result of the integration is scientifically satisfactory, the investment required to complete the product is, however, often grossly underestimated. An integration project generally goes through the following phases:

- *Problem definition*

  Requirements of the finished product are determined. The data inputs and outputs of the whole system are described, decisions made on the treatment of missing data etc.

- *Excitement*

A feeling of well-being among the participants in the project usually precedes the start of the actual work

- *Awareness of problem complexity*

  While actually starting to implement the integration, the true complexity of the system often appears. Awareness rises about the difficulty of having some outside system take over control of the models, situations where the data are such that Model A can run but model B cannot run, etc.

- *Awareness of low data quality*

  Especially with systems that feed on large quantities of measured data. It is quite common that 15% of such data are unusable.

- *Panick*

  The feeling of "how on earth can the product be made to match up with the original requirements" takes over the project team, considering the real complexity and data quality.

- *Who is to blame, who is responsible?*

  Why didn't someone see this, wasn't this obvious from the start? Who said this was doable?

- *Adjusted goals*

  Goals are adjusted, time-limits extended, assumed accuracy of integrated result is reduced.

- *Finish*

  The project finishes, over time, but meets the adjusted goals.

2. Main problems in model integration

In general, problems can arise because of scientific and of technical incompatibility. The scientific compatibility comprises the nature of the interaction variables of the involved models. In other words is it scientifically valid to use the output variable X of model A as input variable for model B and if so, what is the subsequent reliability of the outcome of model B? An example is the use of cloud cover estimates to make (accurate) predictions of water loss on a field scale.

It turns out that most of the time, the major problems in the integration are not of a scientific nature but of a technical one. The reason for this is that the added-value and scientific ins and outs of integration of models are often studied carefully before the actual integration is started. Researchers usually have a clear idea of the type of data that are required to run their models, and the type of data that the model produces. The more frequent technical problems that are encountered depend on the type of integration at hand.

Some **general** problems during integration are:

- *Lack of technical documentation*

  Information about the use of file units, meaning and unit of variables, call tree, general structure of model etc., step control organization.

- *Lack of scientific documentation*

  What is the valid range of the input data, what are the state variables etc.

- *Robustness of system components*

  Often no error checking is done on the input data, causing the system to have abnormal behaviour

- *Occurrence of yet unknown errors in system components*

  Models are often used in an unforeseen (=untested) way in the integrated system. Errors that can be encountered are for instance continuity errors (see Fig 1.).

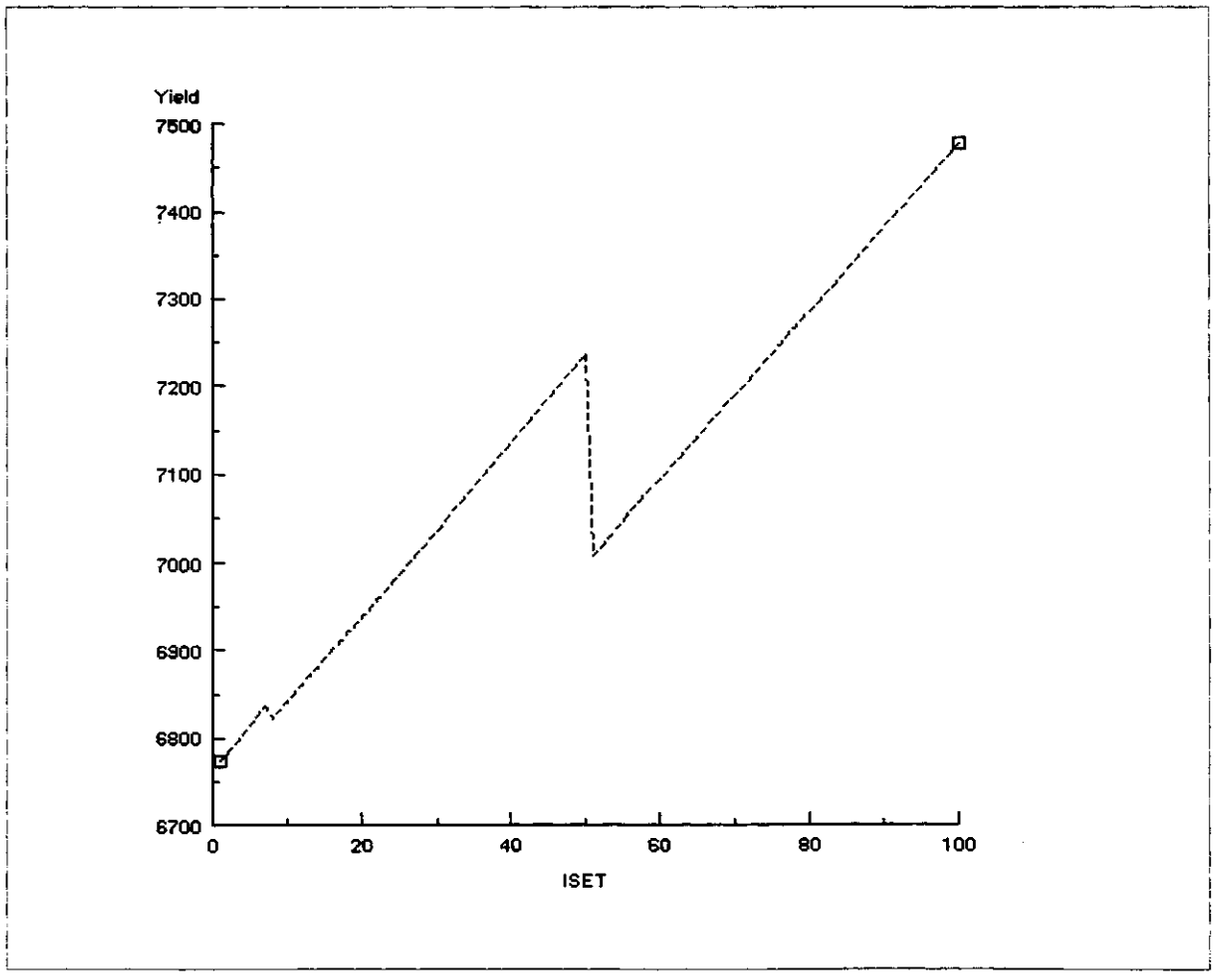


Figure 1. Unforeseen discontinuities in model output when an input variable is increased with small increments

The specific technical problems when integration involves **covering more processes** are

- Difficulties synchronizing the models through time (usually bad model structure: no separation of input and output from calculations, timing, and driving)
- Difficulties getting combined input and output in the appropriate format
- Unit / Dimension inconsistency between system components

Technical problems when integration involves **covering more area** are:

- Quality of input data: what to do when data are missing
- Complexity of overall system
- Model structure

The issue of technical softare quality has also been analysed within SC-DLO and has initiated a project on software process improvement (Van der Velden et al., 1995). The above-mentioned problems are not by any means limited to scientific programming. The software industry has faced these problems many years ago and has adopted new methods for requirements specification, design, programming, testing, documentation, maintenance of software etc., the central problem being how to overcome complexity. The science of "Software Engineering" has grown from the many frustrations that developers and users have had in the past (Van Vliet, 1988; Pressman, 1987). Software Engineering is now taught on many universities and commercial courses are widely available. Not only has Software Engineering contributed to methods to improve the quality of the software itself (SERC-QUINT, 1992), it has also contributed to the analysis and recommendation of organizational conditions that must be met in order to produce high quality software. Among them is the Capability Maturity Model (Paulk et al., 1993), a method by which organizations can be ranked and improved according to their "maturity" with regard to software development.

3. Possible solution to integration problems

With the integration problems as they are described above, integration and also modelling itself could benefit tremendously by applying Software Engineering principles and by making Software Quality Assurance a more integral activity of model development. Areas wherefrom benefits are to be expected are:

- standardization (by means of guidelines for model structure, data formats, documentation, good progamming practices etc.)
- improved Configuration Management (often responsibility of scientific content of a model is unclear, bug reports are not handled properly, no version numbering scheme exists)

- testing (many models are written such that thorough testing becomes very difficult)
- technical reviews of software by others
- applying programming discipline
- have researchers take courses in Software Engineering

An outsider might argue that applying the above approach does not include any guarantee about the correctness of the model. This is certainly true but a basic rule of scientific reseach is that no theory (such as a model) can be proved, it can only be falsified. Any activity that has shown to reduce occurrence of errors of whatever kind is therefore worth carrying out. Applying this approach will in some cases require a mental shift from "*Getting it to work*", to "*Getting it right*".

At AB-DLO a Software Quality Assurance project has recently started that aims at improving the "product quality" of the models, being, the verifyable quality of the product as it is, but also the "process quality", being the methods with which the model was developed. Some of the results of this ongoing project are reported in Van Kraalingen (1995), where a modelling framework is presented, the Fortran Simulation Environment (FSE) system, which is structured in such a way that components can be interchanged. The general design is given in 8 layers of definition in Table 1, in analogy with the OSI seven layer network model (Marciniak, 1994). The principle in this design is the identification of the different required functionalities and their mutual dependencies. The integration of models proves to be greatly simplified if this design hierarchy is reflected in the structure of the model software

The dynamic model level in FSE contains a specific feature that enables synchronization of models while retaining a simple subprocess structure (for details see Van Kraalingen, 1995). The problem being solved by this approach is that the combination of several subprocess models does not imply that all state update calculations have to be put in one block of consecutive statements and that all rate update calculations have to be put in another block (an important requirement from dynamic simulation theory). Such an action would split subprocess information across several places in the model, which adds to many of the integration problems mentioned above. As shown in Fig. 2, the modularity of the subprocess descriptions is preserved by introducing the concept of task-controlled execution. In the case of FSE, the Model driver level has control over the type of calculation that is required (either updating the state variables in the state section or updating the rate variables in the rate section). In this approach, the equations describing the separate subprocesses kan be kept together which facilitates the use of parts of the model to be taken out and used elsewhere. Also submodels that are written with this design philosophy in mind can be coupled more easily with each other.

Table 1: General design philosophy of the FSE system (Van Kraalingen, 1995).

Layer level	Layer name	Explanation
8	User interface	definition of user interface (to be accessed by the user, accesses the Analysis tool level)
7	Analysis tool	Statistical analysis such as DSSAT seasonal and sequential analysis tools (accesses the Run driver level, to be accessed by the User interface level)
6	Run Driver	definition of functionality, Input and Output (I/O) and other standards of 'Run driving' software (to be accessed by the Analysis tool or User interface level, accesses the Model driver level)
5	Model Driver	i.e. software driving one run through time (e.g. the FSE driver, to be accessed by the Run driver level, accesses the Dynamic model level)
4	Dynamic model	i.e. a complete dynamic process description (e.g. a FSE model, to be driven by the model driver, to be accessed by the Model driver level, accesses the Simple arithmetic subroutine and utility routine level)
3	Simple arithmetic subroutine and utility routine	e.g. rate calculating subroutines, interpolation functions, etc., how are the interfaces to the data files, how do date conversion routines work (examples: data read routines in FSE, to be accessed by the dynamic model level, accesses the High level data format level)
2	High level data format	what are the variable names and units for specific processes, what are the data types, lower and upper bounds of variables (to be implemented in the Low level data format level, to be accessed by the simple arithmetic subroutine and utility routine level)
1	Low level data format	what are rules for variable names, what are the data types, how are missing values treated, how is hierarchy organised at the data level etc. (to be used with actual data=high level data format level)

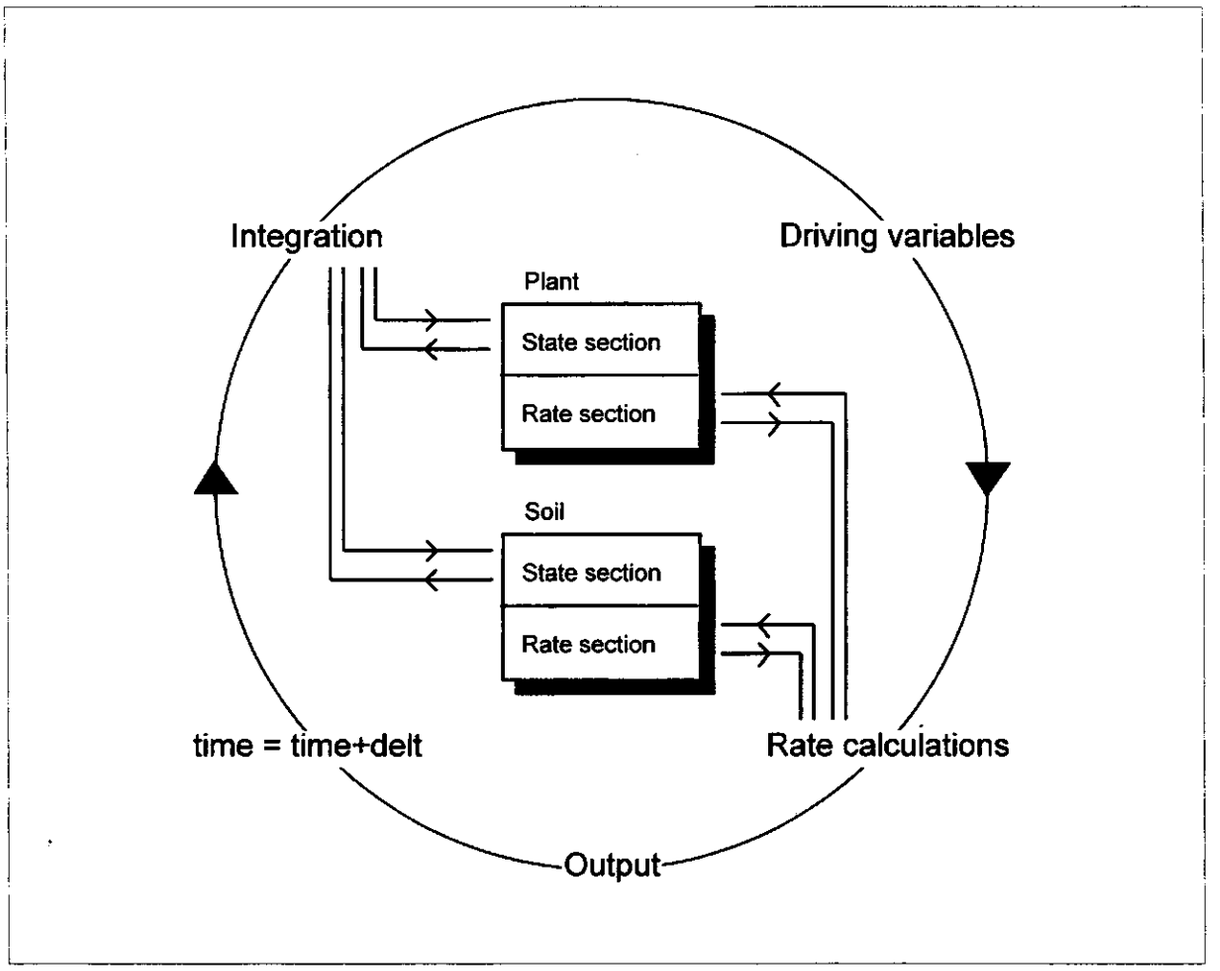


Figure 2. General structure for incorporating several subprocesses, illustrated for a plant and a soil subroutine (shaded rectangles) containing integration and rate calculation into a single simulation model.

4. References

Boote, K.J.,Batchelor, W.D., Jones, J.W., Pinnschmidt, H., Bourgeois, G. 1993. Pest damage relations at the field level. In: F.W.T. Penning de Vries et al. (Eds.) Systems Approaches for Agricultural Development 277-296. Kluwer Academic Publishers. Dordrecht. The Netherlands. 542 pp.

Bouma, J., Brouwer, J., Verhagen, A., Booltink, H.W.G. 1995. Site specific management on field level: high and low tech approaches. In J. Bouma et al. (Eds.) Eco-regional approaches for sustainable land use and food production, 453-473. Kluwer Academic Publishers. Dordrecht. The Netherlands.

Hooijer, A.A., Van der Wal, T. 1994. CGMS version 3.1: User manual. Technical document 13.1. SC-DLO. Wageningen. The Netherlands.

Kropff, M.J., Van Laar, H.H. 1993. Modeling Crop-Weed Interactions. CAB International. Wallingford. UK.

Marciniak, J., 1994. Encyclopedia of Software Engineering. John Wiley & Sons, Inc. New York.

Matthews, R.B., Kropff, M.J., Bachelet, D., Van Laar, H.H. Modeling the impact of climate change on rice production in Asia. CAB International. Wallingford UK.

Paulk, M.C., Curtis, B., Chrissis, M.B., Weber, C.V. 1993. Capability Maturity Model for Software, version 1.1. Report No. CMU/SEI-93-TR-24. Software Engineering Institute. Pittsburg, USA.

Pressman, R.S. 1987. Software Engineering: a practitioner's approach. McGraw-Hill International Editions.

SERC-QUINT, 1992. Het specificeren van software-kwaliteit, een praktische handleiding, Kluwer, Deventer, The Netherlands.

Van der Velden, M.J., Hendriks, P.R.H., Udink ten Cate, A.J. 1995. Software process improvement in a research environment. In: Ross et al. (Eds.) Software Quality Management III. Vol I Quality Management. Computational Mechanics Publication. Southampton Boston.

Van Kraalingen, D.W.G., 1995. The FSE system for crop simulation, version 2.1 Quantitative Approaches in Systems Analysis No. 1.

Van Vliet, J.C., 1998. Software Engineering. Stenfert Kroese. Leiden, The Netherlands.

3.3 Examples of object-oriented design in agricultural research

P.J. Schotman

Department of Computer Science, Wageningen Agricultural University, Dreijenplein 2, 6703 HB Wageningen, The Netherlands

Object-oriented techniques have become widespread in software engineering. One of the newest developments in object-oriented technology is the specification and use of design patterns. This paper shows how two of these design patterns can be used in (agricultural) research. The "Template Method" design pattern applied to modeling and simulation shows the benefits of object-oriented technology with respect to model adaptation. The use of this template together with a proper class hierarchy could be a crucial step in the development of a reusable model base. The "Composite" design pattern is shown as a second example and can be worthwhile in GIS applications.

1. Introduction

The rise of object-orientation reflects and recapitulates the history of computing as a whole. The earliest works in computing concerned themselves with what we now call 'programming'. Only later did a conscious concern with design and analysis as separate issues arise. Similarly, it is object-oriented programming that first attracted attention, later design and analysis issues became dominant (Graham, 1994).

In brief, object-oriented programming languages (OOPL) have extended functionality as compared to their conventional counterparts. In some languages this extension is simply added to the already existing non-object-oriented language specification, e.g. Pascal (Borland, 1995) or C++ (Stroustrup, 1991), while other languages e.g. Java (Gosling & McGilton, 1995; Aitken, 1996) are defined around the concept of object-orientation. Language implementations also differ extensively, in some implementations every language element is an object (e.g. Java or Smalltalk) while in others (e.g. C++, Delphi or Fortran90) this is not the case.

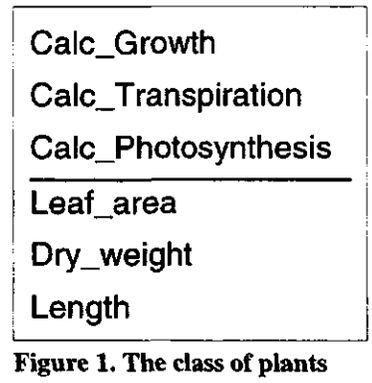


Figure 1. The class of plants with some methods and attributes

Object-oriented programming (OOP) together with its pendants object-oriented design (OOD) and object-oriented analysis (OOA) and their combination OOAD could be seen as a way of thinking when solving problems that require programming. Booch (1994) defines an object as follows: "An *object* has state, behavior and identity; the structure and behavior of similar objects

are defined in their common *class*". Figure 1 shows an example of a class; objects (or instances) are unique *instatiations* of classes in that they combine the structure (attributes or *slots*) and behavior (procedures or *methods*) that is defined for the class. Structure and behavior are defined at the class level so that they can be used at the object level.

2. Object-oriented technology

2.1 Promises of object-oriented technology

Graham (1994) uses 15 pages to elaborate on the advantages of object-oriented methods, the following are highlighted in this paper:

- Easier modeling of the real world. Modeling a problem in terms of *components* that have both structure and behavior seems to be very natural. The meaning and semantics of real world objects can be captured more effectively. Modeling real world objects into hierarchies and reasoning on the basis of these hierarchies has been a prominent research subject in the field of artificial intelligence (Rich & Knight, 1991).
- Productivity improvement. Reuse of earlier software engineering efforts is considered a predominant advantage of object-oriented method. However, this advantage has proven difficult to realize (Pancake, 1995; Meyer, 1995). User interface builders are a classical example of code reusers. Although reuse used to focuse upon code reuse, it was realized recently that design reuse may also be of practical value (Schmidt, 1995).

In agricultural research, modeling and simulation is of increasing importance. Researchers are building simulation models themselves and/or use (and usually modify) models built by others. There are advantages when adopting an object-oriented approach, the most prominent may be the easy insight in the structure and workings of a model and the way parts can easily be modified to suite ones needs. It is not surprising that the first object-oriented programming language (SIMULA67) was specifically intended for simulation (Dahl et al., 1982). Unfortunately, owing to insufficient commercial support this language has not become a success.

2.2 Basic features of object-oriented technology

Apart from the already mentioned coupling between structure and behavior, object-oriented technology is generally centered around the following three attributes: inheritance, polymorphism and encapsulation. The latter two can be considered as special cases of "abstraction" (Graham, 1994). Here they are considered separate.

2.2.1. Polymorphism

Polymorphism applies to the behavior of an object, it refers to the way objects of the same class respond to a *message* that is sent to them. Messages are calls to procedures that belong to the interface of an object. Objects communicate through these calls. Messages can be considered as imperative statements that are to be executed by the receiver (e.g. plant_47.length corresponds to "plant_47 return your length"). Messages to different objects may share the same name although their implementation may be different. Consequentlly, upon execution, the resulting "chain of events" may also be different.

2.2.2. Inheritance

Inheritance refers to the process of obtaining attributes and behavior among classes. Given the Plant class in Figure 1, a *sub*-class **CAM_Plant** could be defined which would inherit all attributes and most of the behavior of the regular **Plant** class. Some behavior will be different, though, and the class **CAM_Plant** may have additional attributes. In this particular case, the photosynthesis process (amongst other things) will be different, therefore, the method that describes the photosynthesis process will have to be different. The standard method for Calc_Photosynthesis defined for the class **Plant** will be *overridden* with a new method (with the same name) belonging to the class **CAM_Plant**. All objects that are instances of the **CAM_Plant** class will apply this new method and behave differently with respect to *this* procedure compared to objects that are instances of the **Plant** class. Note that for behavior that is *not* overridden instances of both classes will behave exactly the same. Object-oriented programming languages have (internal) logics to determine the most applicable method for a given object.

Inheritance relations result in classes hierarchies, sub-classes relate to their parent classes through the "IS-A" relation. Other hierarchies are also possible, for instance, the "PART-OF" hierarchy. This hierarchy relates objects to one another and is prominent in geographical information systems (GIS). An object-oriented design for this relation will be given later.

2.2.3.Encapsulation

Encapsulation refers to the *interface* of objects. The interface of objects is defined at the class level and it contains all messages to which objects of that class can respond. Through encapsulation it is possible to hide implementation details from the object's use. By doing so a software engineer creates a certain amount of freedom. This freedom allows him to easily change details of certain classes without disturbing their use. Some programming languages (e.g. C^{++} or Delphi) offer specific identifiers to mark certain methods and attributes as hidden, others do not (e.g. CLOS (Keene, 1989)). However, some simple software engineering directives (e.g. "Never access an attribute of an object directly.") also suffice to hide the implementation details of a class.

2.3 Support for object-oriented software construction

There currently consists a wide range of methods (or methodologies) and tools that help the software engineer both during the analysis of a problem and design and construction of the software solution. The most prominent methodologies seem to be: object oriented design and analysis (OODA) (Booch, 1994), object modelling techniques (OMT) (Rumbaugh, et al., 1991), combined object oriented analysis and object oriented design (OOA/OOD) (Coad & Yourdon, 1991a; 1991b) and object oriented software engineering OOSE (Jacobson, et al. 1992). Van Bergeijk et al. (1995) show the use of the OOSA methodology (Shlear & Mellor, 1992) in the agricultural domain. Tools that support these methodologies have also been developed (FAQ, 1995). These tools range from informal to formal. Tools that support formal design specification methods sometimes allow for automatic (complete) source code generation or template generation (i.e. name, class and interface of methods are specified but the actual workings have to be filled in). Above and other methodologies have been extensively compared and reviewed, for instance in: Wirfs-Brock & Johnson (1990) , Monarchi & Puhr (1992), Van den Goor et al. (1993) and Wilkie (1993). They will not be discussed further.

3. Design patterns for reusable object-oriented software

Reuse of earlier software engineering results has proven difficult to realize for many reasons (Meyer, 1988). Design patterns may be a promising technique for widespread reuse of software architectures (Schmidt, 1995). Design patterns stem from architecture. Christopher Alexander (in Gamma et al. 1994) describes a pattern as follows: "Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of a solution to that problem in such a way that you can use this solution a million time over, without ever doing it the same way twice.". Design patterns explicitly capture knowledge that experienced developers already understand implicitly. Using this knowledge can lead to better designed software because it contains the craftsmanship of experts in the field. Design patterns also function as a vehicle for communication between the developers and users of the source code. In the next sub-sections two templates from Gamma et al. (1994) are taken as an example for problems that are important in research related to production ecology (PE). Other information on design patterns can be found in Schmidt (1995).

4. Simulation as an example of the "Template Method" design pattern

Reuse and alteration of simulation models built by others frequently occurs in PE research. In many cases, careful inspection of the source code listings (usually FORTRAN) is necessary to

determine the place of one's own changes. Problems become serious when these changes imply changes to the interface of one or more sub-routines in the model. After some time the model may have evolved significantly and, unless the changes have been documented carefully, differences between this adapted model and the original one may be difficult to determine. Applying the Template Method design pattern can help facilitate this well known problem. Changes easily stand out and different versions become apparent through the class structure.

The Template Method design pattern can be used to define the skeleton of an algorithm of an operation deferring some steps to subclasses. The pattern lets subclasses redefine steps (called: primitive operations) of the algorithm without changing its structure (Gamma et al., 1994). As an example, the calculations per time step in a simulation model are being viewed as a template method.

Figure 2 shows, in OMT notation, a simple class hierarchy for which the template method Calc_Photosynthesis is shown. This template is defined for the abstract class **Crop**. For brevity, the figure does not show the attributes of the class. The arrows indicate inheritance relations (i.e. "IS-A" relations). *Abstract classes* and *abstract primitive operations* are shown in *italics*. Abstract primitive operations (also called "hook" operations) do nothing in the abstract class and are meant to be overridden. Abstract classes should not, as a matter of design principle, be used directly. One should first define a concrete sub-class, then define the behavior of the not yet defined hook operations, hereafter the concrete class and its interface are ready for use.

An important goal in designing template methods is to minimize the number of primitive operations that the final concrete sub-class needs to override. With the use of predefined models (i.e. template methods) in a model base this should certainly be the case, otherwise the model base will be of limited value.

The primitive operations that constitute the template method `Calc_Photosynthesis` are: `Get_measured_Light`, `Calc_Light_above_Crop`, `Calc_Light_Distribution_in_Crop`, and `Calc_Photosynthetic_Rate`. The actual implementations of two these primitive operations are also defined for the **Crop** class. As can be seen in the figure, the abstract classes **Greenhouse Crop** and **Field Crop** define different methods for `Calc_Light_above_Crop`. Obviously, the method for **Field Crop** is rather simple, it just passes on the measured light, while the method for **Greenhouse Crop** can be much more complicated because it must take the greenhouse construction into account[1].

In Figure 2, sub-classes of **Greenhouse Crop** define methods for `Calc_Light_Distribution_in_Crop`. When a researcher devises another planting arrangement he or she should create a new sub-class of **Greenhouse Crop** and redefine the method for `Calc_Light_Distribution_in_Crop` for this sub-class.

---

[1] The actual greenhouse will be stored as an attribute in one of the slots of the actual crop object. In an object-oriented implementation this greenhouse will most likely be an object that has specific information concerning its roof structure and a method to calculate its (dynamic) transmissivity.

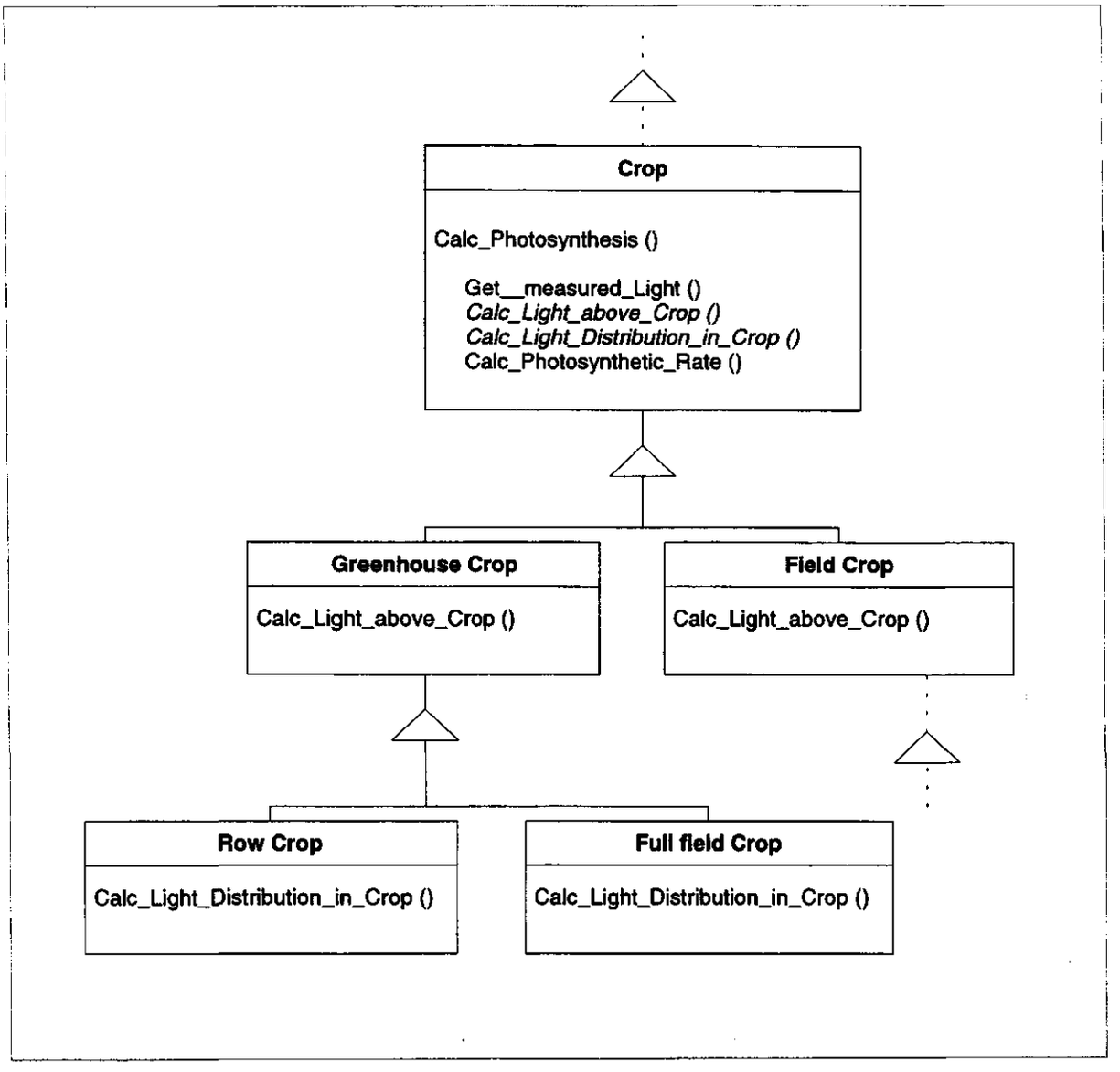


Figure 2. A simple class hierarchy in which classes with methods are shown

The use of the `Calc_Photosynthesis` template method is simple. The use of an object is only possible when it has been created (in Delphi-pascal: `greenhouse-crop-23 := GreenhouseCrop.Create;`) and initialised (i.e. appropriate slots LAI, Dry_Weigth, etc. are being assigned an intial value). The use consists of the following message: `greenhouse-crop-23.Calc_Photosynthesis` being send by some client. Typically, this message is being send during a simulation run.

Some additional remarks:

- Template methods can be nested. The `Calc_Photosynthesis` template method could be part of the `Calc_Growth` template method.
- Elaborate and simple models of the same process can live along side of each other in "parallel" class hierarchies. Differences in template methods are distinguished through the class structure. For instance, under the class **Crop** two subclasses could be defined

Elaborate Crop and **Simple Crop** each having a template method for `Calc_Photosynthesis`. The sub-class structure below both classes can be the same, however it is likely that the **Elaborate Crop** contains a more detailed sub-class hierarchy.

- Since research is a dynamic activity, its results, such as model base growth, but also template methods, change over time. New versions of a model base will contain improved template methods and/or improved primitive operations. Version control procedures for software can be used as a method to document those changes.
- It is possible to opt for different stategies with respect to structural and temporal changes of the class hierarchy. *Temporal* changes carried out by *users* of the hierarchy first involves creating a sub-class below the *lowest* level in the hierachy (in Figure 2 this would mean either below **Full Field Crop** or **Row Crop**). Second, redefinition (overriding) of the method that the user wants to change is carried out. This could be any method of the interface (i.e. all the methods inherited) of the newly defined class. Permanent changes will be carried out by the *maintainers* or designers of the model base and involves restructuring of the hierarchy, that is, inserting new classes at higher levels and implementing appropriate methods for these classes. Temporal changes can become permanent (thus inserted at the appropriate level) when they serve a more general use.
- Data belonging to an object can be stored in the attributes or slots (or *data members* in C++). This can make the list of arguments of methods simple because the attributes of a class are available and need not be supplied as arguments when defining a method. In an object-oriented design forcing function input (e.g. measurements of the environment) may be retrieved by sending a message to an *environment* object that carries this data, for instance: `greenhouse-compartment-23.Temperature` returns the current temperature.

To summerize, the Template Method design pattern shows a clever use of polymorphism. Polymorphism is important in model bases.

5. The "Composite" design pattern as a basis for scaling

The Composite design pattern has been intended to compose objects into tree structures. These tree structures represent part-whole hierarchies (through the "PART-OF" relation). The basic idea behind this pattern is the uniform treatment of individual objects (or leaf nodes in the tree hierarchy) and composites (i.e. sub-trees). The Composite pattern can be found in graphics applications like object-oriented drawing editors and almost every interface toolkit (e.g. Krasner & Pope, 1988) available today. In research, composite structures are typical for GIS applications.

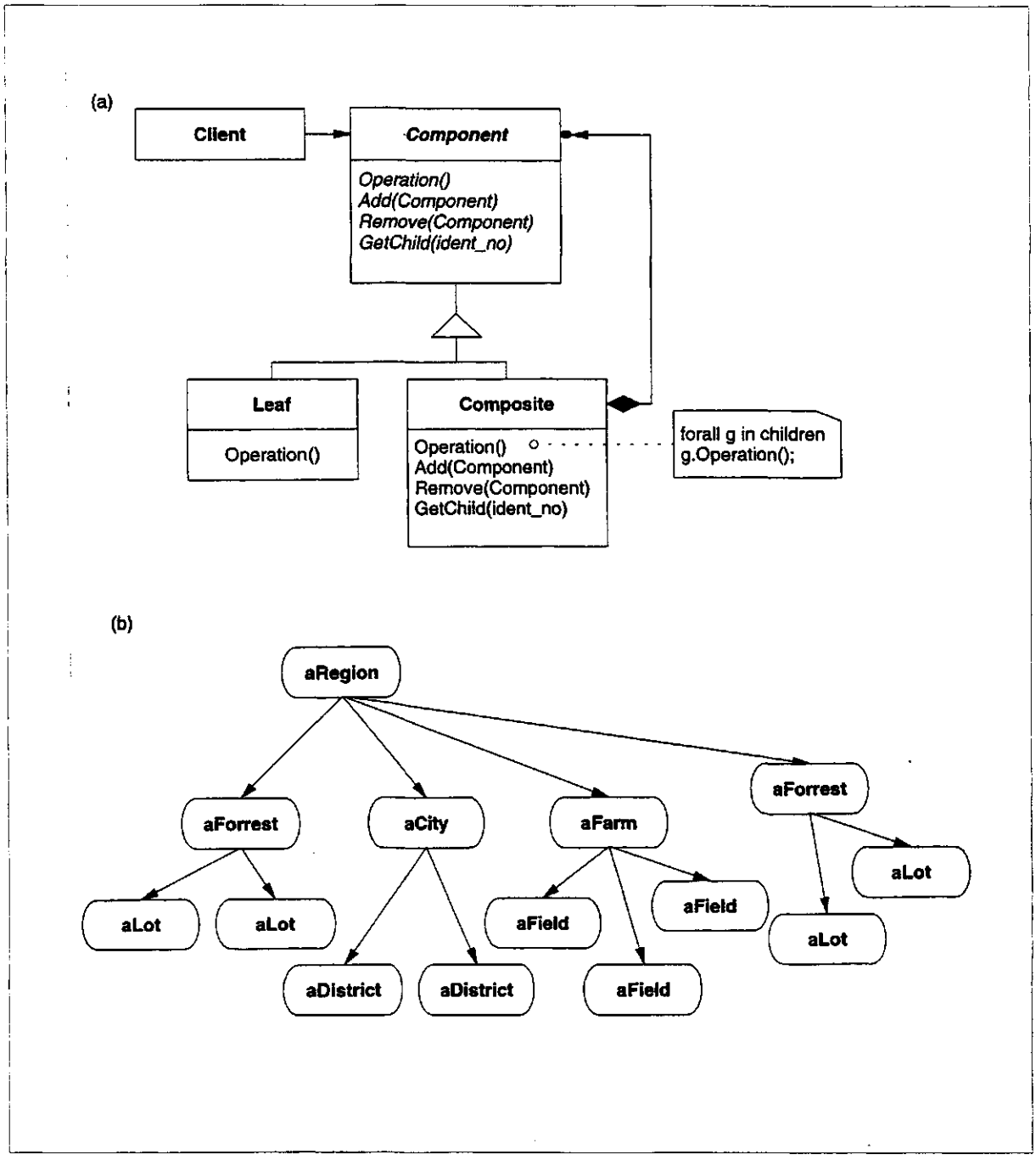


Figure 3. The Composite design pattern (a) and an example (b)

Figure 3a shows the structure of the design pattern. In a geographical information system, the abstract class ***Component*** could be called ***GIS Object***. All kind of abstract operations could be defined for this abstract component. These operations should be filled in for the appropriate composite and leaf classes. It should be noted that there can be many concrete **Leaf** and **Composite** classes. They may also be arranged into class hierarchies themselves. For instance, the **Lot** and **Field** class will have many attributes in common, they could both be sub-classes of

the abstract class ***Rural Unit***. The Composite design pattern allows for an easy interface for a "client" that calls a function on (i.e. sends a message to) an object in a specific part-whole hierarchy (like the one in Figure 3b). Figure 3b shows a simple object hierarchy, the arrows indicate the "PART-OF" relation among objects. An object hierarchy like the one in Figure 3b can be made as detailed as an application requires. There can be many objects of the same type, all representing different individual entities.

With the easy interface for clients, there are some additional responsibilities to attend to. For instance, some of the operations defined for the ***Component*** only apply to composites, appropriate error trapping needs to be defined for primitives (e.g. it is not allowed to add a component to a leaf object). Also, one would like the logic of an application to be automatically enforced which means that one does not want every composite to be part of another composite (e.g. in Figure 3b: **aCity** cannot be part of **aForrest**).

In the simplest case, operations defined for composites could be an accumulation of conceptually the same operation defined for its primitives. However, this does not need to be the case and at least provisions should be made to integrate the outcome of the operation for different leaf classes. An operation like `Calc_Production` called on **aRegion** in Figure 3b will result in a combination of agricultural products (all **aFarms**), an amount of lumber (all **aForrests**) and the production of **aCity** (which will be the combination of industries present in the city). One could also define `Calc_Production` for **Regions** to return something like the economic gross product of the region. Possibilities are infinite, the design pattern supports scaling in a very natural way.

6. Conclusions

Two design patterns have been shown to illustrate the possibilities of object-oriented technology in the domain of production ecology research. The main advantage of the ideas presented in the examples above will be its reuse by people other than the developers. This immediately shows the drawback: object-oriented model bases are not yet available and it will require some effort (i.e. money) to develop them. However, this investment will easily be recovered when researchers are going to use and extend the model base. Within the GIS community the situation may be better object-oriented geographical information systems have already been reported (see for instance the contribution of Burrough in this issue).

Apart from the advantages in the more traditional ways of modeling, newer modeling approaches like individual modeling (e.g. Hogeweg & Hesper, 1990) or multi-agent modeling (e.g. Attonaty & Pasquier, 1996) will profit even more of an object-oriented implementation.

7. References

Aitken, G. 1996. Moving from C++ to Java. Dr.Dobb's Journal March 1996, 52-56.

Attonaty, J.M., Pasquier, G. 1996. Multi-agent models: an approach to be explored to improve individual decision-making aids. Eds. Beers, G., R.B.M. Huirne & H.C. Pruis (eds.) Farmers in small-scale and large-scale farming in a new perspective: objectives, decision making and information requirements. Den Haag, LEI-DLO p.183-197.

Borland, 1995. Dephi user's guide. Borland International Inc.

Booch, G. 1994. Object-oriented analysis and design with applications. Second edition. Benjamin Cummings.

Coad, P., Yourdon, E. 1991a. Object-oriented analysis. Second edition, Yourdon Press.

Coad, P., Yourdon, E. 1991b. Object-oriented design. Yourdon Press.

Dahl, O.J., Myhrhaug, B., Nygaard, K. 1982. Common base language [SIMULA 67]. Fourth edition. Norwegian computing center, Oslo.

FAQ. 1995. Frequently Asked Questions of the newsgroup comp.object. available from: "rtfm.mit.edu/pub/usenet/comp.object".

Gamma, E., Helm, R., Johnson, R. Vlissides, J. 1994. Design patterns, elements of reusable object-oriented software. Addison Wesley.

Gosling, J. McGilton, H. 1995. The Java language environment, a white paper. October, 1995, 86 p. available from: java.sun.com

Graham, I. 1994. Object oriented methods. Second edition. Addison Wesley.

Hogeweg, P. Hesper, B. 1990. Individual modeling in ecology. Mathl Comput. Modeling 13, 83-90.

Jacobson, I. Christerson, M., Jonsson, P., Overgaard, G. 1992. Object-Oriented Software Engineering - A use case driven approach. ACM Press/Addison Wesley.

Keene, S. 1989. Object oriented programming in Common Lisp, a programmer's guide to CLOS. Addison Wesley.

Krasner, G.E., Pope, S.T. 1988. A cookbook for using the model-view controller user interface paradigm in Smalltalk-80. Journal of object-oriented programming. 1, 26-49.

Meyer, B. 1995. On the role of methodology. In: Object-oriented technology for databases and software systems. Eds. Alagar, V.S. & Missaoni, R. World Scientific, Singapore. p.1-6.

Meyer, B. 1988. Object-Oriented Software Construction. Prentice Hall.

Monarchi, D., Puhr, G.I. 1992. A Research Typology for Object-Oriented Analysis and Design. Communications of the ACM 35, 35.

Pancake, C.M. 1995. The promise and cost of object technology: a five year forecast. Communications of the ACM 38, 33-49.

Rich, E., Knight, K. 1991. Artificial intelligence. Second edition. Mc Graw Hill.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., Lorensen, W. 1991. Object-oriented modeling and design. Prentice-Hall.

Schmidt, D.C. 1995. Using design patterns to develop reusable object-oriented communication software. Communications of the ACM 38, 65-74.

Shlear, S., Mellor, S.J. 1992. Object-oriented systems analysis: Modeling the world in states. Yourdon Press.

Stroustrup, B. 1991. The C++ Programming Language (2nd edition). Addison Wesley.

Van Bergeijk, J., Goense, D., Wigboldus, K. 1995. Objectgeorienteerde analyse en design van real-time systemen. Een toepassing van de Shlear-Mellor methode bij het ontwikkelen van een boordcomputer. Agro-informaticareeks no.9 juni 1995.

Van den Goor, G.P.M., Brinkemper, S., Hong, S. 1993. Objectgeoriënteerde ontwerpmethoden. Informatie, 35, 840-851. (in Dutch)

Wilkie, G. 1993. Object-Oriented Software Engineering - The Professional Developer's Guide. Addison Wesley.

Wirfs-Brock, R.J., Johnson, R.E., 1990. Surveying Current Research in Object-Oriented Design. Communications of ACM 33, 104-124.

4. Stages of life of models

This chapter has a more philosophical scope than the previous ones. Central questions involve: is modelling an end-station, or can alternatives be distinguished? When do models age and why? What ages in models and how can this be prevented?

4.1 Emerging maturity or arrested development in crop modeling?

Thomas R. Sinclair[1] and No'am G. Seligman[2]

[1]USDA-ARS, Agronomy Physiology Lab., University of Florida, P.O. Box 110840, Gainesville, FL 32611-0840, USA

[2]The Volcani Center, Institute of Field and Garden Crops, P.O. Box 6, Bet Dagan 50250, Israel.

Crop modeling, the computerized simulation of dynamic crop systems, was born about 30 years ago when systems analysis and modern computers presented a new technique to crop scientists. Since then, crop modeling has gone through a number of development stages similar to that of a biological organism: infancy, juvenility, adolescence, and maturity. The development cycle for crop models began with successes in modeling those factors influencing crop photosynthesis. The development cycle progressed through the burgeoning complexity of models in the juvenile stage. The adolescence stage was marked by confusion and frustrations because modeling expectations were severely constrained by the complexities of reality. Presently, crop modeling seems to be on the one hand in an arrested development, and on the other approaching a more mature acceptance of modest expectations. These modest expectations seem to focus on a heuristic role to facilitate logical, quantitative thinking about crops.

1. Introduction

Crop modeling, defined here as the dynamic simulation of crop growth by numerical integration of constituent processes, emerged just over 30 years ago. Professor C.T. ("Kees") de Wit working in the Agricultural University in Wageningen, The Netherlands, was a pioneer and dominant personality in the development of crop modeling (De Wit, 1970; Rabbinge et al., 1990). His leadership and genius in identifying and quantifying the critical factors that influence plant growth in an agricultural context, set the tone for crop modeling as a scientific activity. Even though he foresaw the pitfalls of crop modeling, his warnings often went unheeded as many models became more and more complex and increasingly irrelevant to real problems in science and crop management.

The development of crop modeling is, in a way, analogous to that of biological organisms that proceed from infancy to maturity (Sinclair & Seligman, 1996). We suggest that in this case, maturity implies a clearer view of the constraints and complexity of biological and agronomic reality together with recognition of the specific role of modeling as a tool for heuristic

investigation of crop behavior in a dynamic and conceptually tractable mode. The implications for model structure, complexity and 'transparency' are briefly discussed.

2. Infancy

After World War II, system analysis and computer science, stimulated by the Cold War and space exploration, became widely available for the study of complex systems. Crop modeling was born in this exciting era of new technologies. The first steps were models developed to estimate light interception and photosynthesis in crop canopies (Loomis & Williams, 1963; De Wit, 1965; Duncan et al., 1967). They facilitated the calculation of the light profile in a canopy & made it possible to assess the sensitivity of canopy photosynthetic rates to sun angles, leaf angle distribution, and the latitudinal position of the crop.

These relatively simple models opened the way to quantitative estimates of maximum attainable growth rates. Crop growth and potential yield became demonstrably linked via biochemical and biophysical mechanisms to the amount of solar energy available for the accumulation of chemical energy and plant mass. After the early successes in modeling photosynthesis of leaf canopies, it seemed only a matter of time before a full description of crop growth and development would be incorporated into models.

3. Juvenility

Like the widening horizons of childhood, the new technologies opened new vistas for many scientists. It seemed straightforward to model the many factors that influence crop yield: weather, soils, crop genetics, plant physiology, pest damage, and agro-technology. The promise was that age-old questions about the linkages among factors could be resolved so that crop yields could be readily predicted and crop manipulations could be optimized. Models offered the promise of abbreviating experimentation in the evaluation of genetic material and new management techniques in the context of a wide range of cropping environments (e.g., Bowen et al., 1973).

The complexity of crop models increased as the various processes that determine crop growth were incorporated into the models. Important advances in describing various sub-components of carbon assimilation in particular were made during this period. The significance of stomatal conductance in regulating leaf gas exchange was quantitatively described (e.g., Cowan, 1977). The fate of photo assimilates in respiratory pathways was carefully analyzed (Penning de Vries, 1975).

Development processes of plants became an important consideration as the time frame of models was lengthened to include the entire growing season. Expressions for the partitioning of

assimilate among various tissues, particularly to the reproductive organs, were developed, often based mainly on speculation. Ultimately, the addition of these various components led to a number of models with daunting complexity, e.g., GOSSYM (Whistler et al., 1986), CERES (Ritchie et al., 1985), SOYGRO (Wilkerson et al., 1985).

Development of complex models was accompanied by stresses and strains frequently associated with juvenility. The number of parameters required to describe the system in detail increased greatly. Complex experimentation was needed to estimate the biological coefficients that describe crucial cultivar characteristics. This in itself gave rise to much valuable research. However, the requirement for large sets of empirical coefficients meant that the inevitable experimental errors propagated and often compounded through the models (Diekkrüger et al., 1995). Variables that could not be measured directly in experiments had to be estimated, often 'guesstimated'. These variables were often "calibrated" with the model, thereby turning the modeling process into a cumbersome curve fitting process. All these complications inhibited implementation of the model and created murky interconnections that hindered understanding of model behavior.

It was not always appreciated that the role and function of models in solving engineering problems does not apply to biological systems (De Wit, 1970). An engineering model is a point of departure that leads to the construction of a structure or a device. All the components are defined and have clear specifications with appropriately low tolerances. In the case of crop models, the point of departure is a very complex biological system and the model is therefore of necessity a highly simplified surrogate system, even when defined in great detail. Consequently, the 'tolerance' between the model and the actual system is inevitably wide and attainment of one-to-one representation is practically unattainable, except in trivial cases. The enormous number of chemical and physical processes that operate simultaneously in the plant and its environment make it futile to attempt a full description of crop performance (Mayr, 1982; Pease & Bull, 1992). In addition, each growing season, each region, each field, each site, offers a new environment where the interplay of new and different factors determine crop performance. In this juvenile stage, the dilemma presented by the attempt to describe the complexity of the crop system with an intrinsically inadequate model became increasingly acute.

4. Adolescence

Adolescence is commonly associated with considerable confusion and turmoil. In this stage of transition from juvenility to adulthood, basic assumptions are questioned and perspectives are changed. The unbounded possibilities of earlier developmental stages shrink as the realities of limited resources, conflicting options, and only partial control over system function become increasingly apparent. So too with crop modeling, some of the original tenets need to be reevaluated in the light of accumulated experience.

4.1. Extensive Reductionism.

An assumption that was implicit in many crop modeling efforts was that scientific rigor required mechanistic rather than descriptive models. This was commonly interpreted as reductionism that required processes to be expressed in basic physical, chemical and physiological terms. Consequently, there was a tendency to adopt an extensive reductionist approach while maintaining an integrated, holistic framework so as to 'extrapolate the laboratory to the field'. As a result, complex systems came to imply complex models that inevitably involved 'descriptive' representations of processes. This included weakly supported assumptions like, for instance, hypothetical pools of compounds that respond to "supply" and "demand". When a high level of plant organization is being modeled, use of such constructs may well give a more distorted representation of organ growth than the use of conservative allometric relationships.

There are many examples where detailed, reductionist models are less reliable than simpler models for simulating observations. A simple water balance model was found superior to COTTAM and GOSSYM in approximating crop water stress and field water balance (Asare et al., 1992). An empirical equation was found superior to CERES in predicting annual potential wheat yields in Mexico (Bell & Fischer, 1994). SOYGRO was found to be inferior to a simple average of a sample in predicting irrigated soybean yield in unsampled populations (Colson et al., 1995).

4.2. Universal Models.

While the development of a community of elements in any biological system must follow a general pattern set by genetic 'blueprints' controlled by negative and positive feedbacks, it is impossible to predict the precise developmental path of each organism even in relatively homozygous crop populations. This is not unlike the uncertainty recognized by physicists in predicting the future path of individual atoms. Environmental heterogeneity further increases unpredictability so that within the physical bounds set by system integrity, the crop system is open and there are unlimited paths for individual plant development. This 'noise' or random component in experimental observations sparked the development of statistical theory for analysis of variance.

The practical consequence is that it is impossible to create universal crop models, even in the environment for which the model was developed (Spitters, 1990). Each new season or new location brings new challenges that were not foreseen in the original model. So, for example, attempts to use existing crop models developed for higher latitudes failed when an attempt was made to simulate crops in the semiarid tropics of Australia (Carberry & Abrecht, 1991). Important deficiencies were found in each of three complex wheat models even after they had been 'calibrated' for a new set of conditions in New Zealand (Porter et al., 1993). Large variation among wheat genotypes in the response of their ontogenetic development to environmental change could not be described by a single generic model (Slafer & Rawson, 1994). Considerable

effort and model modification are required to make models account for discrepancies that derive from changes in cultivars, cropping conditions and peculiarities of the application environment.

4.3. Validation.

Claims for the 'correctness' or 'truth' of a model are commonly based on validation by comparison of model output with observed crop performance. [Although technically model validation only denotes establishment of legitimacy rather than verification, validation and verification are commonly used synonymously (Konikow & Bredehoeft, 1992).] The fundamental difficulty with this claim is that crop models are a collection of hypotheses and not a single falsifiable hypothesis, so they inherently cannot be validated (Pease & Bull, 1992; Oreskes et al., 1994). Other collections of hypotheses may approximate the experimental results equally well.

Validation data themselves are subject to substantial experimental and observational error (Smith, 1995). A validation exercise can only show how well (or badly) a model performs in a particular circumstance. It cannot guarantee the performance of the model under any other environmental condition especially when the model has been 'calibrated' to fit a specific circumstance or set of circumstances (Oreskes et al., 1994).

In short, some of the basic tenets originally proposed as guidelines for crop model construction have not stood the test of time: models are not necessarily improved by extensive reductionism, universal crop models cannot be constructed, and models cannot be validated. The very heavy investment in time and resources in model development have seldom produced anything but meagre, if not trivial results (Seligman, 1990). A new perspective on the role, construction and benefits of crop models may signal an emerging maturity.

5. Maturity

Maturity implies awareness not only of the limits to system behavior but also of the nature of the essential limiting factors. The ability to identify critical signals and relationships, and to separate them from the often very loud background noise, can develop from the accrued life experiences of an organism and from the finite resources it has been able to sequester. Such an analogy seems appropriate to characterize some recent developments in crop modeling.

The limits of crop models as reliable surrogates for reality have become evident. However, crop models do provide a framework in which to test assumptions about potential crop response to environmental, structural, functional, or parameter change. In addition, models may allow the discovery of faulty reasoning or interesting implications about a crop. In this context, crop models are valuable heuristic tools to aid our interpretation of reality (Wüllschleger et al., 1994).

The conclusion that biological models are most useful in a heuristic role was anticipated many years ago by Noy-Meir (Unpublished technical report, 1972). In a retrospective on a

workshop to develop biome models as part of the International Biological Program, Noy-Meir concluded "that the construction of the model was an exercise of great intellectual and education value for every one of the participants". This conclusion was reached in spite of the fact that no working model was produced by the workshop. Noy-Meir goes on to argue that the additional benefits to be gained by making a model operational may not even be worth the effort (Innis et al., 1980). The value of the model was to force logical, quantitative thinking about the variables and processes that influence the performance of the organisms of interest.

The heuristic benefit of crop models in teaching is clear. Crop models were introduced into the classroom more than 15 years ago and upgraded teaching models continue to be developed (e.g., Waldren, 1984; Hart & Hanson, 1990; Wüllschleger et al., 1992). Crop modeling exercises are perceived by students as an effective tool for illustrating the relative importance of factors that influence crop production (Meisner et al., 1991). Learning is likely to be facilitated by using models that are simple and transparent enough in structure to allow students to dissect it and to understand the logic underlying its behavior.

Research on crop systems or subsystems can use models to organize concepts and information that reflect current understanding and to determine their adequacy in explaining relevant phenomena. Shortcomings of the model can highlight important but poorly understood aspects of the crop. In that sense, models can be more useful when they fail than when they succeed! In addition, crop models can be quite useful in analyzing experimental results by virtue of their ability to substantiate possible causes of differences in the results. They provide a level of interpretation that goes beyond a statement of statistical significance in experimental results.

Even the use of crop models in farm management has succeeded more in an heuristic role than as an on-line decision aid. Two examples are the SIRATAC model for cotton pest management in Australia (Ives & Hearn, 1987) and the EPIPRE model for wheat pest management in The Netherlands (Rabbinge & Rijsdijk, 1983). Each model required growers to pay for membership and to supply field observations to a central processing center. At the central processing center, model simulations were done to provide growers with updated pest management recommendations. In each case, there was an initial steady increase in grower membership which resulted in a general improvement in pest management. However, both systems suffered a loss of membership after the initial successes. The decline in participation has been ascribed not to dissatisfaction with the model results but to the fact that the growers felt they had learned the lessons of the models and could now manage on their own (Weiss, 1994). The models were a success in that they taught the growers improved pest management by helping them interpret their own field observations more effectively.

6. Attributes of Heuristic Modeling

Recognizing the role of crop models as heuristic tools in teaching, research, and applied activities can help to define the attributes of a modeling exercise that are more likely to produce

better returns on the investment of time and resources. The following list includes some of the more important attributes.

(1) A clear statement of specific objectives is essential to define the need and nature of a crop model. It is more likely that success will be achieved when the objectives are modest, tractable, and have a clear raison d'etre for a modeling approach. What specific situation is to be investigated and what problem is to be studied with the model?

(2) Criteria to judge the acceptability of a model should be defined in relation to its objectives. In an applications mode, statistical criteria concerning model predictions relative to a sample of observations are appropriate. In a research mode the criteria for acceptability are more ambiguous. High predictive capability may be less important than the need to identify weaknesses in conceptualization of hypotheses about particular processes. A research model is more likely to be concerned with behavioral patterns than with precise quantitative predictions.

(3) The heuristic benefit may be greatest when the modeling approach is not prejudiced by automatically using existing models. While efficiency demands that successful approaches used previously not be ignored, as new problems arise new configurations may need to be developed or an old model may need to be modified for a conceptually new objective. This mode of operation is facilitated when research and extension personnel are able to construct their own models, customized to specific problems and not overloaded with the considerable redundancy necessary to construct a supposedly universal or generalized crop model. These ad hoc models should be as simple as the nature of the objective allows.

(4) The organizational level of the problem (tissue, organ, plant, canopy, or crop) should determine model structure. It is rare that a model objective at one level of organization cannot be achieved by modeling processes at only one subordinate level of description; the 'explainable' and the 'explanatory' levels of De Wit (1970). Increasing crop model complexity by adding more levels and peripheral processes can seldom improve model performance or relevance. Rather the contrary. Excessive complexity will obscure and even distort whatever heuristic benefit may be gained about crop performance at the desired level of interest (Diekkrüger et al., 1995).

(5) Summary models of 'emergent properties' or 'conservative relationships' (Penning de Vries, 1982; Monteith, 1990) should be incorporated into models whenever appropriate. There are several summary relationships that are sufficiently robust to efficiently express underlying empirical patterns or complex processes. Examples of such summary relationships include exponential radiation interception (Monsi & Saeki, 1953), radiation use efficiency (Sinclair & Horie, 1989), transpiration/photosynthesis relationships (De Wit, 1958; Tanner & Sinclair, 1982; Monteith, 1990), and maintenance and growth respiration functions (Penning de Vries, 1975).

Simplification, in contrast to an exaggerated reductionist approach, can be useful even if the simplification is not completely correct (Weiss, 1990; Maddox, 1990). Probably the best example is that of Newtonian physics which is fundamentally incorrect in view of modern

understanding of relativity and quantum physics. Still, the 300-y old equations of Isaac Newton continue to be crucial tools in science and engineering. Also in crop modeling, simplifying assumptions that are appropriate to specific circumstances can give rise to extremely useful relationships even though they lose their validity beyond a certain scale or situation.

7. Conclusion

Crop models have proliferated over the past two decades but have not made an impressive contribution to crop science (Seligman, 1990). In many cases it seems as if the crop modeling effort is not progressing but simply turning out more of the same. However, it can be shown that crop models, when appropriately constructed are useful heuristic tools in teaching, research, and in management and administrative applications. Hypotheses, knowledge, and data can be harnessed to enable the user to reason more consistently and transparently about factors or conditions that deserve thought by students, additional experimental study, or more attention from growers and administrators. Intelligent, clear reasoning, as well as observation, experimentation and experience cannot be replaced by crop models, but they can be well supported by them. Because of the large number of situations where crop models can play a useful heuristic role, we believe that a 'mature' approach to model construction and application will ensure that crop modeling technology be used effectively to further advance crop science.

8. References

Asare, D.K., Sammis, G.I. , Assadian, H., Fowler, J.F. 1992. Evaluating three cotton simulation models under different irrigation regimes. Agric. Water Manag. 22:391-407.

Bell, M.A., Fischer, R.A. 1994. Using yield prediction models to assess yield gains: a case study for wheat. Field Crops Res. 36:161-166.

Bowen, H.D., Colwick, R.F., Batchelder, D.G. 1973. Computer simulation of crop production - potential and hazards. Agric. Eng. 54(10):42-45.

Carberry, P.S., Abrecht, D.G. 1991. Tailoring crop models to the semiarid tropics. In: R.C. Muchow & J.A. Bellamy (eds.), Climatic risk in crop production: Models and management for the semiarid tropics and subtropics. CAB International, Wallingford, U.K. p. 157-182.

Colson, J., Wallach, D., Bouniols, A., Denis, J.B., Jones, J.W. 1995. Mean squared error of yield prediction by SOYGRO. Agron. J. 87:397-402.

Cowan, I.R. 1977. Stomatal behaviour and environment. Adv. Bot. Res. 4:117-228.

De Wit, C.T. 1958. Transpiration and crop yields. Agricultural Research Reports, 64.8. PUDOC, Wageningen, The Netherlands.

De Wit, C.T. 1965. Photosynthesis of leaf canopies. Inst. Biol. Chem. Res. Field Crops Herb. Agric. Res. Rep. No. 663. Wageningen, The Netherlands.

De Wit, C.T. 1970. Dynamic concepts in biology. In: I. Setlik (ed.), Prediction and measurement of photosynthetic activity. Pudoc, Wageningen, pp. 17-23.

Diekkrüger, B., Sondgerath, D., Kersebaum, K.C., McVoy, C.W. 1995. Validity of Agrosystem models: A comparison of results of different models applied to the same data. Ecological Modeling 81:3-29.

Duncan W.G., Loomis, R.S., Williams, W.A., Hanau, R. 1967. A model for simulating photosynthesis in plant communities. Hilgardia. 38:181-205.

Hart, R.H., Hanson, J.D. 1990. PASTORAL grazing simulator. J. Agron. Educ. 19:55-58.

Innis, G.S., Noy-Meir, I., Godron, M., VanDyne, G.M. 1980. Total-system simulation models. In: A.I. Breymeyer & G.M. VanDyne (eds.), Grasslands, Systems analysis and Man. Cambridge Univ. Press, Cambridge. pp. 759-797.

Ives, P.M., Hearn, A.B. 1987. The SIRATAC system for cotton pest management in Australia. pp. 251-268. In: P.S. Teng (ed.), Crop Loss Assessment and Pest Management, APS Press, St. Paul, MN.

Konikow, L.F., Bredehoeft, J.D. 1992. Ground-water models cannot be validated. Adv. Water Res. 15:75-83.

Loomis, R.S., Williams, W.A. 1963. Maximum crop productivity: an estimate. Crop Sci. 3:67-72.

Maddox, J. 1990. Virtue in imperfect models. Nature 347:13.

Mann, T. 1944. Joseph the Provider. A.A. Knopf, New York. p. 157-158.

Mayr, E. 1982. The Growth of Biological Thought. Harvard University Press, Cambridge, MA.

Meisner, C.A., Karnok, K.J., McCrimmon, J.N. 1991. Using crop models in a beginning crop science laboratory. J. Agron. Educ. 20:157-158.

Monsi, M., Saeki, T. 1953. Über den Lichtfaktor in den Pflanzengeselschaften und seine Bedeutung für die Stoff Produktion. Jpn. J. Bot. 14:22-52.

Monteith, J.L. 1990. Conservative behaviour in the response of crops to water and light. In: R. Rabbinge, J. Goudriaan, H. van Keulen, F.W.T. Penning de Vries & H.H. Van Laar (eds.), Theoretical Production Ecology: Reflections and Prospects. Pudoc, Wageningen. pp. 3-16.

Oreskes, N., Shrader-Frechette, K., Belitz, K. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263:641-646.

Pease, C.M., Bull, J.J. 1992. Is science logical? Bioscience 42:293-298.

Penning de Vries, F.W.T. 1975. Use of assimilates in higher plants. p. 459-480. In: J.P. Cooper (ed.), Photosynthesis and productivity in different environments. Cambridge Univ. Press, Cambridge, U.K.

Penning de Vries, F.W.T. 1982. Phases of development of models. In: F.W.T. Penning de Vries & H.H. Van Laar (eds.), Simulation of Plant Growth and Crop Production. PUDOC, Wageningen, The Netherlands.

Porter, J.R., Jamieson, P.D., Wilson, D.R. 1993. Comparison of the wheat simulation models AFRCWHEAT2, CERES-Wheat, and SWHEAT for non-limiting conditions of crop growth. Field Crops Res. 33:131-157.

Rabbinge, R., Rijsdijk, F.H. 1983. EPIPRE: A disease and pest management system for winter wheat, taking account of micrometeorological factors. EPPO Bull. 13:297-305.

Rabbinge, R., Goudriaan, J., Van Keulen, H., Penning de Vries, F.W.T., Van Laar, H.H. 1990. Theoretical Production Ecology: Reflections and Prospects. Pudoc, Wageningen, The Netherlands.

Ritchie, J.T., Godwin, D.C., Otter-Nacke, S. 1985. CERES- Wheat. A Simulation Model of Wheat Growth and Development. Texas A&M University Press, College Station.

Seligman, N.G. 1990. The crop model record: promise or poor show? In: R. Rabbinge, J. Goudriaan, H. van Keulen, F.W.T. Penning de Vries & H.H. VanLaar (eds.), Theoretical Production Ecology, Reflections and Prospects. PUDOC, Wageningen, The Netherlands. pp. 249-263.

Sinclair, T.R., Horie, T. 1989. Leaf nitrogen, photosynthesis, and crop radiation use efficiency: A review. Crop Sci. 29:90-98.

Sinclair, T.R., Seligman, N.G. 1996. Crop modeling from infancy to maturity. Agron. J. (In Press).

Slafer, G.A., Rawson, H.M. 1994. Sensitivity of wheat phasic development to major environmental factors: A re- examination of some assumptions made by physiologists and modellers. Aust. J. Plant Physiol. 21:393-426.

Smith, R.E. 1995. Opus simulation of a wheat sugarbeet plot near Neuenkirchen, Germany. Ecological Modeling 81:121-132.

Spitters, C.J.T. 1990. Crop growth models: Their usefulness and limitations. Acta Horticulture 267:349-362.

Tanner, C.B., Sinclair, T.R. 1982. Efficient water use in crop production: Research or re-search? In: H.M. Taylor, W.R. Jordan & T.R. Sinclair (eds.), Limitations of Water Use in Crop Production. Am. Soc. Agron., Madison, WI. p. 1-27.

Waldren, R.P. 1984. CROPROD: A crop management computer model for undergraduate agronomy courses. J. Agron. Educ. 13:53-56.

Weiss, A. 1990. The role of climate-related information in pest management. Theor. Appl. Climatol. 41:87-92.

Weiss, A. 1994. From crop modelling to information systems for decision making. In: J.F. Griffiths (ed.), Handbook of Agricultural Meteorology, Oxford Univ. Press, NY. p. 285-290.

Whistler, F.D., Acock, B., Baker, D.N., Fye, R.F., Hodges, H.F., Lambert, J.R., Lemmon, H.E., McKinnion, J.M., Reddy, V.R. 1986. Crop simulation models in agronomic systems. Adv. Agron. 40:141- 208.

Wilkerson, G.G., Jones, J.W., Boote, K.J., Mishoe, J.W. 1985. SOYGRO V5.0. Soybean Crop Growth and Yield Model. Technical Documentation. Agricultural Engineering Department, University of Florida, Gainesville.

Wüllschleger, S.D., Hanson, P.J., Sage, R.F. 1992. PHOTOBIO: Modeling the stomatal and biochemical control of plant gas exchange. J. Nat. Resour. Life Sci. Educ. 21:141-145.

Wüllschleger, S.D., Lynch, J.P., Berntson, G.M. 1994. Modeling the belowground response of plants and soil biota to edaphic and climate change - What can we expect to gain? Plant and Soil 165:149-160.

4.2 The future of modelling: trends in automated modelling

J.L. Top

Agrotechnological Research Institute ATO-DLO, PO Box 17, 6700 AA Wageningen, The Netherlands

The art of computer-based modelling is changing rapidly. In this paper we review a number of advances to be expected in the near future. Two important tracks in this development will be: (I) explicit formalization of background knowledge and (ii) structuring of the modelling process. Explicit knowledge will be represented as mathematical and logical descriptions. Yet more important, conceptual languages will be developed to link these highly abstract models to a common sense interpretation. On the other hand, a model is always concrete: it is an answer to a question. Support for asking the right question, thus guiding the modeller to an effective, high-quality answer for his or her problem will be necessary. We claim that automated modelling will exploit information technology to turn the art of modelling into a profession.

1. Introduction

Modelling is a basic human function that pervades all our conscious activities (Rothenberg, 1989). For example, even a simple statement saying: 'It's cold outside' already makes up a model. It tells us something about a part of the world without immediate observation. Viewing this statement as a model, it reveals some important aspects of modelling:

- *Generalisation:* 'Cold' is a generalisation since we can use it without referring to a concrete object. A model is based on abstraction, such that we can share information without giving a complete description of a system. Completeness is even impossible. Moreover, the generalisation is not only across objects, but also across individuals. In other words, modelling implies the need for a language to describe the world so that it can be understood by others.

- *Independent existence*: The abstraction and generalisation only exist in our minds. The model itself is something that exists as a concrete entity in the world. It has its own structure and properties. In this example it is a sentence in English with a certain intonation, loudness etc. The concreteness of a model limits the range of what can be expressed by it.

- *Goal directedness:* The statement *'It's cold outside'* can express different intentions: a warning, an explanation or simply an introduction. A model always exists within the context of a problem, a purpose. Therefore, the context of the model is needed to understand it properly.

These features are also essential for *computer-based modelling*, the subject of this paper. First, by the ability to abstract (in our minds) we are able to understand simulations in terms of the modelled system. However, it is not always clear to see what has (not) been modelled by looking directly at a piece of computer code. Therefore, software engineers use mathematics and logics to set up models in a formal way, such that ambiguity as to how the model should be interpreted and what its limitations are, is least. This is important because of the second factor: computer models consist of hard- and software. They are flexible, powerful and safe, but at the same time they are limited in expressiveness and performance. They can never represent the real system perfectly (Zeigler, 1976).
The third issue, *goal directedness*, is equally important in automated modelling. The process of building a model must be guided, such that the objective is met with as little side effects as possible. This means that ideally it should answer the question posed without loss of time and without the need for additional resources. Understanding the modelling process will help software and knowledge engineers to push the limits of modelling further in this direction.

In this overview we will look at some trends in automated modelling, given the essential tension between generalisation and specification. First, in Sec. 2 we consider the model as a bridge between theory and application. How can the contents of such a small-scale theory be formalized? In Sec. 3 the modelling process will be explicated as an essentially interactive process based on assembling generic pieces of knowledge. We will conclude by summarizing the most important developments in automated modelling to be expected in the near future. The ideas presented in this overview are based on previous work on modelling (Top, 1993; Top & Akkermans, 1994).

2. Formalization: from theory to application and vice versa

Formalization of models - before implementation as a computer programme - will be a major trend in the coming years (Akkermans et al., 1993). Formalization means that the information contained by models is represented in an objective (or better: intersubjective) way, independent of how it will be implemented. By using well-defined general terms, ambiguity can be reduced. Moreover, from the formal model the implementation can be generated almost automatically. In this section we will stake a look at the contents of models and which future developments we can expect with respect to the formal representation of that contents.

2.1 Modelling as small scale theory building

A model can be viewed as a small scale theory (Rosenbluth & Wiener, 1945) dealing with a concrete question, but still general. In a black-box approach (e.g. 'curve-fitting'), the internal structure of a model is irrelevant, if it properly reflects the observed behaviour through its inputs and outputs. One assumes that it will also will work for those input and output values for which no measurements were done: if the model were only valid for the measured values it would be useless. Now, the problem is that a true black box model cannot guarantee in any way that extrapolation is allowed. The usual, implicit, solution is to assume that the behaviour of a system does not change significantly in the ranges considered. However, there

is no possibility to check whether this assumption holds.

More support for the validity of a model (without more observations) can be obtained by employing theoretical knowledge in the form of *domain laws*. Then, the quality of a model is not only guaranteed by validation through observation, but also because by the fact that those theories have proved their value over time. Theory-based validation is sometimes called *confidence building* (Top & Akkermans, 1994). It can be expected that the search for domain invariants and laws ('energy' and its conservation being an extremely powerful example in classical physics) will continue, providing a basis for the small-scale theories that our models are. As for computer science we will see information systems gradually change into knowledge-based systems.

2.2 Mathematics and logics

The most general invariants employed in modelling are the elements of logics and mathematics: variables, equations, clauses, operators, etc. They allow formalization of a model before implementation in a computer programme. The important aspect of this high level of abstraction is that knowledge and data can be manipulated automatically; new values and statements can be derived without reference to the model context. All 'real world' aspects have been removed and the model has been reduced to something that can be complex, but very well defined. Computers happen to be good at doing formalized operations. They can detect syntactical errors and perform logical inferences. Therefore, logics and mathematics will remain primary players with respect to the formalization of knowledge.

In particular the following developments can be expected here:

- *Number crunching*. Increasingly powerful numerical and analytical algorithms will be constructed. This will not only affect the performance of simulators, but also make them less critical to the form the mathematical model.
- *Hybrid systems*. Continuous and discrete approaches will be integrated. This will enable us for example to switch automatically between (parts of) models given their validity constraints.
- *Formal verification and validation*. Automated theorem provers will improve, thus allowing extensive formalization and model checking - before generating the computer code.
- *Nonlinear analysis*. The recent interest in chaotic systems has given an important impulse to nonlinear systems analysis. This will significantly increase the scope of phenomena to be represented.

2.3 Conceptual languages

The problem (and point) of formalization in terms of logics and mathematics is that the meaning of the variables and expressions is completely removed. For example, if we are modelling an electrical circuit we may want to know the voltage across a resistor. Our mathematical model could be: $z = xy$. What does it mean? Without explanation it is unclear. Obviously, the meaning of the variables has to be supplied with the model to allow

interpretation. If the above law had been written as V=IR, using the standard notation, it would have been immediately clear for those who are familiar with the domain. However, what if there are more resistors in the circuit? The need for explanation seems obvious, but all too often the meaning of variables and expressions is left implicit.

A mathematical or logical model without explanation is useless. The assumptions about when and why it can be used are unknown. Thus, a purely mathematical model cannot be reused or shared. What is needed is a conceptual description that provides the link with our common sense understanding of the world. However, this conceptual level must be formal as well, eliminating the ambiguity of natural languages, but less general than the mathematical level. For example, the network description of electrical systems is a kind of conceptual language, provided that the symbols are defined as ideal standard processes (Top, 1993). Different languages will have to be defined for different application areas, with links between them.

An important function of conceptual languages is to hide complexity. For most cases the mathematical model can be derived automatically from the conceptual model. It has even been shown that computational (numerical) schemes derived from proper conceptual models in engineering are more effective (Van Dijk & Breedveld, 1991) than purely mathematical algorithms.

In this context we have to touch on the work done in *Qualitative Reasoning* (QR), a subfield of Artificial Intelligence (MQ&D, 1995; Weld & De Kleer, 1990). The researchers in this field claim to have filled the gap between mathematics and common sense. They propose a method for solving physical and other problems by defining a simple 'common sense' calculus based on a small subset of possible values (eg. 'negative', 'zero' and 'positive'). This approach is supposed to deal with incomplete information in a way that is similar to human reasoning. However, even a simple second order dynamic system leads to a multiple set of solutions, of which most are spurious. In our view this approach can never be successful because it essentially remains a mathematical approach. Human reasoning is based on conceptual thinking and can therefore not be simplified in this rigorous way.

With respect to the conceptual level we may expect the following developments:

- *Graphical conceptual languages.* More and better conceptual languages will become available. In the field of knowledge-based systems much energy is already put into the definition of *ontologies* for several areas (Gruber & Olson, 1994). These languages will be graphical rather than textual. This allows compositionality and direct syntax checking. Moreover, a graphical representation can more easily be inspected by the modeller (Karnopp at el., 1990; Paynter, 1961).

- *Model libraries.* Composite structures in terms of the above conceptual languages can be stored as a whole, with the underlying conditions for their application (Breuker & Van de Velde, 1994; Top et al., 1995).

- *Language integration.* Due to the proliferation of conceptual languages restructuring them will be necessary and to allow for true multidisciplinarity. Specific languages will be constructed on top of more specific languages, and languages of related domains will be linked. This will require an intimate link with theoretical research in the associated sciences, since new *invariants* should be defined.

So far we have only considered what can be expected with respect to the contents of models. However, the process of building a model is equally important and will dramatically change in the future.

3. Asking the right question

In the naive view of what modelling entails, one starts with some observations and from there the model is constructed. This view - which is still dominating - misses the point in two ways. First, usually some domain knowledge is used to start the modelling process. Second and more important: the model construction process is not a single-step process, but essentially iterative. The process of building a model is as important as the resulting model itself, since many decisions about the problem context are being made along the way. During this process the question that is being asked is gradually refined, until an answer can be provided with some assurance. In short: modelling is a *learning* process.

In Fig. 1 we show a schematic view of the modelling process. We distinguish three basic steps in this process: specification of the assumptions, construction of the model and assessment of the model. The modelling process starts with a (possibly vague) intention in the head of the modeller. Moreover, usually an initial model is available in one way or another. In traditional modelling approaches the emphasis has been on the construction step. Here, the building bricks (the *generic components* in Fig. 1) were software statements in Fortran, Pascal or any other computer language. Then the model was assessed by comparing it with the measured behaviour and adjusted where necessary. At some point the model was assumed to be good enough.

However, this approach is inadequate in several ways. First, no guarantee can be given that the model fulfills the requirements of the modeller. This is because these requirements are left implicit, i.e. in the head of the modeller. No specification step is performed. As a result, there is no possibility to control the quality of the models produced. The model cannot be used for a different case, or shared it with other modellers.

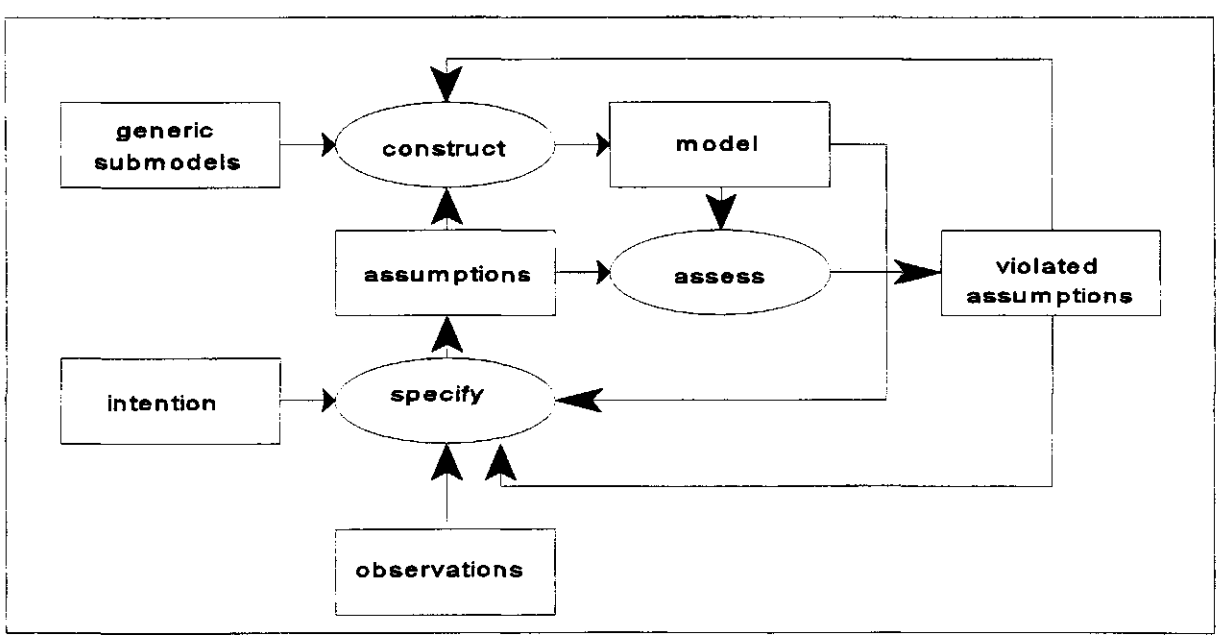


Figure 1 A general decomposition of the modelling task (Top & Akkermans, 1994).

Second, usually no distinction is being made between different viewpoints (see Fig 1). Computational and numerical considerations are easily mixed with conceptual issues. A typical example can be found in physical systems modelling: discrete changes in otherwise continuous model are often inserted as patches in the computer code. However, it has been shown (Strömberg et al., 1993) that introducing an explicit switching concept at the conceptual level eliminates many problems both at the computational and the conceptual level. The complex computer code can even be generated automatically, and the modeller only deals with the much more appealing conceptual switch.

The third reason that this direct coding approach is not the appropriate way of building models is that no guidance is given on how to proceed. The process of specifying assumptions, assembling the model and assessing it against the assumptions is not structured in any way. Therefore, it is a matter of trial and error. Clearly, more structured approaches to modelling, based for example on the elementary scheme presented in Fig. 1, will be needed to help the modeller to answer his or her questions more efficiently.

With respect to computer support of the modelling process we can expect the following developments:

- *Assumption management.* Explicit bookkeeping of the modelling assumptions will be a major step forward in the availability of high-quality models. The background information *about* models that is nowadays given in textual form - or, more common, not given at all - will be an inherent part of the models. This will help the modeller to ask the question he or she intends to ask.

- *Automated abstraction.* Based on general knowledge of what is typically relevant in a certain context, some default operations can be done automatically, provided that the modeller is made aware of them. For example, in modelling dynamic systems it is possible to consider only processes that have a certain time scale. This is called automated timescale abstraction.

- *Cooperative modelling.* By explicitly recording why and how a model is being constructed, the individual model fragments can be shared between modellers. Thus, modelling can become a collective activity rather than individual. The availability of communication networks - in particular Internet - will stimulate cooperation.

- *Adaptive technology.* Transparent shifting between different models, but also between different methods will increase the power of computer models. At the present, an important bottleneck is to how to decide when which technique should be used. Research is needed to formalize the selection process of application techniques.

- *Switching perspective.* As mentioned above, distinguishing between different views of the model contents is necessary. Future developments of conceptual languages and automated modelling will allow transparent switching between different perspectives (physical, mathematical, economical, ...) and propagate information across different views where possible.

4. Tools and techniques

So far we have dealt with the structuring and formalization of models and modelling, and showed some developments with respect to this. All these developments of course heavily depend on what the computer support systems allow us to do. What can we expect from computer science that will allow progress in automated modelling?

As concerns hardware:

- *Growing resources.* Mentioning the expected increase in computational speed and storage capacity is almost needless. Numerical barriers will become a less predominant factor in the modelling process. Note however that increasing computer power adds to the need for proper assumption management, although the opposite is often observed (Truesdell, 1984).

- *Networks.* As mentioned above, models will be developed in a distributed way. This not only refers to computational resources, but also to the cooperation between modellers and designers.

- *Multimedia.* Automated modelling will be less dependent on the textual form. As mentioned above, graphical representations will become more important. In particular a more realistic view will be used, exploiting the power of multimedia systems. Developments in the field of *virtual reality* are important. However, again this technology implies a risk with respect to modelling: is it clear which problem needs to be solved?

- *Sensor technology.* If models are used to solve 'real world' problems, calibration and validation by means of measurement cannot be avoided. The limitation of what can be measured will be pushed further. However, this will require extensive research. The construction of intelligent sensors, which is possible due to new IC-technology, will merge observation with interpretation.

If we consider software technology, the following trends will affect model building:

- *Object orientation.* Object oriented software engineering is particularly useful for building modelling and simulation software (Rumbaugh et al., 1991). Modular, pluggable systems will allow rapid prototyping and continuous, iterative design processes. Recently so-called *patterns* (Gamma et al., 1994) have become popular reusable frames for software design.

- *Database management systems.* To store and retrieve large amounts of model fragments in model libraries, efficient and accessible repositories will be developed. Searching of models is on meta-information (explaining the applicability of the model fragments within the context) rather than internal model structure. The modeller can find models through associative searching.

- *Communication.* Model support systems and databases must be made open, such that different systems can share information and knowledge. This implies exchange protocols and interoperability.

- *Graphical User Interfaces.* The interactive character of the modelling task requires immediate feedback from the computer support system. We mentioned already the importance of graphical modelling languages. Attempts to make incorrect structures can be reported immediately to the modeller. Moreover, the result of the simulation can be reported in a way that is close to our common-sense understanding of the world.

- *Machine learning.* Besides modelling based on explicit conceptual representations, also bottom-up techniques will be developed further. Machine learning techniques (genetic algorithms, neural networks, case based reasoning, etc.) generate black- and grey-box models from data sets. Moreover, the integration of data-driven and concept-driven techniques will significantly improve model support systems.

- *Knowledge-based systems.* We may expect important impact from developments in knowledge engineering. In particular the formalization of domain knowledge is given much attention in this discipline (Akkermans et al., 1993; Breuker & Van de Velde, 1994).

5. Conclusions

We have attempted to elucidate some developments that we can expect - or at least hope for - with respect to automated modelling. Of course, such an overview can never be extensive, if only because we have to rely on what we presently see as an adequate model of the modelling task.
In our view, the following aspects will significantly affect the modelling process in the future:

- Formalization through logics and mathematics.
- Formalization in terms of conceptual languages.
- Structured problem specification.
- Assumption management.

Two major bottlenecks may frustrate this development. First, parametrisation (identification) of models remains a difficult issue. This is due to a mismatch between what can be measured and what should be measured according to the model. Of course, both the model and the measuring system are critical. Second, defining formal models - at the logical-mathematical or at the conceptual level - will be difficult for ill-structured domains. Developing proper languages will be a major challenge in the coming years.

Both problems result from the essential problem in modelling: finding an appropriate tradeoff between generality (theory) and specificity (application). Theories provide a basis for quality models, but they are never applicable without limitations. Practical problems require practical solutions. We will need new ways to push the limits of generalisation *and* goal-directedness.

In that way automated modelling will become more accessible and more widely spread, but also stimulate creativity and innovation.

6. References

Akkermans, J.M., Van Harmelen, F., Schreiber, A.Th., Wielinga. B.J. 1993. A Formalisation of knowledge-level models for knowledge acquisition. Int. J. Intelligent Systems 8, pp 169-208.

Breuker, J. Van de Velde, W. (eds.). 1994. CommonKADS library for Expertise Modelling. IOS-Press, Amsterdam.

Gamma, E., Helm, R., Johnson, R., Vlissides, J. 1994. Design Patterns: Elements of Reusable Object-Oriented Software. Addison Wesley Publishers.

Gruber, T.R., Olson, G.R. 1994. An ontology for engineering mathematics. In: J. Doyle, P.Torasso & E. Sandewall (eds), Fourth Int. Conf. On Principles of Knowledge Representation and Reasoning, Morgan Kaufmann.

Karnopp, D.C., Margolis, D.L., Rosenberg, R.C. 1990. System Dynamics: A Unified Approach. John Wiley & Sons, NewYork.

MQ&D. 1995. Qualitative Reasoning: A Survey of Techniques and Applications. AICOM 8 (3/4).

Paynter, H.M. 1961. Analysis and Design of Engineering Systems. MIT-press, Cambridge, Mass., USA.

Rosenbluth, A., Wiener, N. 1945. The role of models in science. Philosophy of Science, 12(4):316-321.

Rothenberg, J. 1989. The Nature of Modelling. In L. E. Widman et al. (eds), Artificial Intelligence, Simulation and Modelling. John Wiley.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., Lorensen, W. 1991. Object-Oriented Modelling and Design, Prentice-Hall, Englewood Cliffs, NJ.

Strömberg, J.E., Top, J.L., Söderman, U. 1993. Variable Causality Caused by Discrete Effects. Proceedings of ICBGM'93, San Diego, 115-119.

Top, J.L. 1993. Conceptual Modelling of Physical Systems. PhD thesis, University of Twente.

Top, J.L. Akkermans, J.M. 1994. Tasks and Ontologies in Engineering Modelling. Int. J. Human-Computer Studies 41: 585-617.

Top, J.L., Breunese, A.P.J., Broenink, J.F., Akkermans, J.M. 1995. Structure and use of a library for physical systems models. Proceedings ICBGM '95, Las Vegas, 97-102.

Truesdell, C. 1984. The Computer: Ruin of Science and Threat to Mankind. *In:* An idiot's Fugitive Essays on Science: Methods, Criticism, training, Circumstances. Springer-Verlag.

Van Dijk, J., Breedveld, P.C. 1991. Simulation of System Models Containing Zero-order Causal Paths - I. Classification of Zero-order Causal Paths. J. Franklin Institute 328 (5/6), 959-980.

Weld, D.S., De Kleer, J. eds. 1990. Readings in Qualitative Reasoning about Physical Systems. Morgan Kaufmann Publishers, San Mateo, CA.

Zeigler, B.P. 1976. Theory of Modelling and Simulation. John Wiley, New York.

4.3 The end of modelling?

J. Goudriaan

Dept. of Theoretical Production Ecology, Wageningen Agricultural University, Bornsesteeg 47, 6708 PD Wageningen

After a brief history of simulation modelling the value of modelling as a new scientific method as compared to theory and experimentation is discussed. Its role is sketched in science itself, in teaching, in management and in policy. Development of new models may cease in case of dwindling availability of research funds. As long as the society does not break down, no end is foreseen to the use of models.

1. History of modelling

Models, defined as simplified representations of parts of reality (De Wit, 1968; Casti, 1990), are usually imagined as computer models. However, our own thoughts and ideas also satisfy this same definition. Thoughts can have a profound impact on the way people behave and therefore they are the true "Models in Action", of which the activity will never end, unless mankind is eradicated. However, they are volatile, ill-defined, poorly accessible and often even secret. Models as we understand them in an operational meaning should be available to other persons in an objective and reproducible form, so that they can be communicated unequivocally.

Nowadays we tend to see a model as an idealized but imperfect representation of a far more complicated real system. Interestingly, to the ancient Greek philosophers the situation was almost opposite (Popper, 1972; Koningsveld, 1976). To them, real objects were imperfect representations of the perfect "ideas" that would exist in the ideal world behind everything surrounding us. A drawing of a circle was only an imperfect model of the "real" circle. Therefore, even if we did measurements on a drawn circle to determine the value of π, the result would still be imperfect and it would be much better to use the methods of ideal abstract geometry to find the value of π (Barrow, 1992). This Platonic idea has become the first cornerstone of the way we conduct science. In mathematics this method has been extremely fruitful, but on the other hand it has seriously hampered the progress of science of the real world when it tempted "idealists" to think that they could deduce the properties of physical objects, just by careful reasoning. This dead end in science is too familiar to us when we replace "reasoning" by "modelling" and "properties of objects" by "crop growth". The crucial step towards modern science and technology was the insight that studying and measuring real objects can really teach us new things which mere reasoning cannot. This empirism is the second cornerstone of science (Dijksterhuis, 1950). Simulation modelling has the potential of becoming the third cornerstone in future science.

Modelling as a science in its own right began as an engineering activity when it was discovered that rules for small objects are similar in structure to those for larger objects. This was handy in trying out new designs, just by studying a physical scale model (Goudriaan, 1993).

Architectural designs, model ships, lysimeters and even a rhizolab are examples of this methodology (Penrose, 1989).

With the advent of the computer completely new possibilities were created. Complex arithmetic schemes and large constructions of concatenated reasonings could be evaluated, fulfilling the dreams of the "inductionists". The computer models were refined and improved, and over the decades the methods to construct and test them have grown into a new science.

2. What does simulation modelling add?

The reasons to use models instead of a direct study of reality depend on the segment of society: in teaching, in management, in policy making or in scientific research the reasons will be different.

- In teaching simulation modelling will be used to clarify and illustrate concepts, ideas and hypotheses. This works best if the models are simple, and do not have many interacting components. The demonstration of interactions will complicate the material considerably and will really only be useful if the students are on a more advanced level of study.
- In management models are used from a pragmatic point of view. They can support the expert knowledge of the manager, or even partly replace it.
- In policy the emphasis will be more on feasibility studies, and on exploration of options. More than in management the users will be tempted to bend and interpret both the model and its output in such a way that it will support their political view.
- In science model study is firstly a tool to analyze experimental results (Mayr, 1988). There is not enough appreciation of the value of putting the data into a consistent framework. This way of using models leads to whole new possibilities of data quality control and cross-checking of data.

The testing of hypotheses is best done if experiments are designed such that alternative hypotheses are exposed in extreme form. In astronomy, geo-sciences and ecology experiments are not always possible, and models can be invaluable by showing the implications of alternative hypotheses.

2.1 Emerging properties

A new area in science is the discovery that complex model behavior can sometimes show "emerging properties" at a higher level of organization. The new field of so-called computational physics is one result of this discovery. Some examples of emerging properties in model studies are:

- reciprocity relation between directional dependence of radiance and reflection coefficient of a reflecting structured surface (Goudriaan, 1977).
- temporary stabilization of intrinsically unstable modes (especially in meteorology and oceanography). This phenomenon leads to the existence of flip-flop phenomena. An example is the northward flow of warm water in the Atlantic Ocean towards the Arctic Ocean. This flow maintains the mild climate in North Western Europe. There have been occasions in the past (the Lower Dryas 12000 years ago) when this flow was suddenly interrupted and Europe cooled down dramatically.
- various types of sink-source interaction (apical dominance, self-destruction mechanism).

- stability of growth respiration coefficients (Penning de Vries et al., 1983).

2.2 Models and scientific hypotheses

Much of what we can say about models is also true for scientific hypotheses and vice versa. The term model has a very broad meaning. De Wit's definition of a model as a simplified representation of a part of reality does not restrict the size or the kind of model. We may find models as small as a single equation, or as large as thousands of lines of code. In both cases the term model is correctly used. Also, the term does not say whether the model is just conceptual and qualitative (Wegeners model of plate tectonics, Adam Smith's invisible hand), quantitative and deterministic (Keynes' model of economy, most crop growth models), qualitative and stochastic (Darwin's theory of evolution), quantitative and stochastic (General circulation models of the global weather system). The term model carries the notion of implementation, which a scientific hypothesis does not do.

2.3 The problem of model falsification

Scientific hypotheses should be continually subjected to attempts of falsification (Popper, 1959). The longer they stand up to tests (withstand falsification), the more confidence they gain and the more weight they will carry when they are applied in technology and policy.

Scientific hypotheses can be falsified at a qualitative level but also at a quantitative level (Nagel, 1961). At the qualitative level we might prove that a scientific theory is simply wrong, without using numbers or statistics. However, this is only possible if the case is obvious. The problem with a new and revolutionary theory is that the case will only be "clear to those who see", i.e. to those who have altered their viewpoint drastically so that they can see the case in a totally different light (Casimir, 1987; Luyten & Hoefnagel, 1995). Such a shift of paradigm does not happen often, but it is characteristic for major scientific revolutions. One of the best examples is Darwin's theory of evolution as driven by variation and selection (Darwin, 1859). There is also a danger in this kind of scientific progress. The wording "clear to those who see" carries the smell of religion, and it is for good reason that critical and quantitative argumentation is needed before acceptance of a new theory. The popular belief that a "better" economy can be obtained through reduction of the public sector is probably an example of such an unwarranted shift in perception. Indeed, a new theory can only withstand time if extensive quantitative attempts to falsify it are not successful.

Falsification of computer models requires a different approach. The difficulty is that models are usually composed of modules that represent hypotheses in different areas, so that it is not easy to tell what part of the model is responsible in case of failure to reproduce observed data. The problem of model quality is dealt with in another contribution (Van Kraalingen, *this issue*), so that here it may suffice to note three major types of causes for model failure:

a) Technical programming errors
b) Incorrect scientific hypotheses
c) Incorrect model structure

With the further development of simulation languages, the frequency of technical programming errors will be reduced. Also the occurrence of type b errors will be reduced, since referees of scientific papers on modelling will usually scrutinize the underlying theories and equations.

The real problem is the type c modelling errors. Errors of this type are not visible in the scientific model description and easily escape external detection. It is very well possible that each individual model component is technically and scientifically correct, whereas the model as a whole is wrong. Such a situation could be compared to the "eternal stairs" drawing of Escher. Each part of the drawing of the stairs is physically sound, but somewhere along the route a change in meaning occurs that causes the whole picture to become absurd. An output variable of one model segment may be used in another segment in a meaning that is not exactly the same. Such errors easily escape the attention, even upon close inspection of the source code. It is for this reason that the code of trustworthy simulation models must be publicly available so that any critical individual has the possibility to read, see and check.

In case of model failure, it would be unwise to throw the model completely away. The major part of the model may still be correct, and only a minor refurbishing could perhaps result in correct model behavior. What should be done cannot be advised in general and must be based on expert judgement.

2.4 Models in teaching

Models that are used in teaching should be scientifically sound, or at least be temporarily acceptable, in order to convey a reasonable idea of how reality works. This means that the teaching model should be scientifically undisputed, and have disappeared from the forefront of the scientific struggle. Its role in science has ended, and it can only play a role in leading the students towards the scientific frontier of knowledge where other models (perhaps more sophisticated) are still subject of discussion and tests of falsification.

Clear examples of this situation in crop physiology are the theory of efficiency of growth respiration, developed by Penning de Vries, and the theory of energy limitation of evapotranspiration, developed by Penman. Their implementation and parameterization itself has certainly not finished and is still subject of research, but their structural validity is not debated anymore. These are well-established theories and are now part of the larger standard models for plant growth.

It is not necessary that the student him or herself is capable of model building in order to benefit from models when learning about the subject matter. In fact, with the increasing professionalization of computer-aided instruction this capability is becoming less and less a requirement. The user-friendly menu will guide the student through the material without revealing the internal structure of the underlying model. The computer-aided practicum has the large advantage over the real-world experimental practicum of being fast, to the point and reliable. It is also much more cost-efficient. The dangerous other side of the coin is the fact that computer simulation is not the real world. From an education point of view, I believe it is mandatory that students also experience reality as being always more complex, with unexplained variability and with unforeseen problems.

Yet, even these aspects of nature can be simulated to some extent if the designers of the practicum will be able to create a simulation environment that is not deterministic, but that has stochastic features. Troublesome events may add to the educational value of the simulation practicum.

2.5 Life stages

In terms of stages of life of models (Penning de Vries, 1982), a model begins its life as a speculative scientific model ("explorative model"). In this stage it can be around for quite some time, it may then be discarded as being untenable, and it may thus end its life. Many models have probably gone this path and are now lost and forgotten. Another scenario is that the model is not exactly wrong, but it has been overtaken by a more popular model version. It happens more often that parts of a model are extracted and incorporated in other models, so that it finally becomes impossible to tell what part belonged originally to which model. The pedigree of the crop growth models (Bouman et al.,1996) gives many examples of this kind of development.

A full-fledged and working model ("comprehensive model") will normally consist of many components of different origin. The problem of validity of the model has shifted from the validity of the individual components to the model construction as a whole. The size and complexity of a comprehensive model may necessitate the development of a summary model, in a way as a model of a model. This transition may occur together with a shift in application area from a research model to a policy and management model.

3. The end of modelling?

3.1 How do models end?

Models may end their application in different ways:

a) In science, when they have become untenable, being falsified as a theory. This is rarely the case, as models are mostly composed of different hypotheses. Even if one building block (a hypothesis) is altered, the model use can still be continued. However, it is possible that a model is so strongly centered around one single theorem, that it will mean the end of the model if this theorem is falsified. More often a model simply will become obsolete, being no longer a challenge. The life of the model in science will then end, but not necessarily in management or in policy.

b) In teaching. Largely the same situation occurs as in science, but probably with a delay. Also a model will become pedagogically obsolete, being overtaken by more advanced and more fancy implementations.

c) In policy and management a model may have served its purpose and be abandoned after the problem is solved. Paradoxically a model may be kept in use if it does not do its job for the full 100% (very similar to the services of a medical doctor). The economic models have been useful in preventing a deep economic depression, but could not properly help to solve the problem of growing unemployment.

3.2 Will modelling end?

Modelling defined as "the development of new models" will only come to an end when scientific progress ceases. Of course, whether this will happen and when this will happen is not possible to predict. It seems to me that simulation modelling has become so strongly united with the scientific method that a solitary termination of simulation modelling, as separate from science as a

whole, is unthinkable. In contrast, a worldwide dwindling of funds for scientific research is not unthinkable, and may well result in serious damage to science. There may be a growing mood in politics and public "that we know enough" and that further scientific research is a luxury rather than a need. Such a development would seriously threaten modelling, but worse than that it would undermine the rational basic attitude on which our society is still founded. Modelling defined as "the use of models" may then still continue for a while, but the termination of the rationally structured human society as we know it would no longer be in the distant future.

4. References

Barrow, J.D., 1992. Pi in the sky. Penguin books, London.

Bouman, B.A.M., Van Keulen, H., Van Laar, H.H., Rabbinge, R. 1996. The 'School of de Wit' crop growth simulation models: a pedigree and historical overview. In press, Agricultural Systems.

Casimir, H.B.G., 1987. Waameming en visie. Meulenhoff Informatief, Amsterdam.

Casti, J.L., 1990. Searching for certainty. William Morrow & Co., New York, USA

Darwin, C.R. 1859 (reprint 1985). The origin of species. Penguin Classics, London, UK.

De Wit, C.T. 1968. Theorie en model. Inaugural address. Agricultural University Wageningen.

Dijksterhuis, E.J. 1950. De mechanisering van het wereldbeeld. Meulenhoff, Amsterdam.

Goudriaan, J., 1977. Crop micrometeorology: a simulation study. Pudoc, Wageningen.

Goudriaan, J., 1993. Model, schaal en aggregatieniveau. Inaugural address, Agricultural University Wageningen.

Koningsveld, H., 1976. Het verschijnsel wetenschap. Boom, Amsterdam

Van Kraalingen, D.W.G. 1996. Requirements for model integration. This volume.

Luyten, J., Hoefnagel, B. (Eds), 1995. Het oog van de wetenschapsfilosoof. Boom, Amsterdam.

Mayr, E. 1988. Towards a new philosophy of biology. Harvard University Press, Cambridge MA, USA

Nagel, E., 1961. The structure of science, New York, USA

Penning de Vries, F.W.T. 1982. Phases of development of models. *In*: Penning de Vries, F.W.T. & Van Laar, H.H. (Eds) Simulation of plant growth and crop production. Pudoc, Wageningen. pp. 20 - 24

Penning de Vries, F.W.T., Van Laar, H.H., Chardon, M.C.M. 1983. Bioenergetics of growth of seeds, fruits and storage organs. *In*: Potential productivity of field crops under different environments. IRRI Manila, pp 37 - 59.

Penrose, R., 1989. The emperor's new mind. Oxford University Press, Oxford.

Popper, K.R., 1959. The logic of scientific discovery. London, UK.

Popper, K.R., 1972. Objective knowledge. An evolutionary approach. Oxford, UK.