VIDENTE 1.1: a graphical user interface and decision support system for stochastic modelling of water table fluctuations at a single location

VIDENTE 1.1: a graphical user interface and decision support system for stochastic modelling of water table fluctuations at a single location

Includes documentation of the programs KALMAX, KALTFN, SSD and EMERALD and introductions to stochastic modelling

Second, revised, edition

Marc F.P. Bierkens
Wichertje A. Bron
Martin Knotters

Alterra-rapport 613

Alterra, Green World Research, Wageningen, 2002

ABSTRACT

Bierkens, M.F.P., W.A. Bron and M. Knotters, 2002. *VIDENTE 1.1: a graphical user interface and decision support system for stochastic modelling of water table fluctuations at a single location; Includes documentation of the programs KALMAX, KALTFN, SSD and EMERALD and introductions to stochastic modelling.* Wageningen, ALTERRA, Green World Research.. Alterra-rapport 613. 184 pp;. 62 figs.; 5 tables; 70 refs.

A description is given of the program VIDENTE. VIDENTE contains a decision support system (DSS) to choose between different models for stochastic modelling of water table depths and a graphical user interface (GUI) to facilitate operating and running four implemented models: KALMAX, KALTFN, SSDS and EMERALD. In self contained parts each of these models is described. These descriptions also include an introduction to the practice of stochastic modelling. As each of the models can also be used in a stand alone fashion, i.e. outside VIDENTE, input formats for each model are also given.

Keywords: groundwater, time series, stochastic, Kalman filter, DSS, GUI, prediction, simulation

ISSN 1566-7197

This report can be ordered by paying € 27,- into bank account number 36 70 54 612 in the name of Alterra, Wageningen, the Netherlands, with reference to Alterra-rapport 613. This amount is inclusive of VAT and postage.

© 2002 Alterra, Green World Research,
P.O. Box 47, NL-6700 AA Wageningen (The Netherlands).
Phone: +31 317 474700; fax: +31 317 419000; e-mail: info@alterra.nl

No part of this publication may be reproduced or published in any form or by any means, or stored in a data base or retrieval system, without the written permission of Alterra.

Alterra assumes no liability for any losses resulting from the use of this document.

Project 395-11963 [Alterra-rapport 613/EvL/11-2002]

Contents

Preface 7

Preface to the second, revised edition 9

General introduction 11

Part 1: Decision Support System and Graphical User Interface 15

1 Explanation of questions and results of the DSS 17
1.1 General 17
1.2 Results of the DSS 18
1.3 Questions asked in the DSS 19

2 Verklaring vragen en uitkomsten BOS 25
2.1 Algemeen 25
2.2 Uitkomsten 26
2.3 Vragen 27

3 Graphical user interface 33
3.1 Set up 33
3.2 Input files 34
3.3 Example run using KALMAX 36

References 49

Annex: Binary decision trees of the DSS 51

Part 2: Documentation KALMAX and KALTFN 63

1 Model, input variables and parameters 65
1.1 Model description 65
1.2 Input variables and parameters/invoervariabelen en parameters 67

2 Stochastic modelling: prediction, simulation and calibration 73
2.1 Prediction 73
2.2 Stochastic modelling 78
2.3 Calibration 79
2.4 Systematic application of KALMAX and KALFTN: step by step 80
2.4.1 Calibration 81
2.4.2 Verification: prediction without Kalman filter 82
2.4.3 Simulation 83
2.4.4 Estimating fluctuation quantities 83
2.4.5 On line predection monitoring 87

References 89

Annex: Input instructions for KALMAX and KALTFN 91

Part 3: Documentation SSD 97

1 Model, input variables and parameters 99
1.1 Model description 99
1.2 Input variables and parameters/invoervariabelen en parameters 102

2. Stochastic modelling: prediction, simulation and calibration 109
2.1 Prediction 109
2.2 Stochastic modelling 114
2.3 Calibration 116
2.4 Systematic application of KALMAX and KALFTN: step by step 117
2.4.1 Calibration 117
2.4.2 Verification: prediction without Kalman filter 118
2.4.3 Simulation 119
2.4.4 Estimating fluctuation quantities 120
2.4.5 On line predection monitoring 123

References 125

Annex: Input instructions for SSD 127

Part 4: Documentation EMERALD 143

1. Model, input variables and parameters 145
1.1 Model description 145
1.2 Input variables and parameters/invoervariabelen en parameters 151

2. Stochastic modelling: prediction, simulation and calibration 157
2.1 Prediction 157
2.2 Stochastic modelling 159
2.3 Calibration 160
2.4 Systematic application of KALMAX and KALFTN: step by step 162
2.4.1 Calibration 162
2.4.2 Verification: prediction without Kalman filter 164
2.4.3 Simulation 164
2.4.4 Estimating fluctuation quantities 165
2.4.5 On line predection monitoring 168

References 169

Annex: Input instructions for EMERALD 171

Part 5: Input instructions for STATSIM 185

Preface

During almost six years of research several methods were developed for stochastic modelling of the water table depth at a single location. These methods have different levels of complexity, physical basis, data requirements and output variables. However, a rational method to choose between these methods is lacking. Also, because most of the computer programs to apply the stochastic methods were developed in a research environment, no particular attention was paid to making them user friendly. Therefore, a combined graphical user interface (GUI) and decision support system (DSS) has now been developed to apply the stochastic methods and to operate the associated computer programs. The GUI/DSS program is called VIDENTE. Vidente is the Spanish word for "seer" or "soothsayer" (*ziener* in Dutch) which translated into old Greek is *stocastichV* (a person who forecasts a future event in the sense of aiming at the truth). In the modern sense, "stochastic" in stochastic methods refers to the random element incorporated in these methods, i.e. to account for uncertainty. By taking uncertainty into account we have no hope to predict the future water table depth better than deterministic methods. We are however better in estimating how far off our predictions are going to be.

The research that lead to the development of the stochastic methods and their programs was funded by DWK programs 228 and 328 (Spatial Patterns and Variability in Soil and Groundwater) of the Dutch Ministry of Agriculture, Nature Management and Fisheries. Martin Knotters and Dennis Walvoort have contributed significantly to the development of the models implemented in VIDENTE. The computer program EMERALD was developed entirely by Dennis Walvoort.

MB
Wageningen, October 5, 2000

Preface to the second, revised edition

Since Vidente 1.0 was presented about two years ago, it proved to be a useful tool in various studies on the dynamics of the water table depth. The present version, Vidente 1.1, fulfils the wishes for improvement which raised during the first applications of Vidente in practice. The realisation of Vidente 1.1 was financially supported by Research Programme 395 of the Dutch Ministry of Agriculture, Nature Management and Fisheries. In this project I fruitfully co-operated with Tonny Otjens, Wichertje Bron and Jan Wesseling of Wageningen Software Labs (W!SL). Wichertje and Jan worked on the software. I am grateful to my Alterra-colleagues Peter Jansen, Jack van der Horst and Tom Hoogland. They critically tested the performance of Vidente 1.0 in practice, and they initiated the opportunities for further improvements which resulted in the present version, Vidente 1.1.

Martin Knotters, November 2002.

General introduction

To model the fluctuation of shallow water tables at a single location, several stochastic models have been developed at Alterra. The primary use of these models is to extrapolate time series of water table depth in time in order to estimate fluctuation characteristics of the water table depth (e.g. mean highest (MHW) and mean lowest water table depth (MLW)) that are representative for the current climate and hydrological conditions (Knotters and Van Walsum, 1997). Maps of these fluctuation quantities are vitally important for land use planning and regional water management (e.g. Finke *et al*., 1999). More recently, some of these models have been used in a regionalised fashion for space time modelling of water table depth (e.g. Bierkens et al., 2001).

The advantage of using stochastic models rather than deterministic models is that stochastic models are better suited to capture extreme value statistics (such as MHW and MLW), provide more realistically looking time series and can be used for risk analysis: it is possible to estimate the probability that at a given date the water table exceeds a critical value. When combined with costs models, the "expected cost of exceeding" (=risk) can be estimated, which makes it possible to weigh the costs of water management measures against the benefits of risk reduction (=cost of failure). It is likely that risk based cost-benefit analyses will become more important in solving environmental problems and water management under multiple interests.

The common factor between all models is that they describe the fluctuation of the water table depth as a function of time and that they have a deterministic and a stochastic component (noise model). The difference between the various models is based on the amount of physics used to model the deterministic component, the amount of information required to run the models and the number of variables the model is able to generate (i.e. only water table depth or other water balance parameters also). The models have been developed by different people and have been implemented in different computer programs. People using these models therefore raised two questions:
How do I choose between the various models for my specific application?
How do I operate all these different computer programs that have different input an output formats?
To solve both programs simultaneously we have build the Windows application VIDENTE. VIDENTE is programmed in Delphi 5 and consists of a Decision Support System (DSS) to choose between the various models and a Graphical User Interface (GUI) to operate the different programs from similar input and output screens. Parts of this report have been implemented to serve as a help facility. The DSS leads the user through a series of questions to the most appropriate model for his or her problem. Next the GUI can be used to constitute the model input, run the model and analyse the model output.

In VIDENTE four different models have been implemented. Their names, which are in fact the names of the computer programs, are given below as well as key references in English and Dutch that describe the models and their application:

1. **KALMAX** – an auto-regressive exogenous variable model
2. **KALTFN** – a simple transfer-function noise model

   Key references of these two models:

   English

   Bierkens, M.F.P., M. Knotters and F.C. Van Geer, 1999. Calibration of transfer function-noise models to sparsely or irregularly observed time series. *Water Resources Research* 35(6), 1741-1750.

   Knotters, M. and M.F.P. Bierkens, 2000. Physical basis of time series models for water table depths. *Water Resources Research* 36(1), 181-188.

   Dutch

   Bierkens, M.F.P, M. Knotters and F.C. Van Geer, 1999. Tijdreeksanalyse nu ook toepasbaar bij onregelmatige meetfrequenties. *Stromingen* 5(2), 43-54.

   Knotters, M. and M.F.P. Bierkens, 1999. Tijdreeksmodellen voor de grondwaterstand; een kijkje in de black box. *Stromingen* 5(3), 35-49.
3. **SSD** – a stochastic differential equation of the soil water balance

   Key references:

   English

   Bierkens, M.F.P., 1998. Modeling water table fluctuations by means of a stochastic differential equation. *Water Resources Research* 34(11), 2485-2499.

   Dutch

   Bierkens, M.F.P., 1988. Eenvoudige stochastische modellen voor grondwater-standsfluctuaties. Deel 1: Een stochastische differentiaalvergelijking. *Stromingen* 4(2), 5-26.
4. **EMERALD** – A physically based stochastic model of soil- and groundwater flow

   Key references:

   English

   Walvoort, D.J.J. and M.F.P. Bierkens, 1999. A stochastic modelling approach for rapid assessment of groundwater dynamics. Report 171, DLO Staring Centrum, Wageningen.

   Dutch

   Bierkens, M.F.P. en D.J.J. Walvoort, 1998. Eenvoudige stochastische modellen voor grondwaterstandsfluctuaties. Deel 2: Gecombineerd bodem-grondwatermodel met stochastische invoer. *Stromingen* 4(3), 5-20.

Additional references on the application of the KALMAX and SSD models for space-time modelling of water table depth are Bierkens (2001), Bierkens et al. (2001), Knotters and Bierkens (2001) an Knotters (2001), which can be found in the reference list.

The following model has also been developed together with the models above, and will be implemented in VIDENTE in the near future:

5. **TARSO** – a non-linear transfer function-noise model with threshold non-linearity

   Key reference only in English:

Knotters, M. and J.G. de Gooijer, 1999. TARSO modeling of water table depths. *Water Resources Research* 35(3), 695-705.

The following model has not been developed within our group but elsewhere at Alterra (in combination with Wageningen University). When combined with a noise model it is also suited for stochastic modelling of water table depth:

6. **SWAP** – A physically-based model of the soil-water-atmosphere-plant system at a single location, including crop growth and water, heat and solute transport through the soil.

   Under certain circumstances the DSS will point to using SWAP. The program is not implemented in VIDENTE but a 32bit Windows version is available from http://www.alterra.wageningen-ur.nl/fset-onderzoek.htm (price $ 500). The key reference (only in English) is:

   Dam, J.C. van, J. Huygen, J.G. Wesseling, R.A. Feddes, P. Kabat, P.E.V. Van Walsum, P. Groenendijk and C.A. Van Diepen, 1997. Theory of SWAP version 2.0; simulation of water flow, solute transport and plant growth in the soil-water-atmosphere-plant environment. Report 171, Department of water Resources, Wageningen Agricultural University, Wageningen. Technical Document 45, DLO Winand Staring Centre, Wageningen.

Finally, it is possible that the DSS will point to modelling techniques that are available outside Alterra, such as univariate time series models implemented in generals statistical or mathematical packages as MATLAB, SPLUS or GENSTAT.

This report consists of five parts. Part 1 describes the DSS and the GUI. The possible results of the DSS are explained as well as the questions users have to answer while running the DSS. The DSS is described both in English and in Dutch. The binary decision trees upon which the DSS is based are given in the annex. The GUI is mostly self explanatory. We only describe the global set up of the screens. For illustration, KALMAX is applied to an example data and the input and output screen are shown. In parts 2 to 4 the models are described: KALMAX/KALTFN (part 2), SSD (part 3), EMERALD (part 4). Each of these parts is completely self contained. This means that these parts can be read separately from the rest. It also means that there is much redundancy between these parts. For instance, each part provides an introduction to stochastic modelling (chapter 2). Because it is also possible to use each model in a stand alone fashion, we also give the input instructions (formats) for the computer programs of each model. Part 5 gives the input instructions for the program STATSIM that can be used to calculate fluctuation quantities from simulated water table depths.

Parts of this report that contain crucial information to operate VIDENTE have also been written in Dutch. These are the description of the DSS and the lists of parameters of each of the models.

This report comes with a CD. The CD contains the program VIDENTE, as well as the subdirectory “standalone” with sources and executables of KALMAX, KALTFN, SSD, and EMERALD and example input and parameter files. To run EMERALD the screen resolution should be 600×800. The programs on this CD are

distributed in the hope that they will be useful, but WITHOUT ANY WARRANTY. No author or distributor accepts responsibility to anyone for the consequences of using them or for whether they serve any particular purpose or work at all, unless he or she says so in writing. Altering or redistribution of the software should be done while giving proper reference and in accordance with the terms and conditions mentioned in the GENERAL TERMS AND CONDITIONS AS TO THE MAKING AVAILABLE OF COMPUTER SOFTWARE issued by Alterra (see Word document on the CD).

Part 1: Decision Support System and Graphical User Interface

1 Explanation of questions and results of the DSS

1.1 General

The computer program VIDENTE is a graphical user interface (GUI) driving a number of other computer programs. These programs contain methods for stochastic modelling of water table depth fluctuations (phreatic surface) at a single location.

The decision support system (DSS) chooses between the various methods based on the following criteria:

1. Which are the target variables? Is the water table depth the only variable of interest or are additional variables of the soil-groundwater system also important, such as soil moisture content, specific discharge to drains and ditches and crop transpiration?
2. How is the soil-groundwater system to be modelled? For instance, are more drainage levels to be included, are surface water levels varying with time.
3. Limitations to the applicability of methods. For instance, the model implemented in the program SSD cannot be applied to deep water tables.
4. Availability of data. For instance, to apply SSD at a location we need to know the soil moisture retention curve and the drainage levels with respect to surface elevation.

Based upon the answers to yes/no questions pertaining to one of the categories above binary decision trees are built that eventually lead to one of the available methods for stochastic modelling of water table depth (see Annex). Such a method is a mathematical model that is implemented in a computer program. The computer programs are driven by the GUI. The philosophy that it is best to use the most simple method able to solve the problem at hand (parsimony) is implicit in the DSS.

Notice that the accuracy of the various methods or models is not included as a criterion in the GUI. Implicitly the accuracy plays a role in criterion 3, where models are limited to a domain in which they are expected to perform sufficiently accurate. In case that several models are applicable, no relevant differences in accuracy are found. This has been corroborated by several studies comparing the accuracy of transfer function-noise models (TFN models), KALMAX, KALTFN, TARSO, SSD, EMERALD and SWAP (Knotters and Van Walsum, 1997; Bierkens, 1998; Walvoort and Bierkens, 1998; Knotters and De Gooijer, 1999; Bierkens et al., 1999).

1.2 Results of the DSS

The following computer programs are included in VIDENTE (for an extensive description of the implemented methods one is referred to parts 2 to 4):

1. KALMAX (ARX model, precipitation surplus as input : precipitation minus Makkink reference evapo-transpiration (Winter *et al*., 1995), physical interpretation possible);
2. KALTFN (a simple low order TFN model, precipitation surplus as input, physical interpretation possible);
3. SSD (a physical model based on a stochastic differential equation of water table depth derived from the soil-water balance, precipitation and potential evapo-transpiration as input);
4. EMERALD (a quasi-analytical soil-groundwater model with added noise, precipitation and potential evapo-transpiration as input).

Apart from these methods, the DSS may direct the user to one of the following methods that are not implemented in VIDENTE, but elsewhere:

5. TARSO (non-linear time series model with threshold non-linearity, in each regime a simple TFN model, precipitation surplus as input). Although this model has been developed in our research group, it has not been implemented yet. The goal is to do this in the near future;
6. Univariate time series analysis. In this case there is no input time series to explain part of the variation of the observed water table depth. In univariate time series analysis (ARIMA, SETAR (Tong, 1990; Hipel and McLeod, 1994)) the variation of a time series is explained from the values of the time series at previous time steps and a noise process. These models can be used for short time forecasts, interpolation between observations and simulating replicas of the time series for the observation period. However, these methods are not suitable for extrapolating time series of water table depth (e.g. to the past) to correct for the year by year variation of the weather (precipitation surplus) (Knotters and Van Walsum, 1997). Univariate time series models are implemented in GENSTAT, MATLAB and S-PLUS;
7. (Multiple) TFN modelling. For standard TFN modelling the observation frequency of the input time series (e.g. precipitation surplus) have to be the same as for the output time series (e.g. water table depth). For KALMAX, KALTFN, SSD, EMERALD and SWAP this does not have to be the case. However, multiple TFN models can be used to account for more input variables if these are thought to influence variation of water table depth; e.g. river stages and groundwater abstraction rates. Multiple TFN models are implemented in GENSTAT and MATLAB.
8. SWAP. SWAP (van Dam et al., 1997) is a computer program for modelling the Soil-Water-Atmosphere-Plant system. Flow through the unsaturated zone is modelled by numerically solving Richards' equation. SWAP not only provides the water table depth, but also the soil moisture profile with time. Additionally, transport (heat and inert solvents) can be modelled. Also, crop transpiration and evapo-transpiration is modelled in a more sophisticated manner than in SSD and EMERALD. It is even possible to model crop growth using the WOFOST

model (Supit et al., 1994). SWAP is written in FORTRAN77, but a 32bit Windows version is available from http://www.alterra.wageningen-ur.nl/fset-onderzoek.htm. Costs are about $ 500. Of course, application of SWAP requires more information about a site than using SSD or EMERALD. Apart from SWAP many other codes for unsaturated flow are available. However, very few of these are able to model phreatic surface, which is why they are not mentioned here.

One outcome of the DSS may be that none of the above methods is suitable, for instance because certain data are lacking. In particular the following messages can be expected:

1. "Stochastic modelling of the water table depth or related variables is not possible". In this case a time series of water table depth is lacking. This time series should first be recorded, or when a piezometer is present, its data should be achieved from its owner.
2. "What you want is only possible if the necessary information is collected first". Here we have that based on criteria 1 to 3 above the outcome of the DSS is some method or set of methods, but, apart from the water table depth, the necessary data are lacking to apply these methods. If at all possible, an alternative method is proposed. In most cases this will amount to neglecting certain effects (such as varying surface water levels) or limiting the target variables (e.g. water table depth only instead of water table depth and soil moisture content).

Finally, an outcome may be that multiple models are equally suitable for the application at hand. In this case the user has to make his or her own decision.

1.3 Questions asked in the DSS

What are the target variables?
If the water table depth is the only target variable all methods introduced before can be used. If next to the water table depth also the specific discharge or bottom flux (or drainage resistance or effective porosity) is important, one should use KALMAX, KALTFN, SSD, EMERALD or SWAP. If apart from those mentioned the average soil moisture content is a target variable either SSD, EMERALD or SWAP can be used. Finally, if apart from these variables time series of the soil moisture profile, crop growth, soil firmness (capacity to carry machinery) and transport (heat or inert solutes) are required, SWAP is the only option.

Do you have a time series of water table depth?
Without an observed time series of water table depth stochastic modelling of the water table depth or related variables is not possible. This time series is necessary to identify the noise process. For the empirical models such as ARIMA models, KALMAX, KALTFN and TARSO, all parameters are calibrated using observed water table depths, while for the more physical models such as SSD, EMERALD and SWAP a number of conceptual parameters (e.g. effective porosity and drainage resistance) are obtained through calibration.

Do you have time series of precipitation and evapo-transpiration?
Without time series of precipitation and evapo-transpiration (or precipitation surplus) it is not possible to model the water table fluctuations with TFN models, KALMAX, KALTFN, TARSO, SSD, EMERALD or SWAP. In this case, univariate time series analysis (e.g. ARIMA) can be an option. Note however that univariate time series models are not suitable for extrapolating the time series of water table depth to correct for the year to year variation of precipitation surplus.

Are additional temporally variable quantities likely to influence water table fluctuations?
Sometimes the variation of the water table depth not only depends on the precipitation surplus but also on other temporally varying quantities such as river stages or groundwater withdrawals. If this is the case this should be accounted for explicitly. Whether this is possible depends on the presence of time series of these additional input variables.

Do you have time series of additional input variables?
If, apart from precipitation surplus, the water table fluctuations are driven by other variables, but no time series of these variables have been observed, one is forced to account for these influences in an implicit manner. The influence of an extra input variable can manifest itself partly as a trend and partly as additional fluctuation. The trend can be accounted for by trend fitting, removing the fitted trend from the time series of water table depth, conducting the analyses on the de-trended time series and adding the trend afterwards (note that superposition is applied which assumes a linear relation between the water table depth and the additional input variable). The additional fluctuation cannot be accounted for and will lead to a larger noise component (more uncertainty). If time series of the additional input variables are present they can be included using multiple TFN models.

Has the time series of water table depth been observed with constant frequency?
Is the time interval between two successive observations the same for the entire time series or at least approximately so? If this is not the case, standard TFN modelling and TARSO cannot be applied.

Do you want to model the water table depth with smaller time steps than the time interval between observations?
Even if the time series of water table depth is observed with a constant frequency one may be interested in describing the water table depth with smaller time steps than the interval between observations. In this case standard TFN modelling and TARSO cannot be applied.

Are there any threshold non-linearities and do you wish to take these into account?
Examples of thresholds that divide the domain of water table fluctuation into parts where the dynamics are very different: boundaries between soil layers of contrasting texture and drainage levels of trenches. In the latter case, the trench drains the groundwater whenever its level exceeds the trench bottom and is inactive otherwise.

Has the time series of precipitation surplus been observed with constant frequency and is the interval between its observations smaller than or equal to the smallest interval between the observations of water table depth?

Using KALMAX, KALTFN, SSD, EMERALD or SWAP it is possible to model the water table depth fluctuations with a smaller time step than interval between observations of the water table depth. Moreover, it is not even required that the time series of water table depth is observed with constant frequency. However, to do this we require the time series of precipitation surplus to be observed with constant frequency and with an interval between its observations smaller than or equal to the smallest interval between the water table depth observations. If for certain periods the water table depth has been observed with a higher frequency than the precipitation surplus, the water table observations should be aggregated to the frequency of the precipitation surplus.

Do you want to take non-linearity into account?

The relation between water table depth and precipitation surplus is in principal non-linear. This can be caused by a trench that becomes inactive whenever the water table drops below its bottom, soil heterogeneity and variations in soil moisture. If you want to take such non-linearity into account (in this part of the decision tree) SSD, EMERALD or SWAP should be used.

Apart from the water table depth, does the list of target variables also include specific drainage discharge, bottom flux (or drainage resistance or effective porosity)?

If KALMAX or KALTFN is chosen a further choice can be made to interpret the deterministic part of these models in terms of physical parameters. The deterministic part of KALMAX and KALTFN has the following form (with h_t the water table depth and P_t precipitation surplus (precipitation minus evapo-transpiration) at time step t):

$$h_t = c + a\left[h_{t-1} - c\right] + bP_t$$

If the drainage level h_s (surface water level or trench bottom with respect to surface level) is constant (see next question) en known (see question thereafter), the drainage resistance g [T], the effective porosity f [-] and the bottom flux q_b [LT^{-1}] can be estimated from the parameters a,b and c and time step Δt used (Knotters and Bierkens, 2000):

$$g = \frac{b}{1-a} \qquad f = \frac{-\Delta t}{g \ln a} \qquad q_b = \frac{c - h_s}{g}$$

This is called the physical interpretation of KALMAX and KALTFN.

Do you want to take account of more than one drainage level or drainage levels that vary with time?

In principal one could give a physical interpretation to KALMAX and KALTFN when several drainage levels are present (if these are ditches that don't become inactive when the water table drops below their bottom). However, in this case it is

advised to use either SDD or SWAP. If the drainage levels are varying with time (e.g. varying surface water levels) SSD or SWAP should be used.

Is the drainage level (bottom trench or surface water level) with respect to surface elevation known?
As shown in the equations above, a physical interpretation of KALMAX and KALTFN requires the drainage level with respect to surface elevation.

Apart from the precipitation surplus, are there additional variables that influence the water table fluctuation?
This question is to decide between KALMAX and KALTFN. If the precipitation surplus is the only input variable that influences the water table fluctuation, it can be derived that the temporal dependence of the noise process is the same as that of the (deterministic) water table depth (Knotters and Bierkens, 2000). In this case KALMAX is preferred. If other variables such as river stages are influencing the water table depth, and these variables have not been explicitly taken into account (otherwise the choice would not have been between KALMAX or KALTN), these influences end up in the noise so that the noise is likely to have a different correlation structure than the water table depth. In this case it is more appropriate to use KALTFN.

Is the water table fluctuating at depths larger than 2 m below the surface?
In these type of systems the infiltrating rain water will show a noticeable time shift and a dampening as it travels through the large unsaturated zone to the phreatic surface. This effect cannot be modelled with SSD so that EMERALD en SWAP are the remaining options.

Do you want to model the reduction of evapo-transpiration in detail?
The reduction in evapo-transpiration is modelled rather simply in SSD and EMERALD, i.e. as a loss function that depends on the average soil moisture content. If you want the reduction to depend on the root distribution and the soil moisture profile, to use a loss function that varies with time due to crop growth or to divide evapo-transpiration into soil evaporation and crop transpiration SWAP should be used.

Do you have information about unsaturated conductivity per soil layer?
When applying EMERALD or SWAP one needs information about the unsaturated conductivity function per soil layer. This information can either be obtained directly by taking laboratory measurements, or standard curves can be applied by linking these to texture using pedo-transfer functions (Wösten and Van Genuchten, 1988; Wösten, 1997). In this case we need to have a texture profile description of the soil.

Do you have the following data: land use, soil moisture retention curve, drainage level?
To apply SSD, EMERALD or SWAP one needs (among other things) data about the land use. Based upon land use the actual evapo-transpiration can be calculated from the reference evapo-transpiration (e.g. using crop factors) and rainfall interception can be determined. One is referred to the descriptions of EMERALD and SSD for a table with interception factors and crop factors for various types of land use (see also

the HELP capability of VIDENTE). Like the unsaturated conductivity $k(y)$, the soil moisture retention curve $q(y)$ can be obtained from laboratory measurements or indirectly from texture descriptions and pedo-transfer functions (Wösten and Van Genuchten, 1988; Wösten, 1997). Finally, the drainage level(s) with respect to surface elevation are required.

Are most of the water table depths of the time series smaller than 50 cm?
EMERALD is not very suitable for modelling very shallow water table depths. In that case either SSD or SWAP should be chosen.

Do you have the following data: land use, soil moisture retention curve, unsaturated conductivity, drainage level?
To apply SWAP one needs (among other things) data about the land use. Based upon land use rainfall interception, actual evapo-transpiration and, if needed, crop growth can be calculated. The unsaturated conductivity $k(y)$ and the soil moisture retention curve $q(y)$ are needed for each soil layer. These can be obtained from laboratory measurements or indirectly from texture descriptions and pedo-transfer functions (Wösten and Van Genuchten, 1988; Wösten, 1997). Drainage level(s) with respect to surface elevation are also required.

2 Verklaring vragen en uitkomsten BOS

2.1 Algemeen

Het computerprogramma VIDENTE is een grafische user interface rond een aantal computerprogramma's met verschillende methoden voor de stochastische modellering van de (freatische) grondwaterstandsfluctuatie op een locatie.

Het beslissingsondersteunend systeem (BOS) selecteert tussen de verschillende methoden op basis van de volgende criteria:

1. Wat is de doelvariabele? Is dit alleen de grondwaterstanden, of zijn ook andere variabelen die betrekking hebben op het bodemwater-grondwater systeem van belang, zoals bodemvochtgehalte, gewasverdamping en drainageafvoer?
2. Op welke wijze dient het bodemwater-grondwater systeem worden gemodelleerd? Bijvoorbeeld, moeten er meerdere drainageniveaus worden gemodelleerd, of dienen de slootpeilen in de tijd te kunnen variëren?
3. Beperkingen die aan de toepasbaarheid van de methode zitten. Bijvoorbeeld, het model dat is geïmplementeerd in het computerprogramma SSD kan niet worden toegepast op te diepe grondwaterstanden.
4. De beschikbaarheid van de gegevens. Bijvoorbeeld, om SSD op een locatie toe te passen moet daar informatie over de pfcurve beschikbaar zijn.

Op basis van ja/nee vragen die onder één van de bovenstaande categorieën vallen wordt een binaire determinatieboom opgebouwd die uiteindelijk leidt tot de keuze voor één van de methoden om de grondwaterstandsvariatie te modelleren. Een methode bestaat uit een model dat de variatie van de grondwaterstand beschrijft en is geïmplementeerd in een computerprogramma. De computerprogramma's kunnen worden bedient via de grafische *user interface* (GUI). Impliciet is bij het BOS als filosofie gekozen dat een gegeven probleem met een zo simpel mogelijke methode die voldoet moet worden opgelost.

Opvallend is dat in het BOS de nauwkeurigheid van de modellen geen rol speelt. Impliciet zit dit reeds in criterium 3, waarin modellen worden beperkt tot een toepassingsgebied waarbinnen, op basis van de modelveronderstellingen, voldoende nauwkeurige resultaten kunnen worden verwacht. In het geval dat meerdere modellen toepasbaar zijn blijkt uit meerdere studies (Knotters en Van Walsum, 1994; Bierkens, 1998; Bierkens en Walvoort, 1998; Knotters en De Gooijer, 1999; Bierkens *et al.*, 1999) dat er geen noemenswaardige verschillen in nauwkeurigheid optreden tussen transfer-ruismodellen, KALMAX, KALTFN, TARSO, SSD, EMERALD en SWAP. De nauwkeurigheid van de modellen is dus niet in de beslisregels meegenomen.

2.2 Uitkomsten

De volgende computerprogramma's zijn in VIDENTE opgenomen (voor een uitgebreide beschrijving van de methoden wordt verwezen naar de HELP optie van VIDENTE):

1. KALMAX (ARX model, neerslagoverschot als invoer, fysische interpretatie mogelijk);
2. KALTFN (een eenvoudig, lage orde transfer-ruismodel, neerslagoverschot als invoer, fysische interpretatie mogelijk);
3. SSD (een fysisch model, gebaseerd op een stochastische differentiaalvergelijking van de grondwaterstand, afgeleid op basis van bodemwaterbalans, neerslag en verdamping als invoer);
4. EMERALD (een quasi-analytisch bodem-grondwatermodel met toegevoegd ruisproces, neerslag en verdamping als invoer).

Naast deze methoden is het ook mogelijk uit te komen bij de volgende methoden die niet in VIDENTE zijn geïmplementeerd:

5. TARSO (niet-lineair tijdreeksmodel met drempel niet-lineariteiten, in elk regime een eenvoudig transfer-ruismodel, neerslagoverschot als invoer). Hoewel TARSO wel is ontwikkeld binnen onze onderzoeksgroep, is dit model nog niet geïmplementeerd. Het is de bedoeling dit in de nabije toekomst te doen.
6. Univariate tijdreeksanalyse. In dit geval is er geen neerslagoverschot of andere invoerreeks om de variatie van de grondwaterstand te verklaren. Bij univariate tijdreeksanalyse (bijv. ARIMA, SETAR (Tong, 1990; Hipel en McLeod, 1994)) wordt de variatie van de reeks verklaard uit de waarden van de variabelen op vorige tijdstappen en een ruisproces. Met deze modellen kunnen korte termijnvoorspellingen worden gedaan, er kan worden geïnterpoleerd tussen waarnemingen, en voor de duur van de waarnemingsperiode kunnen replica's worden gesimuleerd (voor ontwerpdoeleinden). Het is met deze methoden echter niet mogelijk om een tijdreeks van grondwaterstanden te verlengen, bijv. naar het verleden, om op deze wijze weersvariaties uit de reeks te filteren (zie Knotters en Van Walsum, 1994). Univariate tijdreeksmodellen zijn geïmplementeerd in pakketten als GENSTAT, MATLAB en S-PLUS;
7. (Multiple) Transfer-ruismodellering. De frequenties van de invoerreeksen moeten hier gelijk zijn aan die van de uitvoerreeksen (i.t.t. KALMAX, KALTFN, SSD en EMERALD). Er kunnen echter wel multiple invoerreeksen worden gemodelleerd, bijvoorbeeld als het grondwaterstandsverloop niet alleen afhangt van het neerslagoverschot, maar ook van de standen in een nabijgelegen rivier of het onttrekkingsdebiet van een pompstation in de buurt. Pakketten waarin multiple transfer-ruismodellen zijn geïmplementeerd: GENSTAT, MATLAB;
8. SWAP. SWAP (Van Dam *et al.*, 1997) is een model van het bodem-water-plant-atmosfeer systeem. Eéndimensionale stroming door de onverzadigde zone wordt gemodelleerd door het numeriek oplossen van de Richards' vergelijking. Met SWAP kan daarom, behalve de grondwaterstand, ook het vochtprofiel worden berekend, alsmede het transport van warmte en inerte stoffen. Verder is in vergelijking met SSD en EMERALD de verdampingsreductie veel gedetailleerder te modelleren, met name de gewasverdamping. Het is zelfs mogelijk om in

SWAP gewasgroei te modelleren met het WOFOST model (Supit *et al.*, 1994). SWAP is geschreven in FORTRAN77, maar er is nu ook een Windows versie (kosten $500) die tegen betaling kan worden gedownload via http://www.alterra.wageningen-ur.nl/fset-onderzoek.htm. Vanzelfsprekend betekent het gebruik van SWAP dat nog meer informatie over de locatie nodig is dan gebruikt in SSD en EMERALD. Naast SWAP zijn er natuurlijk nog veel meer programma's waarmee de stroming van water door de onverzadigde zone op een locatie kan worden berekend. De meeste van die modelcodes berekenen echter geen vrije grondwaterspiegel, zodat ze hier niet verder worden genoemd.

Het resultaat van het doorlopen van het BOS kan zijn dat geen van bovenstaande modellen voldoet, omdat bepaalde benodigde gegevens ontbreken. In het bijzonder zijn de volgende twee meldingen te verwachten:

1. "Geen stochastische modellering van de grondwaterstanden of aanverwante variabelen mogelijk". In dat geval ontbreekt een tijdreeks van de grondwaterstand. Deze zal eerst moeten worden verzameld (een langere tijd regelmatig waarnemen) of, in het geval er reeds een grondwaterstandsbuis aanwezig is, moeten worden aangeschaft bij NITG-TNO;
2. "Wat u wilt is alleen mogelijk indien u de ontbrekende informatie verzamelt". In dat geval is de gebruiker op basis van criteria 1 t/m 3 bij een bepaalde methode uitgekomen en ontbreekt er andere informatie dan de grondwaterstand die nodig is om deze methode toe te passen. In dat geval wordt er, indien mogelijk, ook een alternatieve methode aangedragen. In de meeste gevallen zal dit alternatief betekenen dat bepaalde zaken moeten worden verwaarloosd (bijv. homogene in plaats van heterogene bodem) of dat men genoegen moet nemen met minder doelvariabelen (alleen de grondwaterstand i.p.v. grondwaterstand en vochtgehalte).

Tenslotte kan het zijn dat meerdere modellen net zo geschikt zijn voor het betreffende probleem. In dat geval zal de gebruiker zelf moeten kiezen. Deze keuze zal voornamelijk berusten op smaak.

2.3 Vragen

Wat zijn de doelvariabelen?
Als alleen de grondwaterstand de doelvariabele is, dan kunnen alle boven beschreven methoden worden toegepast. Is men daarnaast geïnteresseerd in specifieke afvoer, kwel en infiltratie en hydraulische parameters als drainageweerstand en effectieve porositeit, dan kunnen KALMAX, KALTFN, SSD, EMERALD en SWAP worden toegepast. Behoort het gemiddeld vochtgehalte van de bodem tot de doelvariabelen, dan voldoen alleen SSD, EMERALD en SWAP. In het geval men een tijdreeks van het gehele vochtprofiel wil berekenen of tijdreeksen van gewasgroei, draagkracht en transport (van warmte of inerte stoffen) kan alleen SWAP worden toegepast.

Is er een tijdreeks van de grondwaterstand beschikbaar?
Om een stochastisch model van de grondwaterstandsfluctuatie toe te passen moet er een tijdreeks van de grondwaterstand zijn. Alleen in dat geval kan men het stochastisch gedeelte van het model (het ruisproces) identificeren. Overigens geldt dat bij de meer empirische modellen (bijv. ARIMA, transfer-ruismodellen, KALMAX, KALTFN, TARSO) alle parameters uit waargenomen grondwaterstanden worden geschat, en bij de meer fysische modellen (SSD, EMERALD, SWAP) een aantal moeilijk te meten parameters zoals de drainageweerstand en de effectieve porositeit.

Is er een neerslag en verdampingsreeks?
Als er geen tijdreeksen van neerslag en verdamping (of het neerslagoverschot: neerslag minus verdamping) zijn dan kunnen transfer-ruismodellen, KALMAX, KALTFN, TARSO, SSD, EMERALD en SWAP niet worden toegepast. Univariate tijdreeksanalyse met bijvoorbeeld ARIMA modellen is nog wel een optie. Het probleem hierbij blijft dat extrapolatie ten behoeve van het corrigeren voor jaar tot jaar fluctuaties van het neerslagoverschot niet mogelijk is.

Zijn er naast het neerslagoverschot nog andere variabelen die de grondwaterstand kunnen beïnvloeden en die variëren in de tijd?
Als er sterke aanwijzingen zijn dat de fluctuaties van de grondwaterstand niet alleen verklaard worden door het neerslagoverschot, maar ook door andere in de tijd variërende variabelen zoals onttrekkingen of een rivierpeilen, dan is het verstandig om hiermee rekening te houden. Of dit feitelijk mogelijk is hangt af of er tijdreeksen van deze exetra invoervariabelen aanwezig zijn.

Zijn er van deze extra invoervariabelen tijdreeksen aanwezig?
Als er wel sprake is van de beïnvloeding van de grondwaterstand door een andere invoervariabele dan het neerslagoverschot, zoals bijv. een ontrekking, maar er zijn geen tijdreeksen daarvan, dan is men gedwongen om deze invoerreeks niet expliciet mee te nemen. De invloed van een extra invoervariabele kan zich doen gelden op twee manieren. Als een trend en als een extra fluctuatie. Het trendmatige gedeelte kan worden verdisconteerd door een trend te fitten en van de grondwaterstandstijdreeks af te trekken, de analyse op de residuen te doen en later de trend er weer bij op te tellen (let op: dit is superpositie en verondersteld dus een lineaire relatie tussen de grondwaterstand en de extra invoervariabele). De extra fluctuatie kan niet worden verdisconteerd en manifesteert zich dan ook door een grotere ruiscomponent (meer onzekerheid). Als tijdreeksen van extra invoervariabelen wel aanwezig zijn, dan kunnen die worden meegenomen via multiple transfer-ruismodellen.

Is de grondwaterstandstijdreeks equidistant?
De vraag is hier of de tijdsintervallen tussen twee waarnemingen constant zijn, of in ieder geval (zoals bij grondwaterstanden in Nederland) bij benadering constant. Is dat niet het geval dan kunnen standaard transfer-ruismodellen en TARSO niet worden toegepast.

Moeten de grondwaterstanden met een kleinere tijdstap dt worden gemodelleerd dan de tijdstap tussen de waarnemingen?
Zelfs als de tijdreeks van de grondwaterstand equidistant is kan men de wens hebben om de grondwaterstanden met een kleinere tijdstap te willen modelleren dan die is waargenomen. In dat geval kan men geen standaard transfer-ruismodellen en TARSO toepassen.

Is er sprake van drempel niet-lineariteiten, bijv. bodemfysische grenzen of drainageniveaus en wilt u hiermee rekening houden?
Typische voorbeelden van dergelijke drempels die het bereik van de grondwaterstandsflucuatie indelen in verschillende domeinen (de reactie op het neerslag-overschot in elk domein is anders) zijn contrasterende bodemlagen zoals klei op zand en het voorkomen van drainageniveaus die droogvallen als de grondwaterstand eronder komt (bijv. greppels).

Is de tijdreeks van het neerslagoverschot equidistant en is het interval ervan kleiner of gelijk aan het kleinste interval van de grondwaterstandsreeks?
Bij de toepassing van KALMAX, KALTFN, SSD, EMERALD en SWAP is het mogelijk met een kleinere tijdstap te rekenen dan de tijdstap van de grondwaterstandsreeks. De reeks grondwaterstanden hoeft zelfs niet equidistant te zijn. Wat we hiervoor wel nodig hebben is een equidistante tijdreeks van het neerslagoverschot met een tijdstap kleiner dan de kleinste tijdstap van de grondwaterstanden. In de Nederlandse praktijk zal dit meestal betekenen dat dagwaarnemingen van neerslag en verdamping nodig zijn. Als de grondwaterstand tijdens sommige perioden met een nog hogere frequentie dan het neerslagoverschot is gemeten, bijvoorbeeld elk uur met divermetingen, dan zullen die grondwaterstanden moeten worden uitgemiddeld tot de tijdstap van het neerslagoverschot. Grondwaterstandsfluctuaties van hogere frequentie dan de meetfrequentie van de reeks van het neerslagoverschot kunnen namelijk niet uit deze reeks worden verklaard.

Wilt u rekening houden met niet-lineariteit?
In het algemeen is de relatie tussen het neerslagoverschot en de grondwaterstand niet-lineair. Dit heeft te maken met bijvoorbeeld droogvallende waterlopen, contrasten tussen bodemlagen en variaties in het bodemvocht. Als u daarmee rekening wenst te houden moet u (op deze plek in de determinatieboom) SSD, EMERALD of SWAP gebruiken.

Zijn naast de grondwaterstand ook de drainageafvoer, de kwelflux (of de drainageweerstand of de effectieve porositeit) doelvariabelen?
Als gekozen wordt voor KALMAX of KALTFN dan kan men er verder voor kiezen om deze te koppelen aan een fysische interpretatie of niet. Het deterministische gedeelte van zowel KALMAX als KALTFN ziet er als volgt uit (h_t is de grondwaterstand en P_t het neerslagoverschot op tijdstip t):

$$h_t = c + a\left[h_{t-1} - c\right] + bP_t$$

Als het drainageniveau h_s (ten opzichte van maaiveld) constant is (zie volgende vraag) en bekend is (zie vraag daarna) dan kunnen de drainageweerstand g [T], de effectieve porositeit f [-] en de onderrandflux (kwel/infiltratie) q_b [LT^{-1}] uit de parameters a,b en c en de tijdstap Δt als volgt worden berekend (Knotters en Bierkens, 1999):

$$g = \frac{b}{1-a} \qquad f = \frac{-\Delta t}{g \ln a} \qquad q_b = \frac{c - h_s}{g}$$

Dit is de fysische interpretatie van KALMAX en KALTFN.

Wilt u rekening houden met meerdere drainageniveaus of variërende drainageniveaus?
Het is in principe wel mogelijk om bij meerdere drainageniveaus (als deze niet droogvallen en constant zijn) een fysische interpretatie van KALMAX en KALTFN te geven. In dat geval is het echter toch aan te raden SSD of SWAP te gebruiken. Bij variërende drainageniveaus moeten SSD of SWAP in ieder geval worden gebruikt.

Is het drainageniveau (greppeldiepte of slootpeil) ten opzichte van maaiveld bekend?
Zoals blijkt uit bovenstaande formules is het drainageniveau h_s (ten opzichte van maaiveld buis) nodig om een fysische interpretatatie van KALMAX of KALTFN te geven.

Zijn er behalve het neerslagoverschot nog andere invloeden die in de tijd variëren?
Dit gaat om de keuze tussen KALMAX en KALTFN. Als het neerslagoverschot de enige invoervariabele is die de grondwaterstand beïnvloedt, dan kan men afleiden dat de afhankelijkheid in de tijd van het ruisproces hetzelfde is als die van de (deterministische) grondwaterstand (Knotters en Bierkens, 1999). In dat geval kunnen we KALMAX gebruiken. In het geval dat er nog meer invloeden zijn (bijv. rivierpeilen) die niet expliciet zijn gemodelleerd (anders was de keuze niet op KALMAX of KALTFN gevallen) dan zitten die in de ruis. Dit betekent dat de tijdsafhankelijkheid van de ruis anders zal zijn dan die van de grondwaterstand zelf en is het gebruik van KALTFN op zijn plaats.

Zijn alle grondwaterstanden dieper dan twee meter beneden mv?
Bij dat soort diepe grondwaterstanden speelt de vertraging en demping van het neerslagoverschot in de onverzadigde zone een rol. Dat wordt niet gemodelleerd door SSD, zodat EMERALD en SWAP de overblijvende opties zijn.

Wilt u de verdampingsreductie gedetailleerd modelleren?
In EMERALD en SSD is de verdampingsreductie zeer eenvoudig geïmplementeerd als een verliesfunctie die afhangt van het gemiddeld vochtgehalte in de wortelzone. Wilt u de verdampingsreductie modelleren als functie van de wortelverdeling en het vochtprofiel, de reductiefunctie variëren in de tijd ten gevolge van gewasgroei of de verdamping opsplitsten in bodemverdamping en gewasverdamping, dan zult u SWAP moeten gebruiken.

Heeft u informatie over de onverzadigde doorlatendheid per bodemlaag?
Bij de toepassing van EMERALD of SWAP dient u informatie te hebben over de onverzadigde doorlatendheidskarakteristiek, $k(y)$, per bodemlaag. Deze informatie kan direct beschikbaar zijn uit metingen of deze kan indirect worden geschat via een vertaalfunctie gekoppeld aan de bodemtextuur via klassen (Wösten *et al.*, 1994) of continue vertaalfuncties (Wösten, 1997) van de Staringreeks, in welk geval de textuur van de bodem(lagen) bekend moet zijn. Is deze informatie niet aanwezig, maar de verlangde informatie in de volgende vraag wel, dan zult u SSD moeten toepassen.

Zijn de volgende gegevens beschikbaar: bodemgebruik, de bodemvochtretentiecurve, drainageniveau?
Om SSD, EMERALD of SWAP toe te passen moet u (onder andere) beschikken over het bodemgebruik om de neerslag minus interceptie te berekenen uit de KNMI neerslag en om de potentiële verdamping te berekenen uit de Makkink referentieverdamping. U wordt verwezen naar de beschrijvingen van SSD of EMERALD (zie ook de HELPfunctie van VIDENTE) voor een tabel met interceptieparameters en gewasfactoren. De bodemvochtretentie- of pF-curve, $q(y)$, van de bodem(lagen) is nodig. Net als $k(y)$ kan deze direct worden gemeten of worden afgeleid uit de textuurbeschrijving van de bodem(lagen) via klassen (Wösten *et al.*, 1994) of continue vertaalfuncties (Wösten 1997). Tenslotte moet u de drainageniveau(s) ten opzichte van maaiveld kennen.

Bevinden de meeste grondwaterstanden zich binnen 50 cm beneden mv?
EMERALD is niet erg geschikt voor het modelleren van zeer ondiepe grondwaterstanden. In dat geval moet gekozen worden tussen SSD en SWAP.

Zijn de volgende gegevens beschikbaar: bodemgebruik, bodemvochtretentiecurve, onverzadigde doorlatendheid, drainageniveau?
Om SWAP toe te passen moet u (onder andere) beschikken over het bodemgebruik. Dit is nodig voor het berekenen van interceptie, verdamping en eventueel gewasgroei. De pf-curve $q(y)$ en de onverzadigde doorlatendheid $k(y)$ van de bodem(lagen) zijn nodig. Deze kunnen direct worden gemeten of worden afgeleid uit de textuurbeschrijving van de bodem(lagen) via klassen (Wösten *et al.*, 1994) of continue vertaalfuncties (Wösten 1997). Tenslotte moet u de drainageniveau(s) ten opzichte van maaiveld kennen.

3 Graphical user interface

3.1 Set up

The DSS as well as the four models are controlled using a graphical user interface (GUI) programmed in Delphi 5. The working of the GUI is mostly self explanatory. Therefore, we will not provide an elaborate user manual, but give a global description of the set up and the input requirements.

In the start up screen the user can choose between running the DSS (option “choose model” selected with the mouse), and each of the four models. Before any actions can be performed a new project should be defined, or an existing project should be loaded first.

The DSS starts by choosing first from one of the four sets of target variables and then moving down a binary decision tree by answering questions with “yes” or “no” by a mouse click. The outcome of the DSS is either some method implemented in VIDENTE, a method not implemented in VIDENTE or the conclusion that further data gathering is required. If the outcome is one of the models implemented in VIDENTE, a button appears which leads the user directly to the input screen of that model by a click of the mouse.

For each model there is an input screen and an output screen. The input screen is divided into four tab sheets, each of which is used to provide input (parameters and input file locations) for calibration, prediction and simulation with the model. These three modes are explained in the model descriptions in the subsequent parts. After all entries have been filled in, the model can be run by clicking the “Run” button. Also, a tab sheet is reserved for providing the input to the program STATSIM, which is used for statistical analysis of realisations simulated with the models (the output from “simulation”).

The output screen consists of two tab sheets, one showing graphical output (if any) and one output text files (if any).

The help files can be accessed using F1, the speed button or the pull down menu.

VIDENTE stores the output in the directory (= folder) that has been indicated by the user as the directory where the new project should be stored (called “data directory” in the dialog box that appears when creating a new project). Project meta-information is stored in this directory in a file with “[name piezometer]” as root and “.vpr” as extension. For each of the models, subdirectories are created and within these subdirectories sub-subdirectories named “calibration”, “prediction”, “simulation” and “statsim”. In these sub-subdirectories the associated temporary input files for the models as well as the output of the model runs are stored.

3.2 Input files

Model parameters are provided by filling in the screens. The input time series have to be provided in the form of ASCII files. These files can be placed anywhere on the computer. When a new project is created, the files are copied to the subdirectory at which the new projects is stored.

File with meteodata (see Figure 1 for an example file)
Precipitation and evapo-transpiration (mm/d) of the following format:
Record 1: *ndata* - number of records with meteodata (integer)
Record 2, *ndata*+1: *year* (integer, four digits, e.g. 1988)
month (integer, {1,2,…,12})
day (integer, {1,2,…,31})
precipitation (if necessary minus interception) (mm/day) (real)
evapo-transpiration (mm/day) (real) (Makkink reference crop evapo-transpiration in case of SSD and EMERALD)

KALMAX, KALTFN and SSD assume the values to be constant in the interval $<t,t+Dt]$. In case of EMERALD time steps should always be one day ($Dt=1$), so that a record is needed for every day number.

```
4018
    1980       1       1     5.8     0.3
    1980       1       2     0.6     0.3
    1980       1       3     1.3     0.1
    1980       1       4     9.1       0
    1980       1       5     4.2     0.1
    1980       1       6     6.2     0.1
    1980       1       7     0.3     0.1
    1980       1       8     0.8     0.1
    1980       1       9       0     0.1
    1980       1      10       0     0.1
    1980       1      11       0     0.1
    1980       1      12       0     0.4
    1980       1      13       0     0.4
    1980       1      14       0     0.2
    1980       1      15       0     0.1
```

Figure 1 Example of input meteofile for VIDENTE

File with groundwater levels (see Figure 2 for an example file)
File with water table depth (cm reference level)
Record 1: *ndata* - number of records with water table observations (integer)
Record 2, *ndata*+1: *year* (integer, four digits, e.g. 1988)
month (integer, {1,2,…,12})
day (integer, {1,2,…,31})
water table depth (cm reference level) (real)

Note: intervals between observations do not have to be regular.

```
210
  1981    1    9    -103.00
  1981    2   23    -126.00
  1981    3   10     -96.00
  1981    3   25     -97.00
  1981    4    9    -122.00
  1981    4   24    -132.00
  1981    6   23    -137.00
  1981    7    8    -125.00
  1981    7   23    -134.00
  1981    8    7    -131.00
  1981    9   21    -139.00
  1981   10    6    -137.00
  1981   10   21     -82.00
  1981   11    5    -105.00
```

Figure 2 Example of file with water table depths for VIDENTE

File with surface water levels (see Figure 3 for an example file)

Only used by SSD. For each drainage level a time series of either a surface water level (time varying, draining and infiltrating) or a ditch or trench bottom (constant in time, only draining) is given. Levels in cm with respect to the reference level (=surface level for SSD).

Record 1: *ndata* - number of records with surface water levels (integer)
nlevels – number of drainage levels recorded (integer)

Record 2, *ndata* +1: *year* (integer, four digits, e.g. 1988)
month (integer, {1,2,…,12})
day (integer, {1,2,...,31})
For i = 1, *nlevels*
drainage level i (cm reference level) (real)

```
24 2
 1980  3  31 -180   -50
 1980  9  31 -140   -50
 1981  3  31 -180   -50
 1981  9  31 -140   -50
 1982  3  31 -180   -50
 1982  9  31 -140   -50
 1982 12  31 –180   -50
```

Figure 3 Example input file with varying surface water levels as used by VIDENTE (only used by SSD)

SSD assumes the levels to be constant between two observations, where the record given provides the last day of a period. So in this case, water levels of drainage level 1 are 180 cm below reference level between January 1 1980 and March 31 1980, 140 cm below reference level from April 1 1980 to September 31 1980, etc. Note that if the last record had not been added and we were modelling water table depth until December 31 1982, that the period October 1 1982 to December 31 1982 would have been given the value of –140, instead of –180. Also, note that we have two drainage levels here. One surface water level varying with time and one trench or drain with a bottom depth of 50 cm below reference.

3.3 Example run using KALMAX

In this section we present an example application of VIDENTE where it is used to run the model KALMAX.

We have created a subdirectory called “demo” at which we have put our data files, i.e. “debilt.met” containing 30 years of precipitation and potential evapo-transpiration from meteorological station “De Bilt” (Netherlands) (period January 1 1961 – December 31 1990) and the file “debilt.gws” containing the water table depths observed two times a month for the period 1985-1990. Our goal is to obtain fluctuation quantities (mean highest water table (MHW) and mean lowest water table (MLW), etc) for the entire 30 year period.

After starting up VIDENTE the following start up screen (Figure 4) appears:

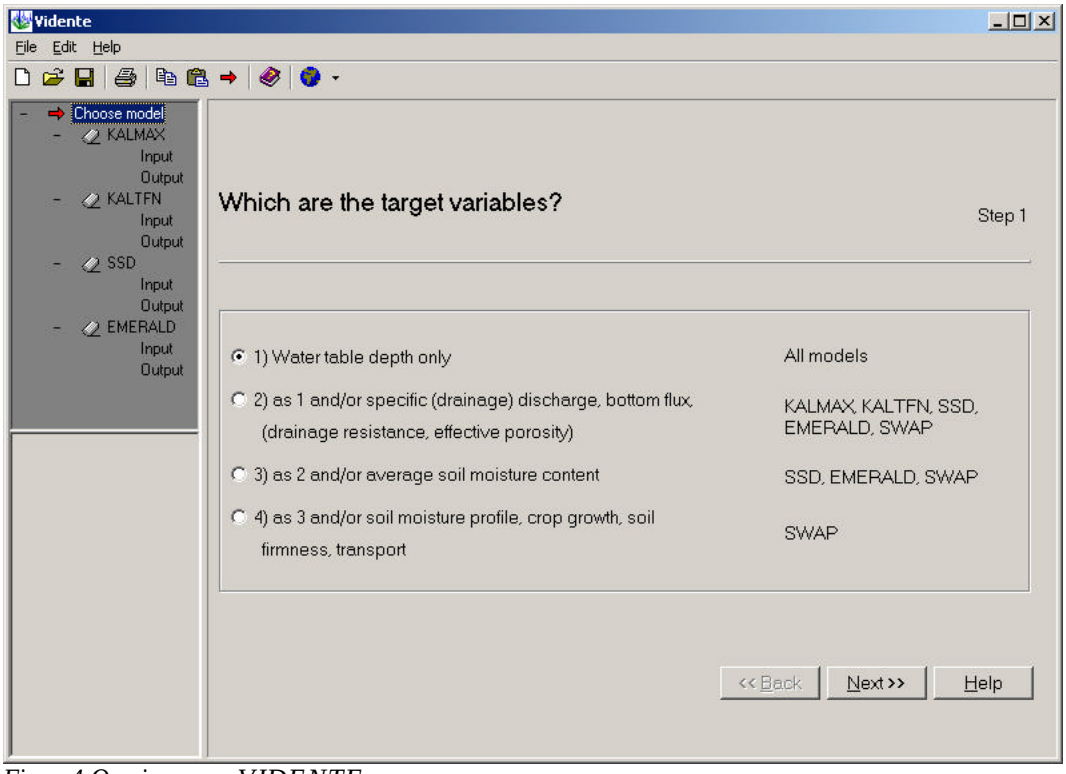


Figure 4 Opening screen VIDENTE

As can be seen, the user’s first action is to create or open a new project. For this move the cursor to <file/new project> or <file/load project> or use one of the speed buttons. The only action possible before a project is loaded is to change the language from English to Dutch or vice versa. If a project is loaded or created and the user wants to change the language again, the project has to be closed and saved first. After the language has been changed, the saved project can be loaded again to continue working on that project. Figure 5 shows the dialog box used to define a new project.

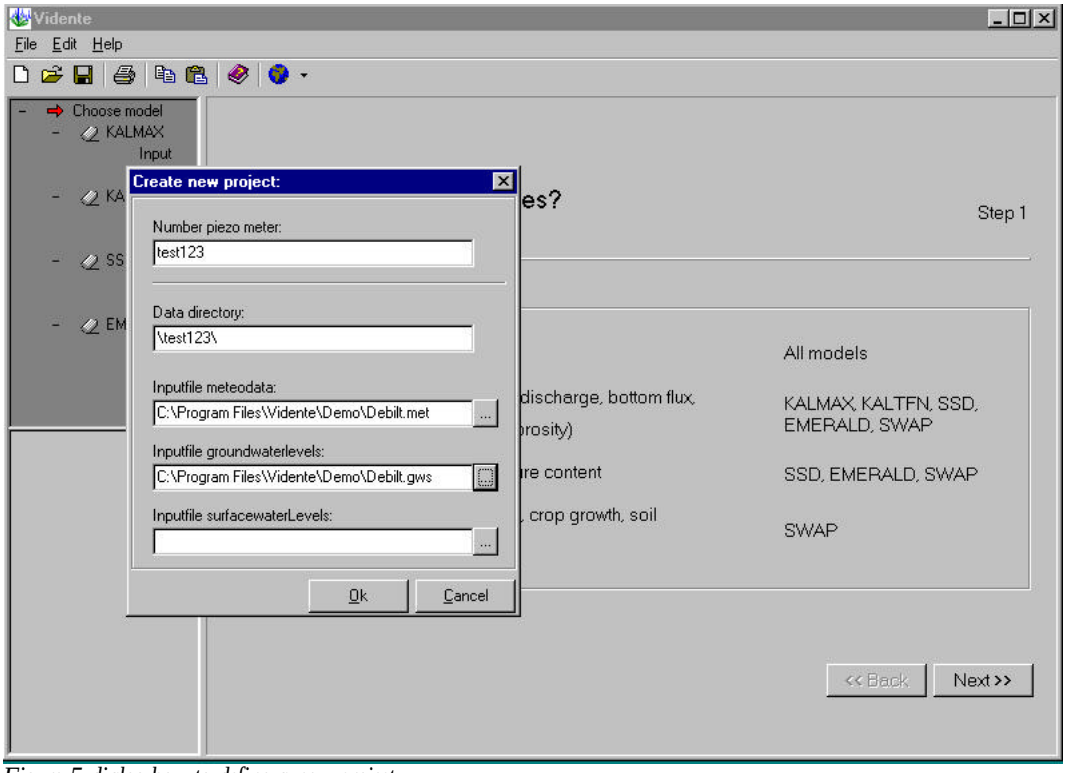


Figure 5 dialog box to define a new project

Here a subdirectory c:\test123 has been created where we have copied the data files from the subdirectory c:\Program Files\Vidente\Demo. The project information will be stored in a file test123.vpr.
Now we can proceed with either entering the DSS or going directly to one of the models. To give some idea about the workings of the DSS we show two screens here (Figure 6), one showing a yes/no question and one ending screen that leads to either the model KALMAX or KALTFN.

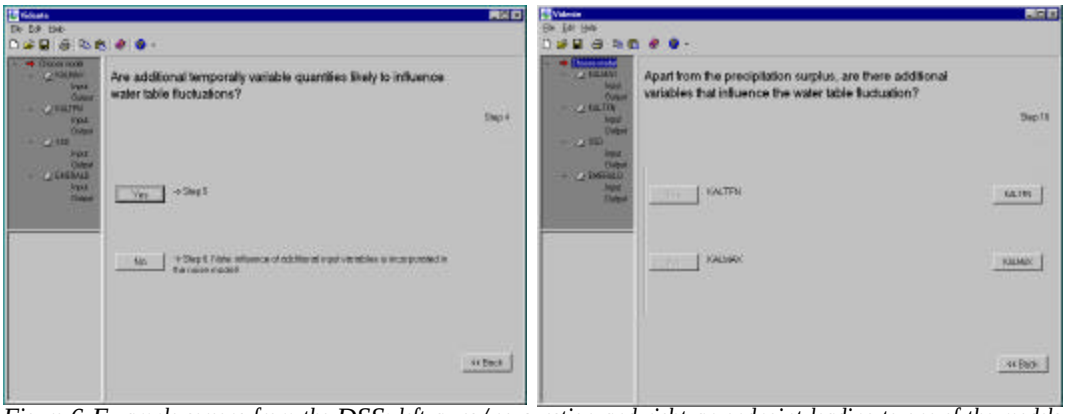


Figure 6 Example screens from the DSS; left a yes/no question and right an endpoint leading to one of the models

If the user wants to remember the choices made that lead to the model, these can be printed out. Clicking on the KALMAX button will start the input menu of the KALMAX model. In Figure 7 we start with the calibration tab sheet. We calibrate

the ARX model using the water table depths from 1985-1990. First the deterministic calibration is performed. The closure criterion is used in the optimisation algorithm and should be increased if obtaining a solution takes too long. Figure 7 shows the input screen, where “mode = deterministic” is checked. Note that the output files automatically get the extension “.dat” so that no extension have to be given. If one is given, it is replaced by “.dat”. This entails that for the output files different names have to be given.

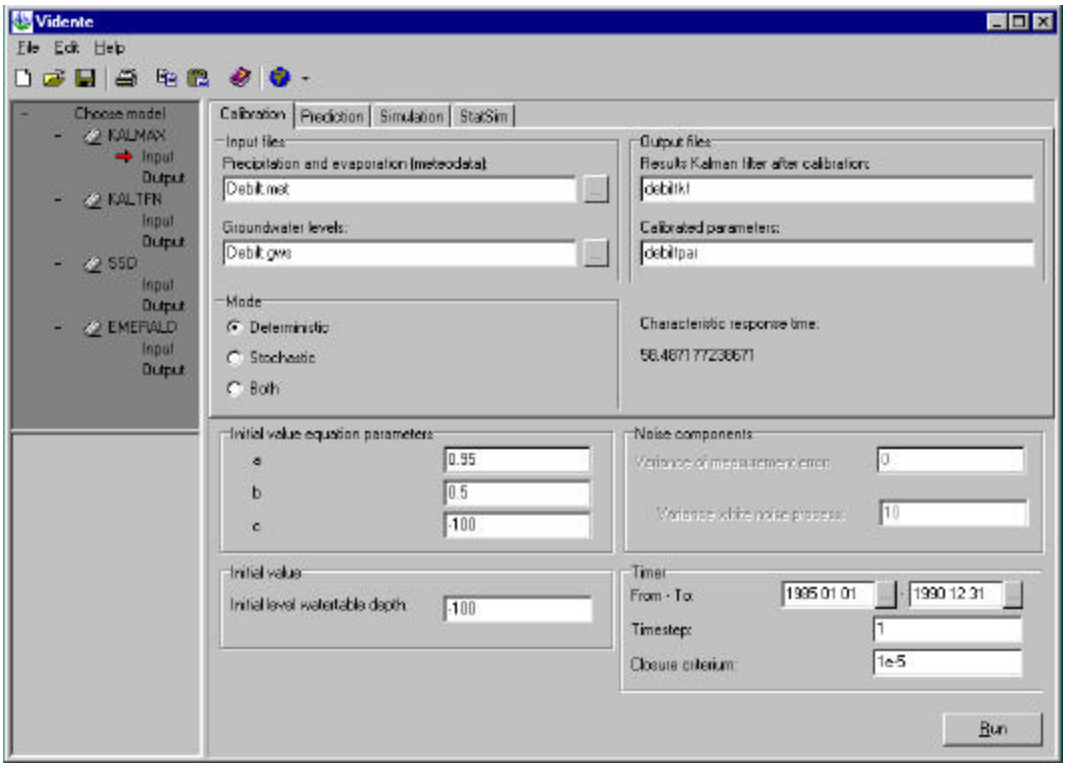


Figure 7 Example input screen deterministic calibration with KALMAX

During the calibration, the value of the closure criterium is monitored (Figure 8). The calibration can be stopped by pressing the ``stop" button.

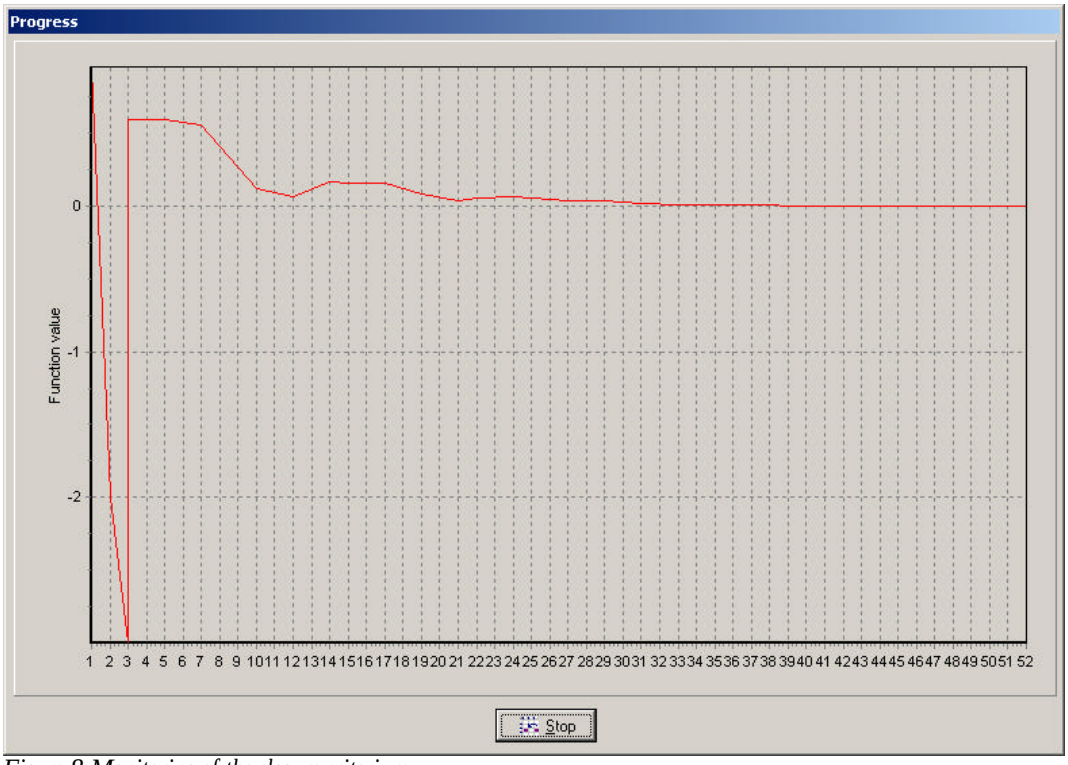


Figure 8 Monitoring of the closure criterium

After running the deterministic calibration, the program jumps to the output screen (see hereafter) and a box appears (Figure 9) with the calibrated parameters, some calibration statistics (ME, RMSE and MAE of the calibrated deterministic model in cm) and the question whether we want them to replace our initial values. If we agree the parameters are replaced.

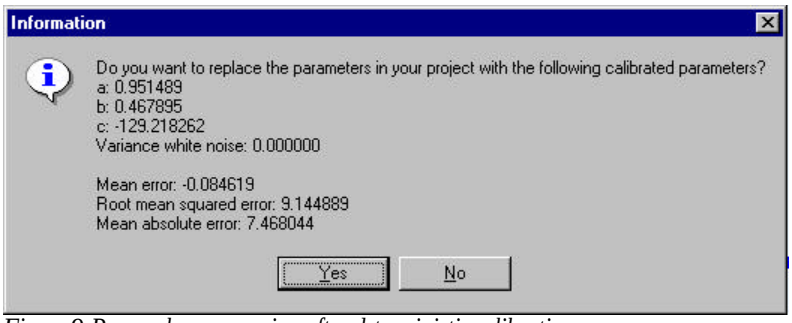


Figure 9 Pop up box appearing after deterministic calibration

The next step is to run the stochastic calibration while using and fixing the already obtained parameters from the deterministic calibration. The input screen can be seen in Figure 10. Notice that also the characteristic response time (days) is given which can be calculated from the a-parameter (Bierkens et al., 1999; see Equation (34) in section 2.4.4 of part 2). Observation of water table depth should not be taken further apart than 60 days in order to calibrate the ARX model with a = 0.951.

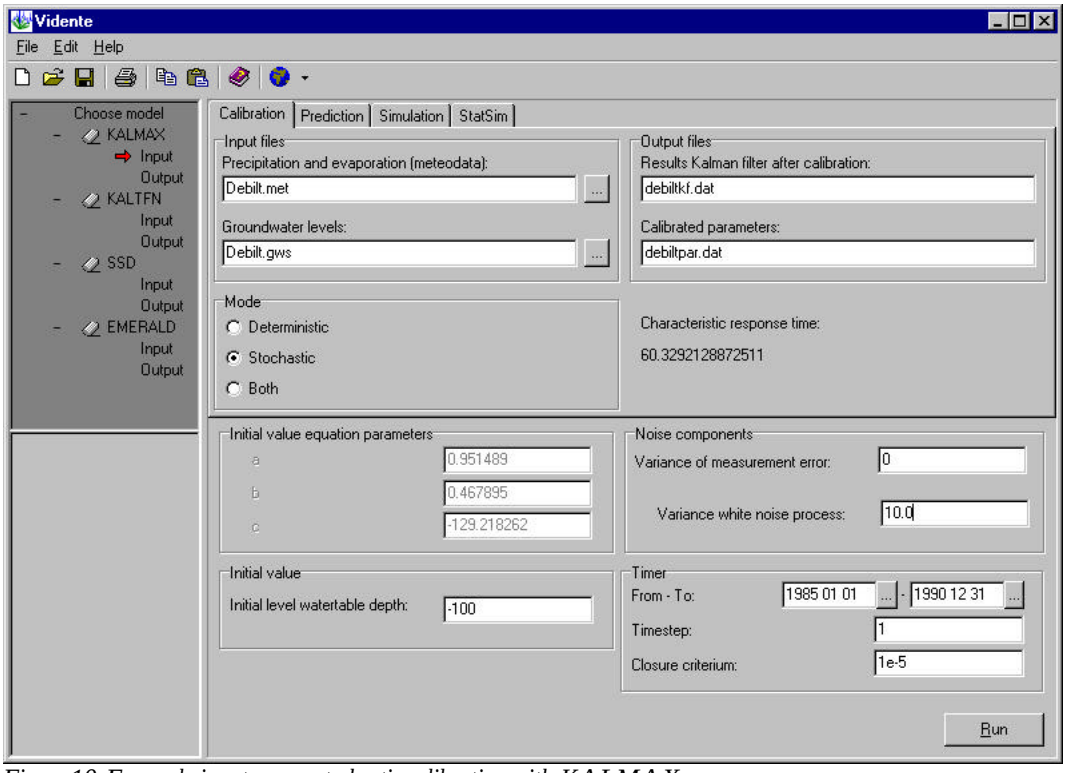


Figure 10 Example input screen stochastic calibration with KALMAX

After the stochastic calibration a similar pop up box appears with the question whether the parameters are to be preserved. Finally a calibration round is performed with the mode “both”. Here in front of each parameter a check box appears (Figure 11). If the box is checked, the parameter is calibrated, if not it is fixed. For unbiased estimates of the parameters all boxes should be checked. After all parameters have been calibrated, again a pop up box appears (Figure 13) showing the final parameter estimates and the final statistics. To preserve these parameters and use them for prediction and simulation later on the user should click “yes”. Note that in case of KALMAX these are predictions with the Kalman filter and are therefore slightly better than those of the deterministic model (Figure 9) due to the updating (see chapter 2 of Part 2 of this report).

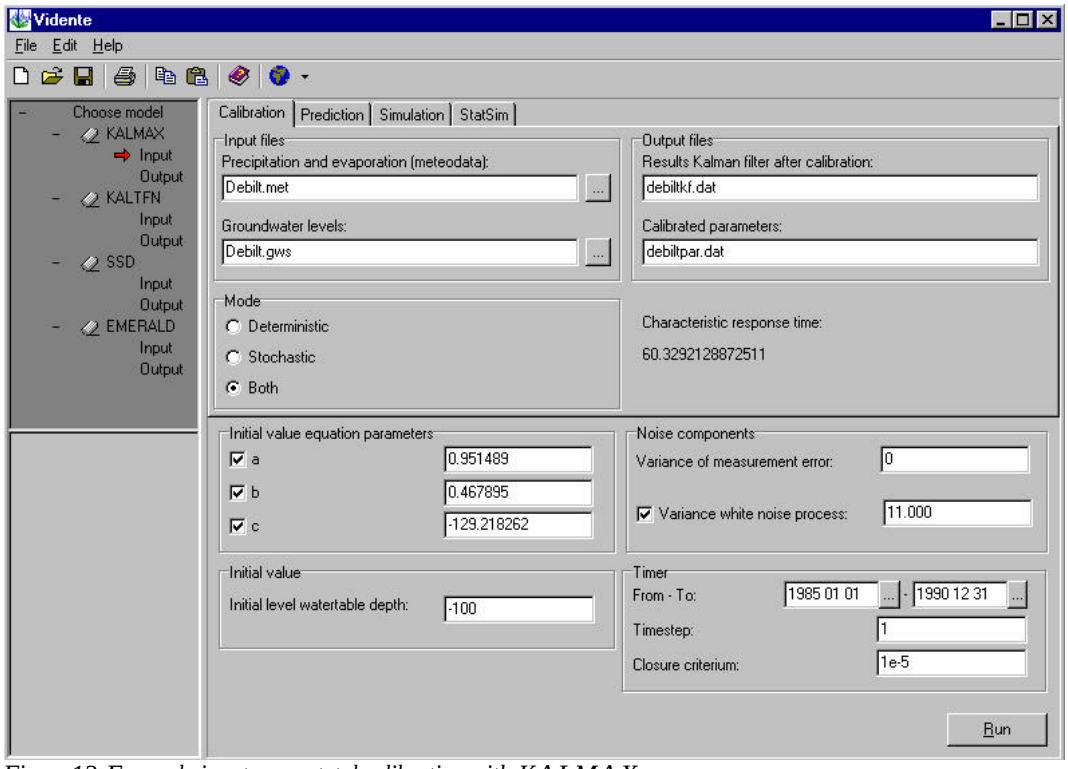


Figure 12 Example input screen total calibration with KALMAX

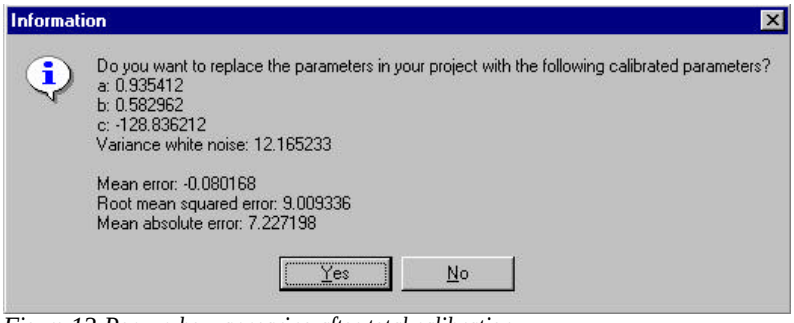


Figure 13 Pop up box appearing after total calibration

The results of the calibration can be looked at in the output screen. Each tab sheet of each model has a separate input screen and output menu screen. For instance, the calibration tab sheet of KALMAX has the input screen (Figure 12) and an output screen (Figure 14). Which one the user is looking at is indicated by the arrow on the left. The output screen itself consists of two tab sheets, a “graphics” sheet showing output graphs if any are produced and “text” in which the content of text files produced can be viewed. Figures 14 and 15 show the results of above calibration. In Figure 14 a graph of observations and Kalman filter predictions is shown, as well as innovations (differences between Kalman filter predictions and observations) and the 95% prediction limits of these innovations. The text file shows the content of the file with calibration results, i.e. the parameter values (the line for which the number in the last column is smallest) and on the bottom line ME, RMSE and MAE respectively.

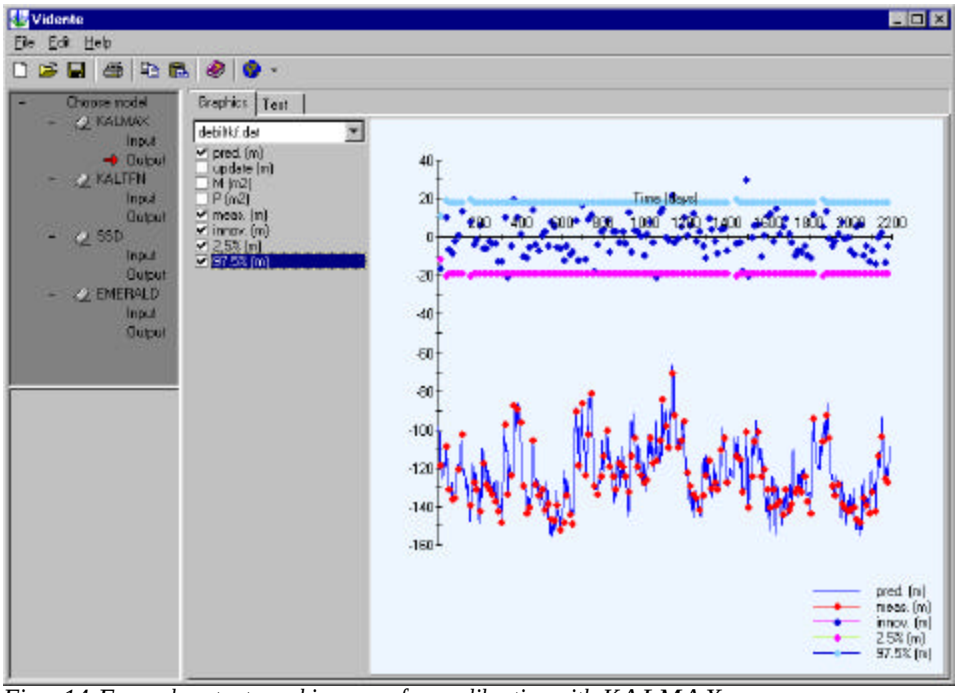


Figure14 Example output graphics screen from calibration with KALMAX

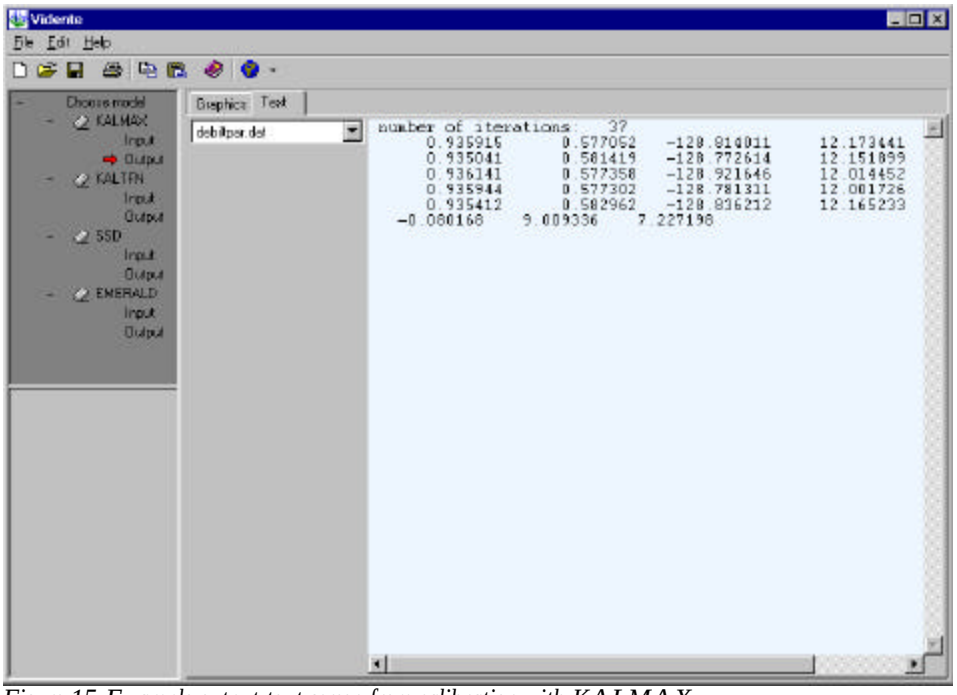


Figure 15 Example output text screen from calibration with KALMAX

Once the calibration is performed it is always wise to check whether the deterministic part of the model is suitably parameterised (see section 2.4 of Part 2). To do this, a prediction run can be made without the Kalman filter updating using the calibrated parameters. Figure 16 shows the prediction input screen, Figures 17 and 18 the output screens. Note that on the input screen the theoretical variance of the prediction error (without updating) is given. This variance is calculated with Equation

(18) in section 2.1 of Part 2. So a RMSE of about 9.9 cm is expected. The output files show that the fit of the deterministic model looks satisfactorily and that the actual RMSE is somewhat smaller than the theoretical one. (The bottom line in the screen of Figure 18 shows ME, RMSE and MAE, respectively.) Also notice that the validation statistics are slightly higher than those of Figures 13 and 15, as can be expected as no updating is used here. The graphics in Figure 17 can be printed directly or included into an Office document (Word, Powerpoint) by using copy and paste. Besides a graph, Figure 17 gives a percentage of variance accounted for ('percentage explained') and the variance of residuals (predicted minus observed value), as measures of relative and absolute accuracy, respectively.

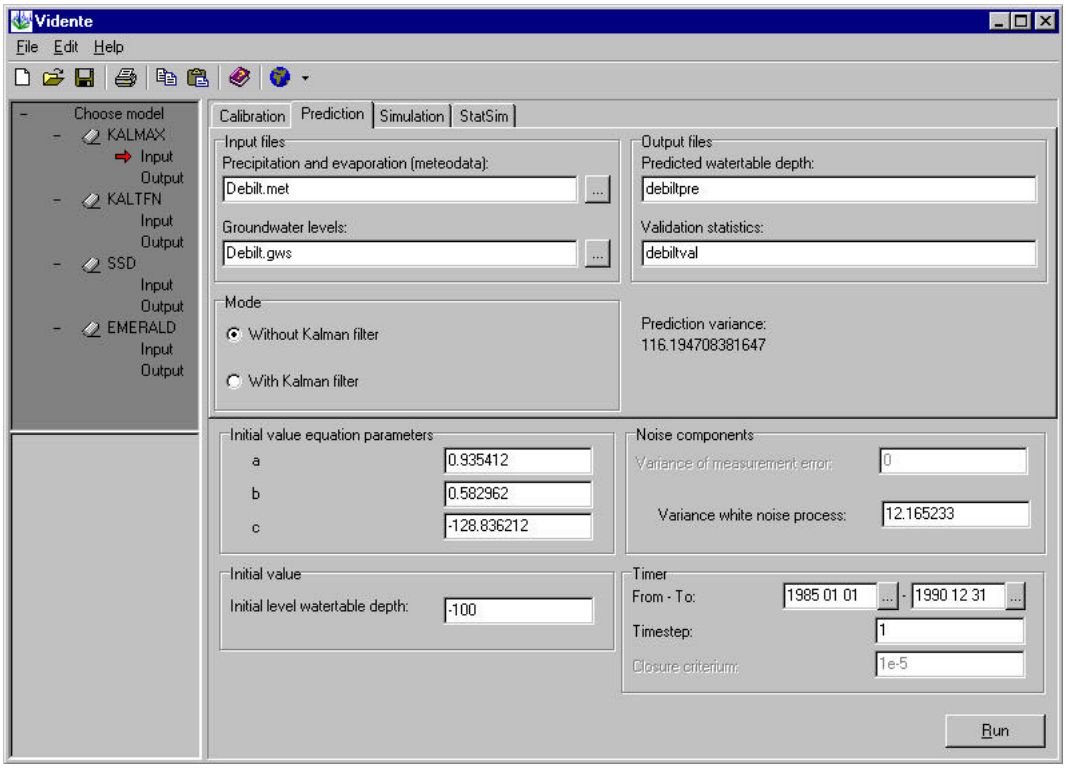


Figure 16 Example input screen from prediction with KALMAX

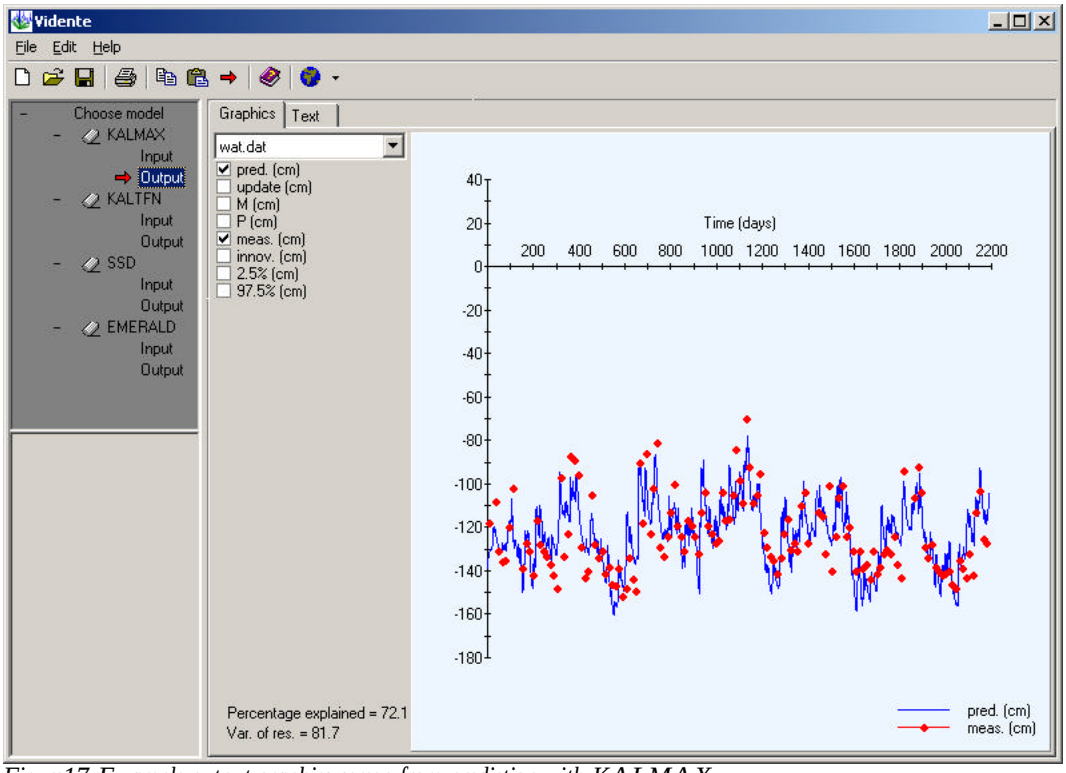


Figure17 Example output graphics screen from prediction with KALMAX

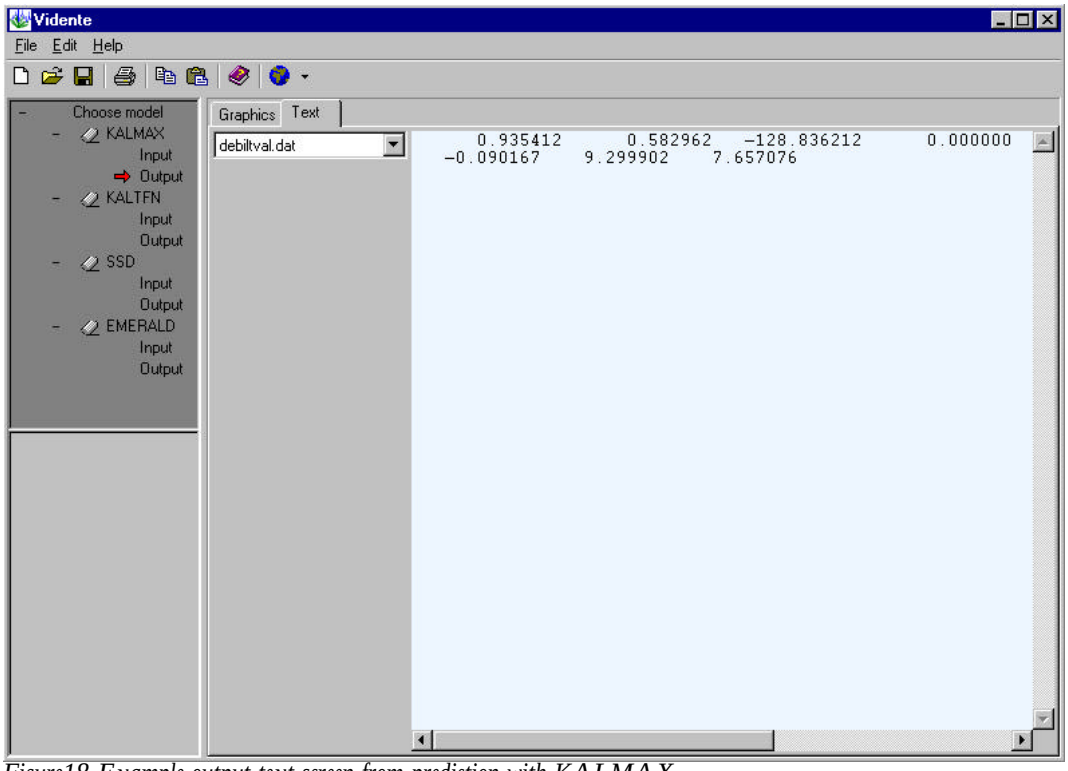


Figure18 Example output text screen from prediction with KALMAX

Finally, using the calibrated parameters we simulate a sufficient number of realisations of 30 years water table depth using the meteodata from 1961 to 1990 (see input screen Figure 19). The results are analysed with the program STATSIM. The input screen of STATSIM is shown in Figure 20. Results of STATSIM are a number of graphic files and a textfile with univariate statistics (see section 2.4.4 of Part 2 for a description). Shown here are the text file with statistics (Figure 21) and two graphics, i.e. the frequency of exceeding graph (FOE graph) in Figure 22 and a regime graph in Figure 23.

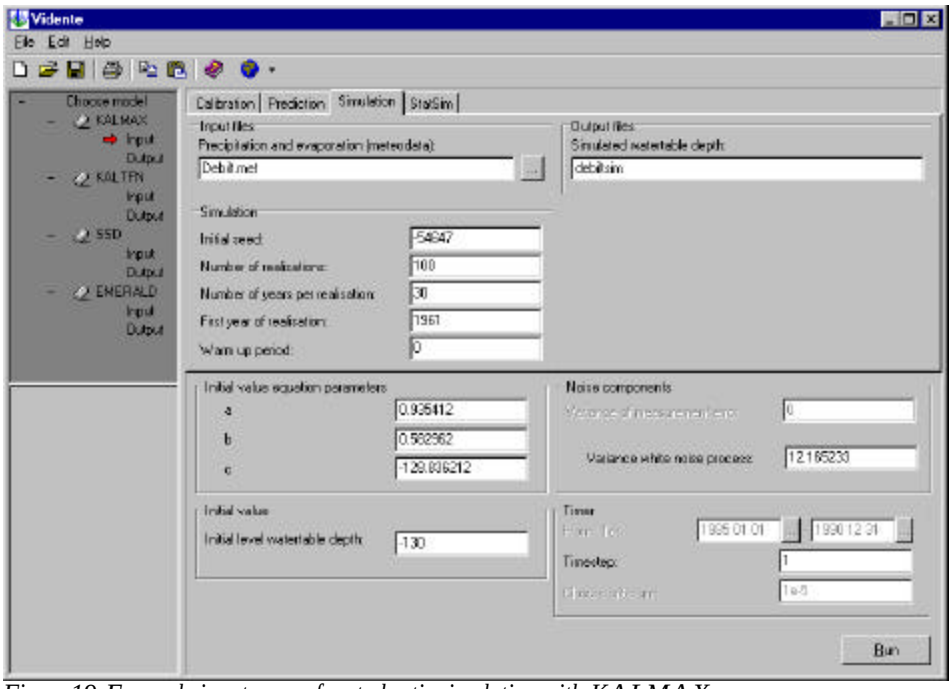


Figure 19 Example input screen for stochastic simulation with KALMAX

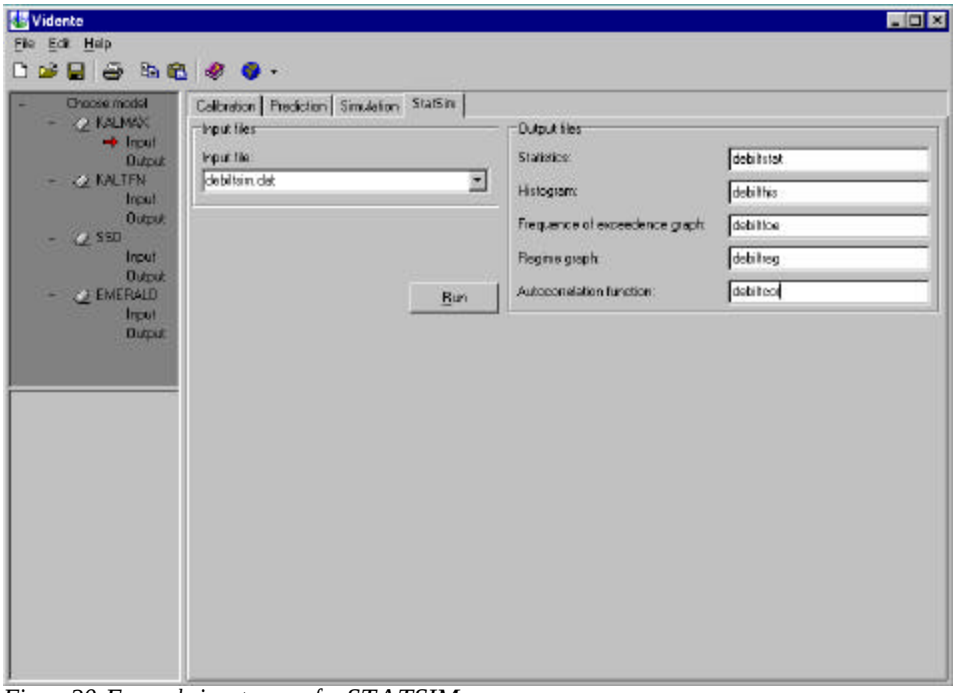


Figure 20 Example input screen for STATSIM

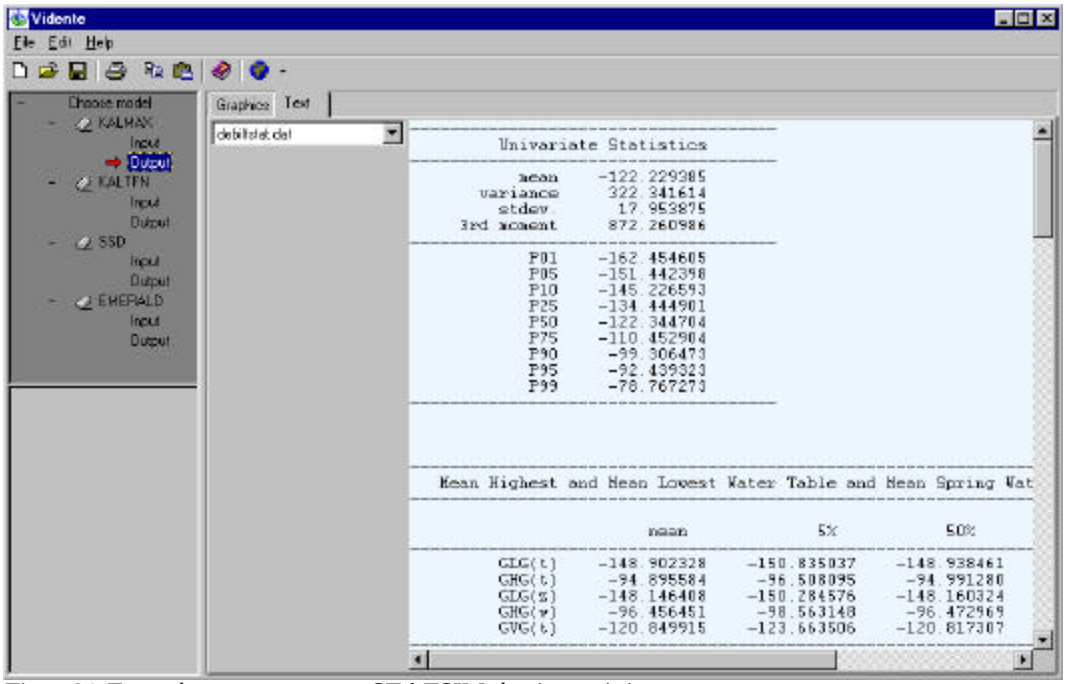


Figure 21 Example output text screen STATSIM showing statistics

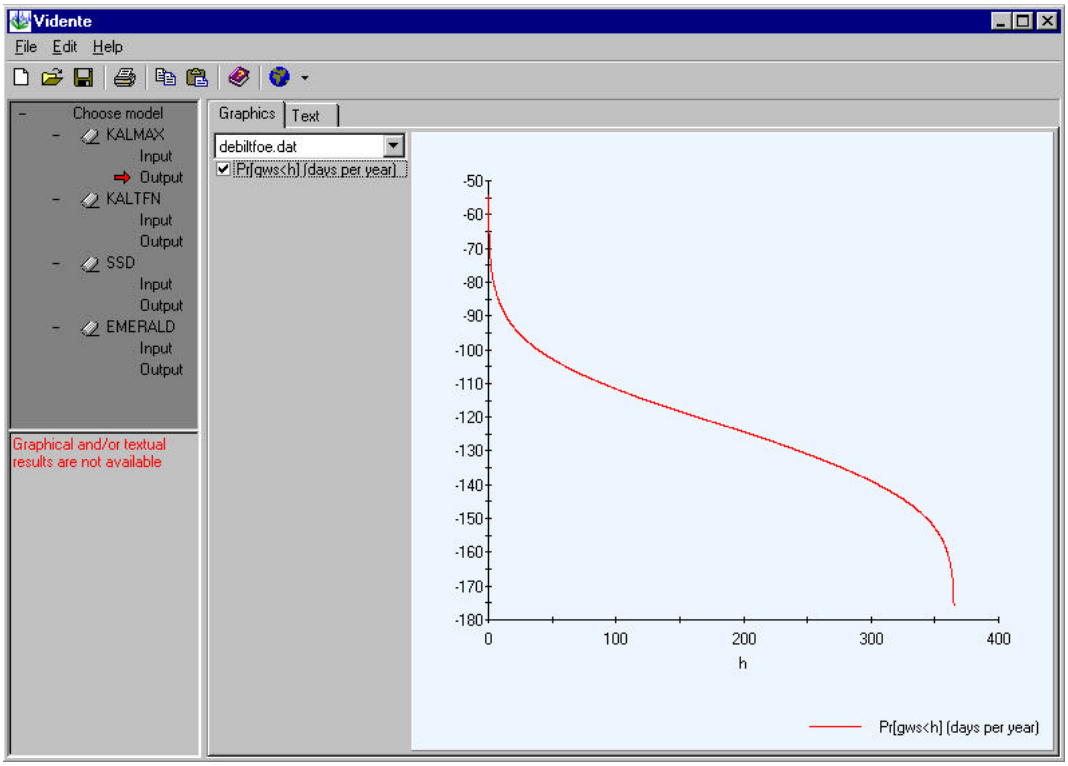


Figure 22 Example output text screen STATSIM showing FOE graph

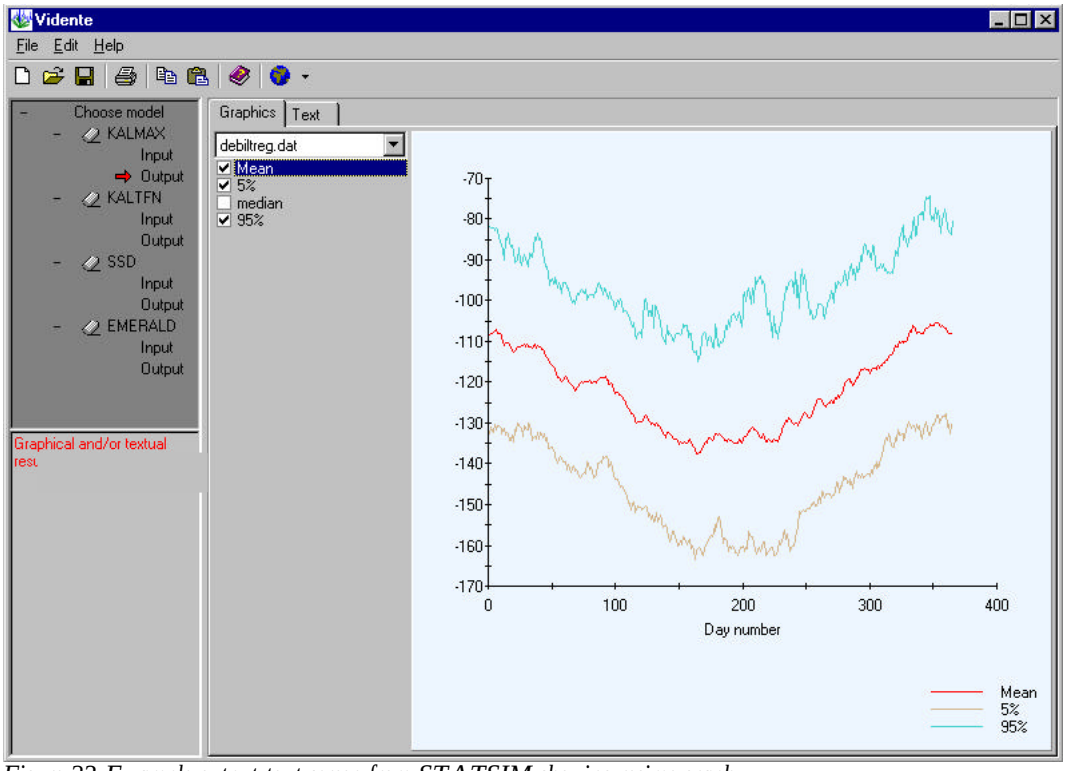


Figure 23 Example output text screen from STATSIM showing regime graph

References

Bierkens, M.F.P., 1998. Modeling water table fluctuations by means of a stochastic differential equation. *Water Resources Research* 34(10), 2485-2499.

Bierkens, M.F.P., 1998. Eenvoudige stochastische modellen voor grondwaterstandsfluctuaties; deel 1: Een stochastische differentiaalvergelijking. *Stromingen* 4(2), 5-26.

Bierkens M.F.P. and D.J.J. Walvoort, 1998. Eenvoudige stochastische modellen voor grondwaterstandsfluctuaties; deel 2: Gecombineerd bodem-grondwatermodel met stochastische invoer. *Stromingen* 4(3), 5-20.

Bierkens M.F.P., M. Knotters and F.C. van Geer, 1999. Calibration of transfer function-noise models to sparsely or irregularly observed time series. *Water Resources Research* 35(6), 1741-1750.

Bierkens, M.F.P, M. Knotters and F.C. Van Geer, 1999. Tijdreeksanalyse nu ook toepasbaar bij onregelmatige meetfrequenties. *Stromingen* 5(2), 43-54.

Bierkens, M.F.P., 2001. Spatio-temporal modelling of the soil water balance using a stochastic model and soil profile descriptions. *Geoderma* 103, 27-50.

Bierkens, M.F.P., M. Knotters and T. Hoogland, 2001. Space-time modelling of water table depth using a regionalized time series model and the Kalman filter. *Water Resources Research* 37, 1277-1290.

Dam, J.C. van, J. Huygen, J.G. Wesseling, R.A. Feddes, P. Kabat, P.E.V. Van Walsum, P. Groenendijk and C.A. Van Diepen, 1997. Theory of SWAP version 2.0; simulation of water flow, solute transport and plant growth in the soil-water-atmosphere-plant environment. Report 171, Department of water Resources, Wageningen Agricultural University, Wageningen. Technical Document 45, DLO Winand Staring Centre, Wageningen.

Hipel, K.W. and A.I. McLeod , 1994. *Time Series Modelling of Water Resources and Environmental Systems*. Developments in Water Science 45, Elsevier Science, Amsterdam.

Knotters M. and P.E.V. van Walsum, 1997. Estimating fluctuation quantities from time series of water-table depths using models with a stochastic component. *Journal of Hydrology* 197, 25-46.

Knotters M. and P.E.V. van Walsum, 1994. Uitschakeling van weersinvloeden bij de karakterisering van het grondwaterstandsverloop. Rapport 350, DLO Staring Centrum, Wageningen.

Knotters M. and J.G. de Gooijer, 1999. TARSO modeling of water table depths. *Water Resources Research* 35(3), 695-706.

Knotters, M. and M.F.P. Bierkens, 1999. Tijdreeksmodellen voor de grondwaterstand; een kijkje in de black box. *Stromingen* 5(3), 35-49.

Knotters, M. and M.F.P. Bierkens, 2000. Physical basis of time series models for water table depths. *Water Resources Research* 36(1), 181-188.

Knotters, M. and M.F.P. Bierkens, 2001. Predicting water table depths in space and time using a regionalised time series model. *Geoderma* 103, 51-77.

Supit, I., A.A. Hooyer en C.A. van Diepen (Eds.), 1994. *System description of the WOFOST 6.0 crop simulation model implemented in CGMS. Vol. 1: theory and algorithms.* EUR publication 15956, Agricultural Series, Luxembourg, 146 pp.

Tong, H., 1990. *Non-linear Time Series. A dynamical System Approach*. Clarendon Press, Oxford.

Walvoort, D.J.J. and M.F.P. Bierkens,1998. Emerald: A stochastic modelling approach for rapid assessment of groundwater dynamics. Report 171, DLO Winand Staring Centre, Wageningen.

Winter, T.C., D.O. Rosenberry, and A.M. Sturrock, 1995. Evaluation of 11 equations for determining evaporation for a small lake in the north central United States. *Water Resources Research* 31(4), 983-993.

Wösten, J.H.M., Van Genuchten, M.Th, 1988. Using texture and other soil properties to predict the unsaturated hydraulic conductivity. *Soil Science Society of America Journal* 52(6), 1762-1770.

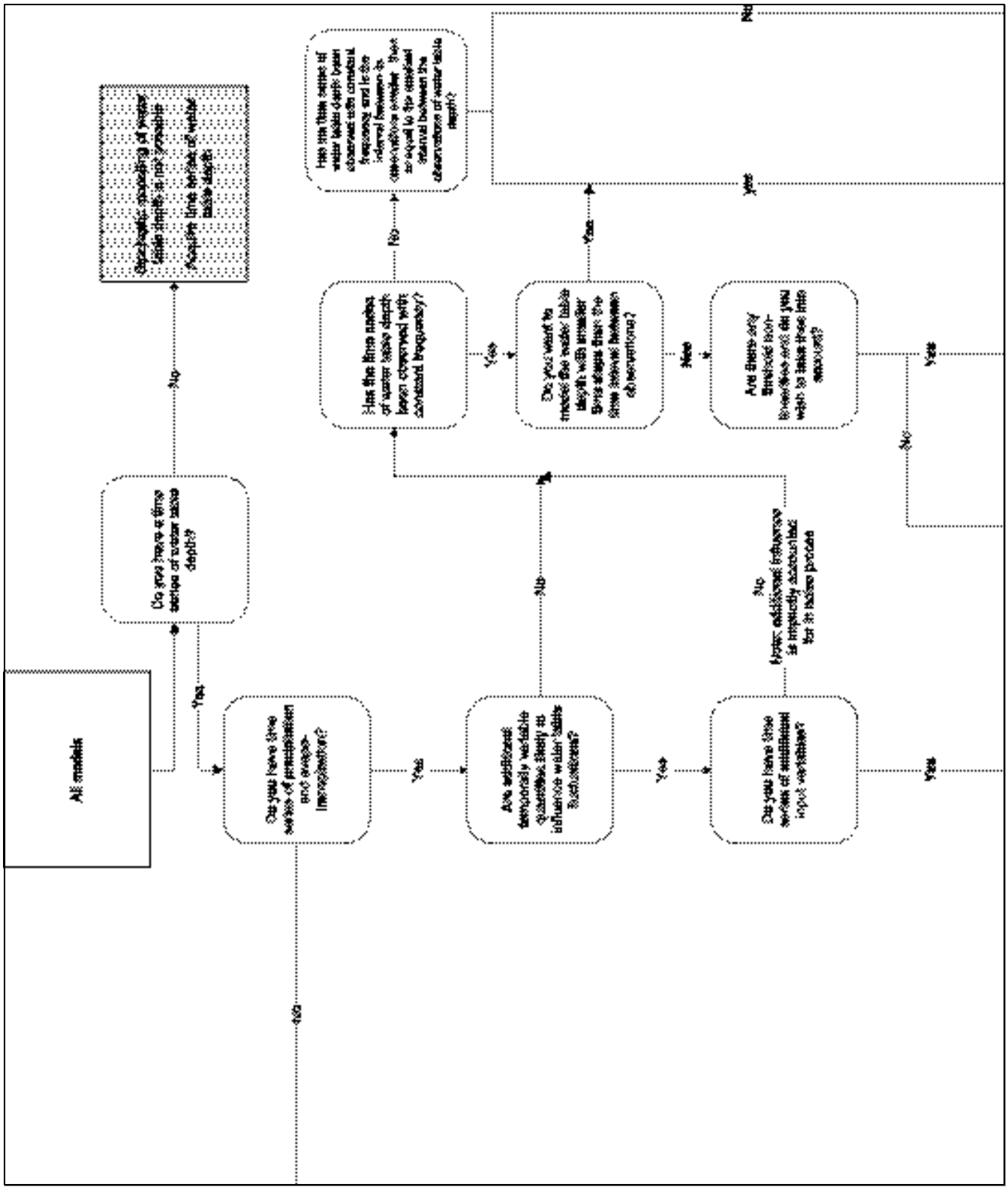
Wösten, J.H.M., G.J. Veerman and J. Stolte, 1994. Waterretentie- en doorlatendheidskarakteristieken van boven- en ondergronden in Nederland: de Staringreeks; vernieuwde uitgave 1994. Technisch Document 18, DLO Staring Centrum, Wageningen.

Wösten, J.H.M, 1997. Pedotransfer functions to evaluate soil quality. In: Gregorich and Carter (Eds.), *Soil Quality for Crop Production and Ecosystem Health.* Developments in Soil Science 25, Elsevier, Amsterdam.

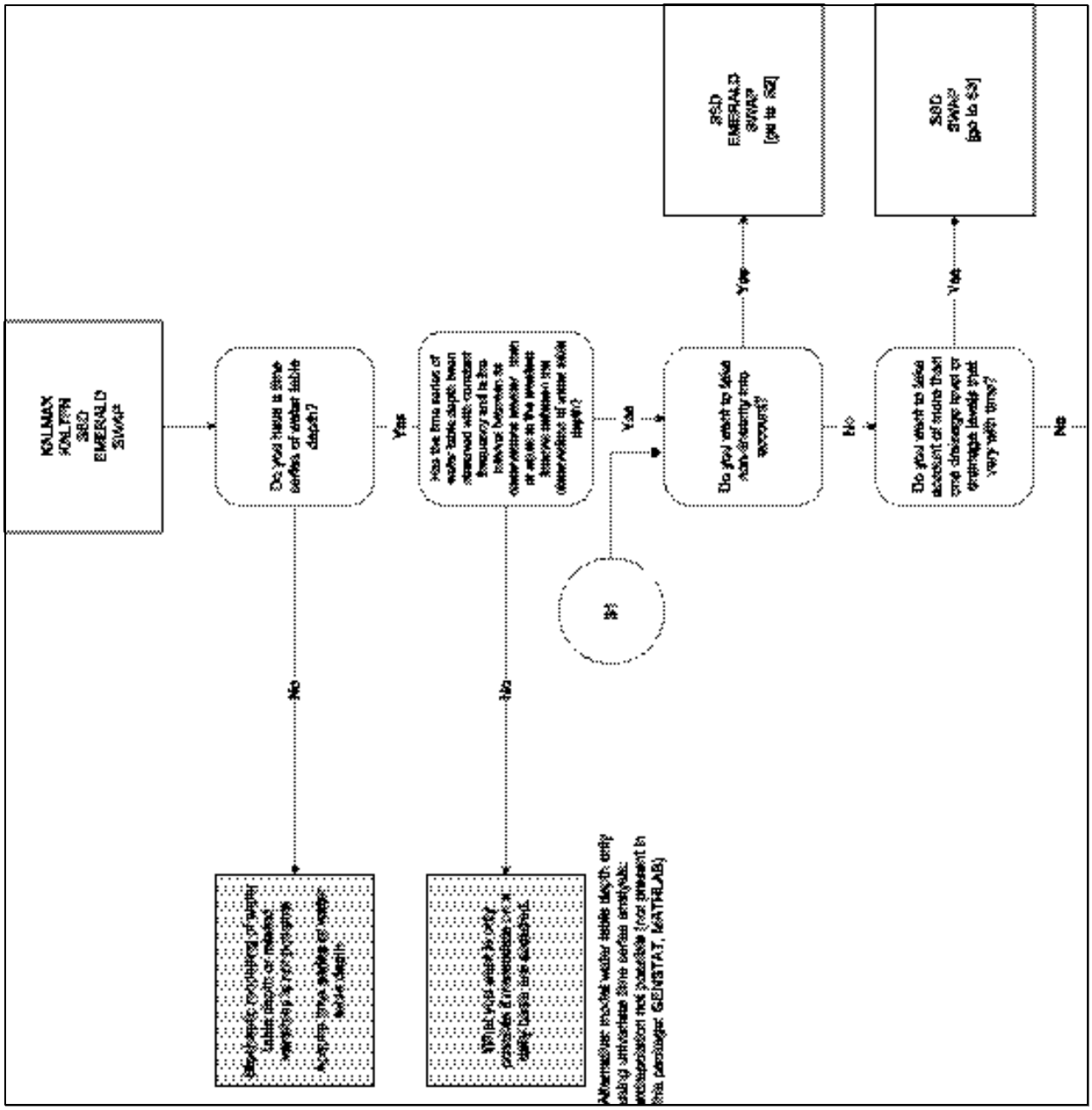
Annex: Binary decision trees of the DSS

Following is a set of binary decision trees that form the basis of the DSS. The DSS starts with choosing one of the four options, depending on the target variables variables. Each options is the root of a binary decision tree that leads to the choice of a method based on answering a series of yes/no to questions. These questions as well as possible outcomes of the decision trees are treated in chapters 2 (in English) and 3 (in Dutch).

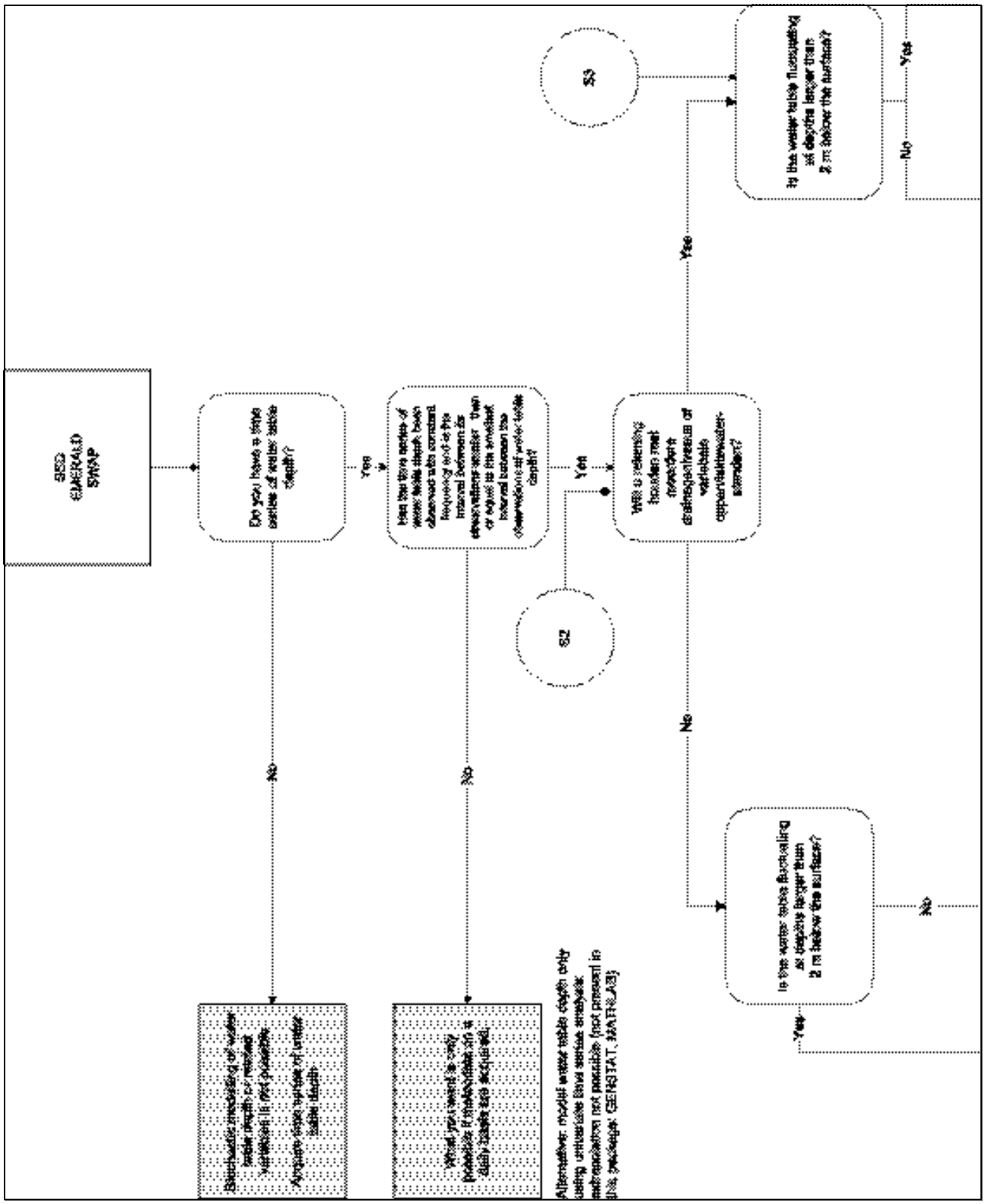
What are the target variables?
(1) water table depth only
All models
(2) as 1 and/or specific (discharge) discharge, bottom flux, (drainage resistance, effective porosity)
KALMAN MACRO SGD EMERALD SWAP
(3) as 2 and/or average soil moisture content
SGD EMERALD SWAP
(4) as 2 and/or soil moisture profile, crop growth, soil fluxes, transport
SWAP

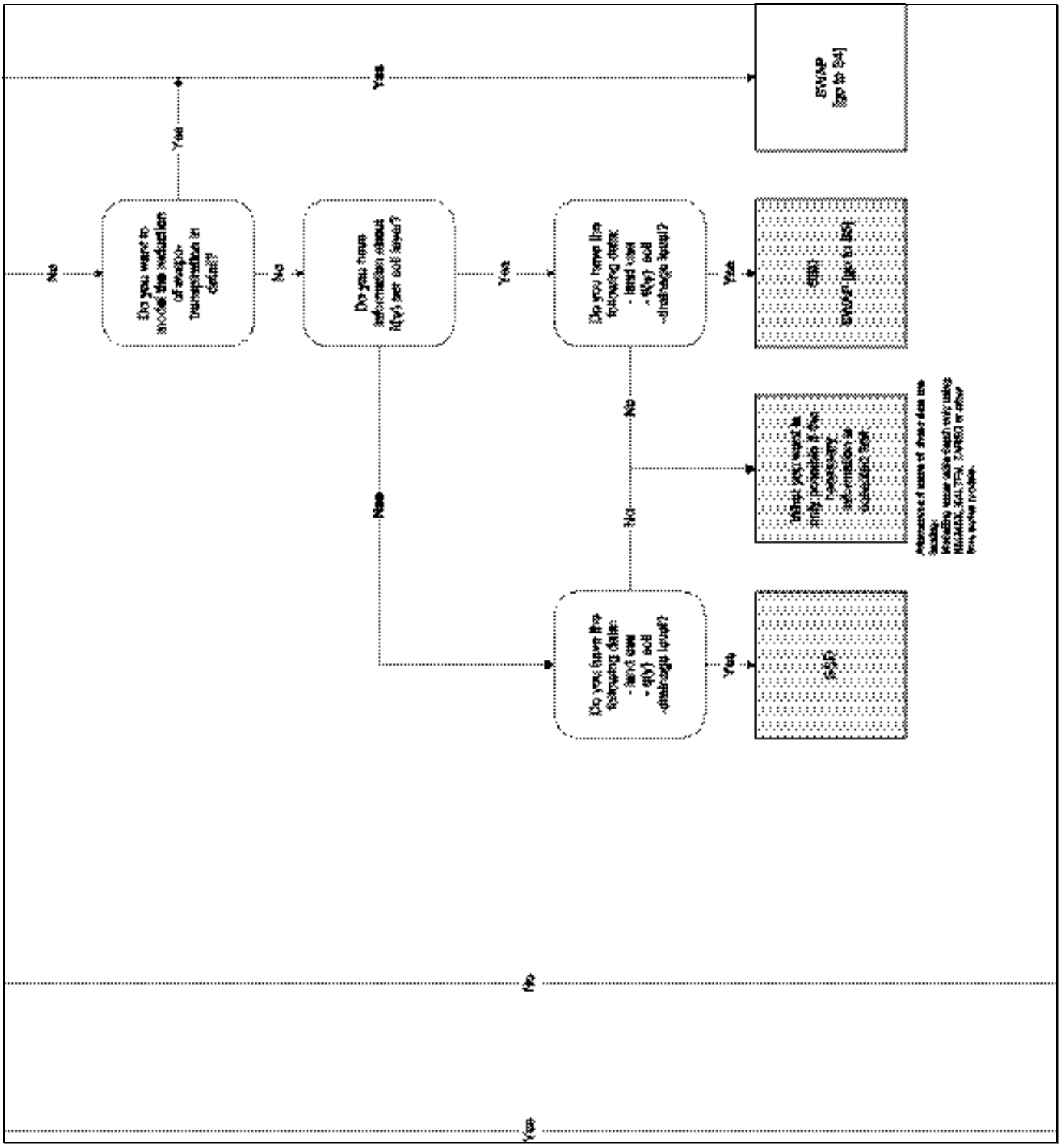

All models
Do you have a time series of water table depth?
Yes
No
Do you have time series of precipitation and evapo-transpiration?
Yes
No
Are additional temporally variable quantities likely to influence water table fluctuations?
Yes
No
Do you have time series of additional input variables?
Yes
No
Has the time series of water table depth been observed with constant frequency and is the interval between its observations equal to the standard interval between the observations of water table depth?
Yes
No
Do you want to model the water table depth with smaller time steps than the time interval between observations?
Yes
No
Are there any threshold non-linearities and do you wish to take these into account?
Yes
No
Has the time series of water table depth been observed with constant frequency and is the interval between its observations equal to the standard interval between the observations of water table depth?
Yes
No
Yes
No

No
Yes
Yes
No
Yes
No



No
Apart from the water table depth, does the list of target variables also include specific drainage attributes, bottom slope for drainage assessment or effective porosity?
No
Apart from the precipitation surplus, are there additional variables that influence the water table fluctuation?
Yes
KALTEN
No
KALMAX
Yes
Is the drainage level (bottom trench or bottom water level) with respect to surface elevation known?
Yes
Apart from the precipitation surplus, are there additional variables that influence the water table fluctuation?
Yes
KALTEN
Complete measurements
No
KALMAX
Complete measurements
No
Probably not useful or only possible if the necessary information is collected first
alternatives: modelling water table depth only with KALMAX or KALTEN


SSG
EMERALD
SWAP
Do you have a time series of water table depth?
Yes
No
Will a referring boundary met drainage/stream of variable upper/lateral water standard?
Yes
No
S2
S3
Is the water table fluctuating at depths larger than 2 m below the surface?
Yes
No
Is the water table fluctuating at depths larger than 2 m below the surface?
Yes
No
Alternative: model water table depth only using unitarizable time series analysis; extrapolation not possible (not present in this package: GENSTAT, MATHLAB)



Part 2: Documentation KALMAX and KALTFN

Programs for modelling the water table depth using transfer function-noise models

1 Model, input variables and parameters

1.1 Model description

An extensive description of the models and their use can be found in:
English
Bierkens, M.F.P., M. Knotters and F.C. Van Geer, 1999. Calibration of transfer function-noise models to sparsely or irregularly observed time series. *Water Resources Research* 35(6), 1741-1750.
Dutch
Bierkens, M.F.P, M. Knotters and F.C. Van Geer, 1999. Tijdreeksanalyse nu ook toepasbaar bij onregelmatige meetfrequenties. *Stromingen* 5(2), 43-54.

Here follows a brief description of the models and their parameters.
The basic model used in KALTFN is the following transfer function-noise model:

$$h_k = h_k^* + n_k \tag{1}$$
$$h_k^* = a h_{k-1}^* + b P_k \tag{2}$$
$$n_k = c + f(n_{k-1} - c) + e_k \tag{3}$$

with
k discrete time steps of size Δt, i.e. $t = k\Delta t, k = 0,1,2,...$;
h_k water table depth [L] at time step k;
h_k^* deterministic part of the water table depth [L] at time step k;
n_k auto-regressive noise process [L] at time step k;
e_k zero mean discrete Gaussian white noise process at time step k;
P_k precipitation surplus, i.e. average precipitation minus average potential evapotranspiration between $(k-1)\Delta t$ and $k\Delta t$.

Equation (1) shows that the water table depth consists of a "deterministic" part (Equation (2)), which is explained by variations of precipitation surplus, and a "stochastic" part (Equation (3)) which describes the part that is not explained by precipitation surplus. Equation (2) is called the "transfer function model", and Equation (3) the noise model. The total model is therefore called a transfer function-noise model (TFN model). A TFN model is therefore a stochastic model and h_k a stochastic process. In chapter 2 the nature of stochastic processes is described.

The following parameters are distinguished:
a auto-regressive parameter of the transfer model [-];
b moving average parameter of the transfer model [T];
c average level of the noise process, which is the average water table depth in case $P_k = 0$ for all k;

f auto-regressive parameter of the noise model [-];

s_e^2 variance of the white noise process e_k [L^2].

The KALMAX model is somewhat simpler. Here we have that $f = a$ so that the TFN model (1)-(3) reduces to a so called auto-regressive exogenous variable model (ARX model):

$$h_k = c + a(h_{k-1} - c) + bP_k + e_k \tag{4}$$

Knotters and Bierkens (2000) give a physical interpretation of the ARX model (and the transfer component of the TFN model). Figure 1 shows a simplified water balance of a soil column. Apart from the precipitation surplus P_k and the water table depth h_k the following variables and parameters determine the water balance of a soil column (see Figure 1):

h_s the drainage level (surface water level or trench bottom) [L];

$q_{d,k}$ the flux from/to the surface waters [LT^{-1}], which is assumed to be linearly related to the water table depths as:

$$q_{d,k} = \frac{h_k - h_s}{g} \tag{5}$$

g the drainage resistance [T];

m effective porosity (specific yield) [-];

q_v flux from/to deeper groundwater system [LT^{-1}];

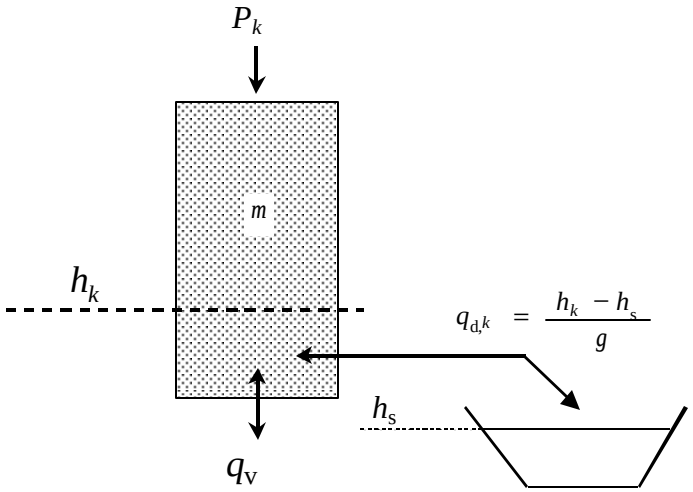


Figure 1. Schematic representation of the water balance of a soil column

Based on the simple model of Figure 1 Knotters and Bierkens (2000) derived the following relationships between the ARX (transfer) parameters and the hydrological parameters above:

$$a = \exp\left(\frac{-\Delta t}{fg}\right)$$
$$b = g\,(1-a) \qquad (6)$$
$$c = gq_{\mathrm{v}} + h_{\mathrm{s}}$$

So, if the ARX model (4) or the transfer model (2) is calibrated to a time series of water table depth and the local drainage level h_s is known, we can invert relationships (6) to calculate hydrological parameters that characterise the local water balance and groundwater dynamics:

$$g = \frac{b}{1-a}$$
$$f = \frac{-\Delta t}{g \ln a} \qquad (7)$$
$$q_{\mathrm{v}} = \frac{c - h_{\mathrm{s}}}{g}$$

From the time averaged precipitation excess $<P_k>$ the time averaged water table depth $<h_k>$ can be calculated, using either the hydrological parameters or the ARX parameters:

$$< h_k > = h_s + g\left[q_{\mathrm{v}} + < P_k >\right] = c + \frac{b}{1-a} < P_k > \qquad (8)$$

1.2 Input variables and parameters / invoervariabelen en parameters

Next we list the input variables and parameters used in the programs KALMAX and KALTFN (units between square brackets). First the English version is given, thereafter the Dutch version.

English:
Input variables

P_k average precipitation surplus (mm/d) between $(k-1)\Delta t$ and $k\Delta t$, calculated as the difference between precipitation P_k^* and potential evapo-transpiration E_k^{p}:

$$P_k = P_k^* - E_k^{\mathrm{p}} \qquad (9)$$

Because we are dealing with statistical models, we are generally satisfied with using the observed values of precipitation and potential evapo-transpiration from nearby meteorological stations. In the Netherlands this means that for potential evapo-

transpiration the so called reference crop evapo-transpiration of Makkink E_k^{r} is used (see Winter et al., 1995), which gives the transpiration of a full grass cover under optimal conditions of water supply. For very different forms of land use (e.g. forest), or if one seeks a physical interpretation of the ARX parameters (Equations 7) it may be worthwhile to correct the precipitation surplus for rainfall interception and different potential evapo-transpiration:

$$P_k^{*} = (1 - F_{\mathrm{i}}) P_k^{**} \tag{10}$$

and

$$E_k^{\mathrm{P}} = F_{\mathrm{c}} E_k^{\mathrm{r}} \tag{11}$$

where

F_{i} interception fraction (-);

F_{c} crop factor (-);

P_k^{**} precipitation (mm/d) as observed by the meteorological station without interception.

Table 1 list for a number of different land use classes the crop factor F_{c} (Equation 5) as well as the interception fraction F_{i}. Note that Equations (10) and (11) are not executed in KALMAX or KALTFN. Thus, if one desires to correct precipitation for interception and evapo-transpiration for crop type, these corrections should be performed outside KALMAX or KALTFN and the corrected figures put into the input file.

Parameters KALMAX (ARX model (4))

a auto-regressive parameter (possibly between -1 and 1; typically between 0.9-0.99);

b moving average parameter (days). Values typically range between 0.1 and 1 if P_k is in mm/d;

c level parameter (cm with respect to surface, e.g. -130);

s_e^2 variance of the white noise process e_k (cm^2). Typical values range between 5 and 50 cm^2;

s_y^2 variance of measurement error (cm^2); only used in case the Kalman filter is used (see chapter 2). In most time series analyses set to zero.

Parameters KALTFN (TFN model (1)-(3))

a auto-regressive parameter of the transfer model (possibly between -1 and 1; typically between 0.9-0.99);

b moving average parameter of the transfer model (days). Values typically range between 1 and 10;

c level parameter (cm with respect to surface, e.g. -130);

f auto-regressive parameter of the noise model (between 0 and 1); Typical values range between 0.5 and 0.99;

s_e^2 variance of the white noise process e_k (cm^2). Typical values range between 5 and 50 cm^2;

s_y^2 variance of measurement error (cm^2); only used in case the Kalman filter is used (see chapter 2). In most time series analyses set to zero.

Nederlands:

Invoervariabelen

P_k gemiddeld neerslagoverschot (mm/d) tussen $(k-1)\Delta t$ and $k\Delta t$, berekend als het verschil tussen neerslag P_k^* and potentiële verdamping E_k^{p}:

$$P_k = P_k^* - E_k^{\mathrm{p}} \tag{9}$$

Omdat het hier gaat om statistische modellen is het meestal voldoende om te werken met de waarden van neerslag en potentiële verdamping zoals die worden gemeten op nabijgelegen meteostations. In Nederland betekent dit dat voor de potentiële verdamping de Makkink referentieverdamping E_k^{r} wordt gebruikt. Deze geeft de verdamping van gras dat optimaal van water wordt voorzien. Als men rekening wil houden met andere vormen van landgebruik en de ARX of TFN parameters fysisch wil interpreteren (zie Vergelijking (7)) dan kan het de moeite waard zijn om de neerslag te corrigeren voor interceptie en de verdamping te corrigeren met een gewasfactor:

$$P_k^* = (1 - F_{\mathrm{i}})P_k^{**} \tag{10}$$

en

$$E_k^{\mathrm{p}} = F_{\mathrm{c}} E_k^{\mathrm{r}} \tag{11}$$

waarbij

F_{i} interceptiefractie (-);

F_{c} gewasfactor (-);

P_k^{**} neerslag (bij voorkeur cm/d) zoals waargenomen op het meteorologisch station.

Tabel 1 geeft voor een aantal verschillende landgebruiksklassen de gewasfactor F_{c} en de interceptiefractie F_{i}. We merken hier op dat vergelijkingen (10) en (11) niet in KALMAX of KALTFN zijn geprogrammeerd, zodat deze berekeningen buiten KALMAX en KALTFN moeten worden uitgevoerd en als gecorrigeerde waarden aan deze programma's aangeboden door ze in de invoerfile te zetten.

Parameters KALMAX (ARX model (4))

a autoregressieve parameter (-) (kan tussen -1 en 1 liggen, maar ligt meestal tussen 0.9 en 0.99);

b *moving average* parameter (dagen). Waarden variëren meestal tussen 0.1 en 1 als P_k in mm/d is;

c niveauparameter (cm referentieniveau, bijv. -130);

s_e^2 variantie witte ruis proces e_k (cm^2). Waarden variëren meestal tussen 5 en 50 cm^2;

s_y^2 variantie meetfout (cm^2); wordt alleen gebruikt bij toepassing van het Kalman filter is (zie hoofdstuk 2). Bij tijdreeksanalyse is deze meestal gelijk aan 0.

Parameters KALTFN (TFN model (1)-(3))

a autoregressieve parameter van het transfermodel (-) (kan tussen -1 en 1 liggen, maar ligt meestal tussen 0.9 en 0.99);

b *moving average* parameter van het transfermodel (dagen). Waarden variëren meestal tussen 0.1 en 1 als P_k in mm/d is;

c niveauparameter (cm referentieniveau, bijv. -130);

f autoregressieve parameter van het ruismodel (kan tussen 0 en 1 liggen, maar ligt meestal tussen 0.9 en 0.99);

s_e^2 variantie witte ruis proces e_k (cm^2). Waarden variëren meestal tussen 5 en 50 cm^2;

s_y^2 variantie meetfout (cm^2); wordt alleen gebruikt bij toepassing van het Kalman filter is (zie hoofdstuk 2). Bij tijdreeksanalyse is deze meestal gelijk aan 0.

Table 1. Crop factors and interception fractions for various crops and land use (sources in footnotes)

Crop/land use	F_c	F_i
Grassland	0.96[1]	0.00
Potatoes	1.03[1]	0.00
Beets	0.98[1]	0.00
Grain	0.95[1]	0.00
Maize	0.92[1]	0.00
Other crops	0.88[1]	0.00
Fallow	0.70[2]	0.00
Deciduous wood	1.00[3]	0.20[3]
Coniferous wood	0.80[3]	0.40[3]
Tree nurseries	1.00[3]	0.20[3]
Other trees	0.90[3]	0.30[3]
Heath	0.70[2]	0.00
Wetland vegetation	1.04[4]	0.00
Dryland vegetation	0.70[2]	0.00
Other vegetation	0.87[5]	0.00
Orchard (soil not covered)	0.91[6]	0.06[7]
Orchard (grass strips)	0.97[6]	0.06[7]
Sports field	0.96[8]	0.00
Public garden	0.97[9]	0.06[9]
Horticulture under glass	1.30[10]	1.00
Horticulture not under glass	0.88[11]	0.00

[1] Crop factor of average growing season according to Feddes, R.A., 1997. Crop factors in relation to Makkink reference-crop evapo-transpiration. In: Verslagen en Mededelingen 39, pp. 33-45, CHO-TNO, The Hague.

[2] Jansen, P.C., 1986. De potentiële verdamping van (half-)natuurlijke vegetaties. ICW nota 1703, Wageningen (in Dutch).

[3] Moors, E.J., A.J. Dolman, W. Bouten en A.W.L. Veen, 1996. De verdamping van bossen H_2O 19(16), 462-466 (in Dutch). Furthermore, the parameters for "tree nurseries" have been taken the same as for deciduous wood, and those for "other wood" as the average of deciduous and coniferous wood.

[4] Seasonal average of Molinia from Moors, E.J., J.N.M. Stricker and G.D. van den Abeele, 1998. Evapo-transpiration of cut over bog covered by Molinea Caerulea. Agricultural University, Department of Environmental Sciences, report 73, Wageningen.

[5] Average of wetland and dryland vegetation.

[6] Assuming a tree coverage of 30% we calculate:

Orchard (soil not covered): $F_c = 0.7 \times 0.88$ (other crops) + 0.3×1.0(deciduous trees) = 0.91

Orchard (grass strips): $F_c = 0.7 \times 0.96$ (grassland) + 0.3×1.0(deciduous trees) = 0.97

[7] Assuming a tree coverage of 30% we also assume 30% of the interception of deciduous wood yielding an interception factor of $0.3 \times 0.20 = 0.06$.

[8] Taken the same as "grassland".

[9] Taken the same as "orchard (grass strips)"

[10] All water is intercepted. The evapo-transpiration for horticulture under glass in the western part of the Netherlands is about 700 mm/year, which is about 1.3 times the average reference evapo-transpiration for that area (personal communication Philip Hamaker). The excess water needed is obtained from the intercepted precipitation water that is collected in tanks.

[11] Taken the same as "other crops".

2 Stochastic modelling: prediction, simulation and calibration

There are two different ways of using the stochastic models: prediction and stochastic simulation. In the following sections these two different ways are further explored. The section thereafter explains how the ARX or TFN parameters can be estimated from time series of water table depth through calibration of the ARX or TFN models. Finally, the last section recapitulates how the application of the ARX/TFN models to a practical problem proceeds.

2.1 Prediction

Before we start with explaining prediction, we have to explain the nature of h_k when it is described with stochastic models. It means that we are uncertain about the exact variation of h with time. We do know that it is likely to be lower in the summer time and higher during the winter, but there is still a lot of unknown variation left. This variation is due to errors in our inputs, model parameters and the fact that our model itself is only an approximation of reality. Therefore, we vision that h_k is not described by a single function of time, but as a collection of possible functions, each of which are equally probable of describing the real but unknown variation of h. Figure 2 illustrates this concept, showing four equally possible functions (note that for convenience we have drawn them as continuous functions).

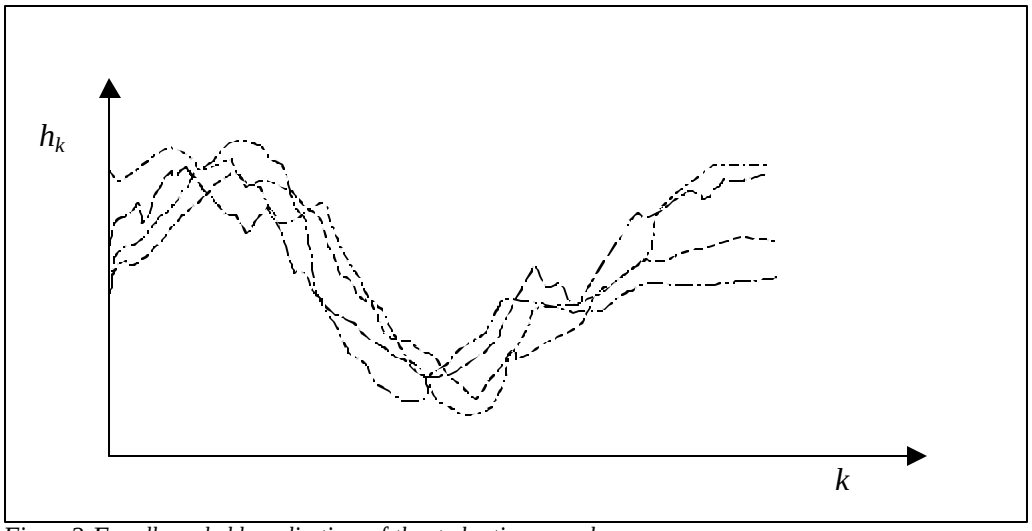


Figure 2 Equally probable realisations of the stochastic process h_k

One particular function is called “a realisation” and the whole collection of equally probable realisations, usually an infinite number of them, is called “the ensemble” and h_k is said to be “a stochastic process”. Now reality is assumed to be on of these possible realisations, however which one exactly is unknown. We would want our prediction to be such that the prediction error is minimal. However, because we do

not know which of the realisations is reality it is not possible to evaluate the prediction error either. Instead, we are forced to look at every time step at the difference between our prediction and the values of all possible realisations. Suppose that $\hat{h}_k$ is the prediction and $h_k^{(i)}$ the value of realisation number i, then we seek a prediction for which the following properties hold:

1. The prediction error, i.e. the difference between our prediction and the realisations, is on average equal to zero:

$$\sum_{i=1}^{\infty}\left[\hat{h}_k - h_k^{(i)}\right] = 0 \tag{12}$$

2. The average squared prediction error is minimal (as small as possible):

$$\sum_{i=1}^{\infty}\left[\hat{h}_k - h_k^{(i)}\right]^2 \Rightarrow \text{minimal} \tag{13}$$

It turns out that we achieve this if we take as prediction at every time the average of all realisations:

$$\hat{h}_k = E[h_k] = \sum_{i=1}^{\infty} h_k^{(i)} \tag{14}$$

The average of all possible realisations is called "the expectation" or "expected value" of the stochastic process h_k and is usually denoted with the operator $E[]$. If a prediction is such that it has properties (12) and (13), it is said to be "unbiased" (Equation 12) and "optimal in least squared sense" (Equation 13). If we describe the stochastic process h_k with the TFN model or the ARX the expected value is the deterministic part of these models:

$$\hat{h}_k = c + a(\hat{h}_{k-1} - c) + bP_k \tag{15}$$

Figure 3 shows the realisations with the optimal prediction. Also shown is the 95%-prediction interval which gives for every time step the boundaries that contain 95% of the realisations. Because the noise e_k has a Gaussian (normal) distribution, the 95% prediction interval can be calculated from the standard deviation of the prediction error $s_e(t)$ as follows:

$$\left[\hat{h}_k \pm 1.96 s_e\right]$$

with

$$s_e = \sqrt{\sum_{i=1}^{\infty}\left[\hat{h}_k - h_k^{(i)}\right]^2} \tag{16}$$

For the TFN model the variance of the prediction error s_e^2 is equal to the variance of the noise process n_k (s_n^2) and can be calculated as (Bierkens *et al.*, 1999):

$$s_e^2 = s_n^2 = \frac{s_e^2}{1-f^2} \tag{17}$$

For the ARX model this variance is given by

$$s_e^2 = \frac{s_e^2}{1-a^2} \tag{18}$$

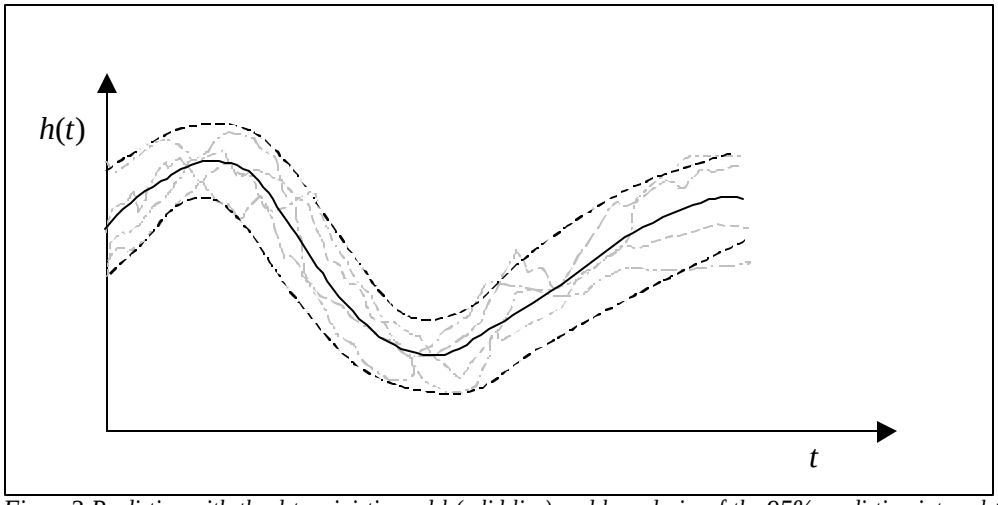


Figure 3 Prediction with the deterministic model (solid line) and boundaries of the 95% prediction interval (dashed lines); grey lines are a number of realisations.

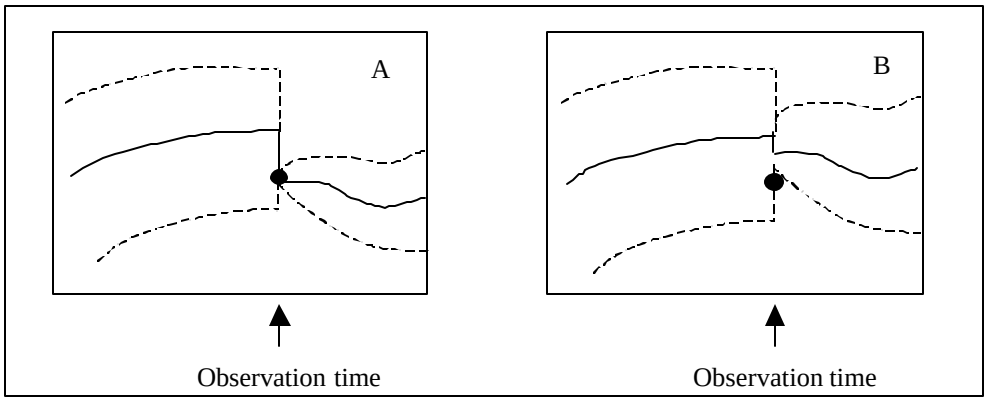


Figure 4 Correcting a prediction (solid line) with an observation without an observation error (Figure 4a) and with an observation error (Figure 4b). The dashed lines show the 95% prediction interval, which will be of zero width at the measurement time in case 4a and of limited width in case 4b.

In case we have observations of h_k we can use these to further improve the predictions. This is shown in Figure 4. Here we have a clue about the realisation that stands for reality because it has been observed a number of times. If the realisations

were without error, than the best strategy would be to simply set the predicted value equal to the observation as soon as the observation comes in (Figure 4a) and proceed from there. At that time the prediction error would be zero. If there is an observation error, the truth would be somewhere in between and an improved prediction would be some weighted average of the prediction and observation. The prediction error would than be somewhat smaller than the observation error (Figure 4b).

An algorithm that performs such a correction is called the "Kalman filter". Suppose that $\hat{h}_{k|k-1}$ is the prediction at time step k that has been obtained using (15) and corrections for all observations up to and including time step k-1 and at time step k an observation y_k is available. The corrected prediction is then obtained as follows:

$$\hat{h}_{k|k} = \left(1 - \frac{s^2_{k|k-1}}{s^2_{k|k-1} + s^2_{y_k}}\right)\hat{h}_{k|k-1} + \left(\frac{s^2_{k|k-1}}{s^2_{k|k-1} + s^2_{y_k}}\right)y_k \tag{19}$$

where

$\hat{h}_{k|k-1}$ the "time update", which is the optimal prediction at time step k given all observations up to including time step k-1. The time update is such that the difference between the time update and all realisations is on average zero and that the variance of this difference is minimal:

$$\sum_{i=1}^{\infty}\left[\hat{h}_{k|k-1} - h_k^{(i)}\right] = 0 \tag{20}$$

$$s^2_{k|k-1} = \sum_{i=1}^{\infty}\left[\hat{h}_{k|k-1} - h_k^{(i)}\right]^2 \Rightarrow \text{minimal} \tag{21}$$

$s^2_{k|k-1}$ the variance of the error in the time update as defined in (21), which is calculated by the Kalman filter;

$\hat{h}_{k|k}$ the "measurement update", which is the optimal prediction at time step k given all observations up to including time step k. The measurement update is such that the difference between the measurement update and all realisations is on average zero and that the variance of this difference is minimal:

$$\sum_{i=1}^{\infty}\left[\hat{h}_{k|k} - h_k^{(i)}\right] = 0 \tag{22}$$

$$s^2_{k|k} = \sum_{i=1}^{\infty}\left[\hat{h}_{k|k} - h_k^{(i)}\right]^2 \Rightarrow \text{minimal} \tag{23}$$

$s^2_{k|k}$ the variance of the error in the measurement update as defined in (23) which is also calculated by the Kalman filter;

$s^2_{y_k}$ the variance of the error in the observation.

the time update is calculated by application of the deterministic model (15) between time steps k-1 and k with measurement update $\hat{h}(k-1|k-1)$ as initial condition:

$$\hat{h}_{k|k-1} = c + a(\hat{h}_{k-1|k-1} - c) + bP_k \tag{24}$$

From Equation (19) the workings of the Kalman filter become clear. If the observation error is small compared to the model prediction error (i.e. the time update), the updating in (19) will be such that the measurement update is close to the value of the observation. On the other hand, if the error in the time update is small compared to the observation error, i.e. the model is very precise when compared to the observations, the measurement update resembles very much the time update. The weight used in the weighted average (19), i.e. $[s^2_{k|k-1}/(s^2_{k|k-1} + s^2_{y_k})]$ is called the "Kalman gain".

If at time step k there are no observations available, we simply have:

$$\hat{h}_{k|k} = \hat{h}_{k|k-1} \tag{25}$$

$$s^2_{k|k} = s^2_{k|k-1} \tag{26}$$

If for a long time no observations are taken we have that $s^2_{k|k} = s^2_{k|k-1} \rightarrow s^2_e$, i.e. the Kalman filter prediction variances become equal to the prediction variance of the deterministic models, as given in Equations (17) (TFN) or (18) (ARX).

Figure 5 is similar to Figure 3, but now we have a number of observations present and predictions are performed with the Kalman filter. For convenience, we have shown the case with no observation errors. At a number of time steps observations are present. Again, our unknown reality is modelled as a large number of equally probable realisations. However, because we have observations, only those realisations are eligible that pass through the observations (in case of a small observation error, the realisations do not have to pass through the observations exactly, but should pass close by). The time update is shown, as well as the 95% prediction interval. Because the predicted water table depth is updated, the variance of the prediction error is smaller than without updating (Figure 3); to put it in another way: the averaged squared difference between the prediction (time update) and the realisations is smaller than without the use of the Kalman filter.

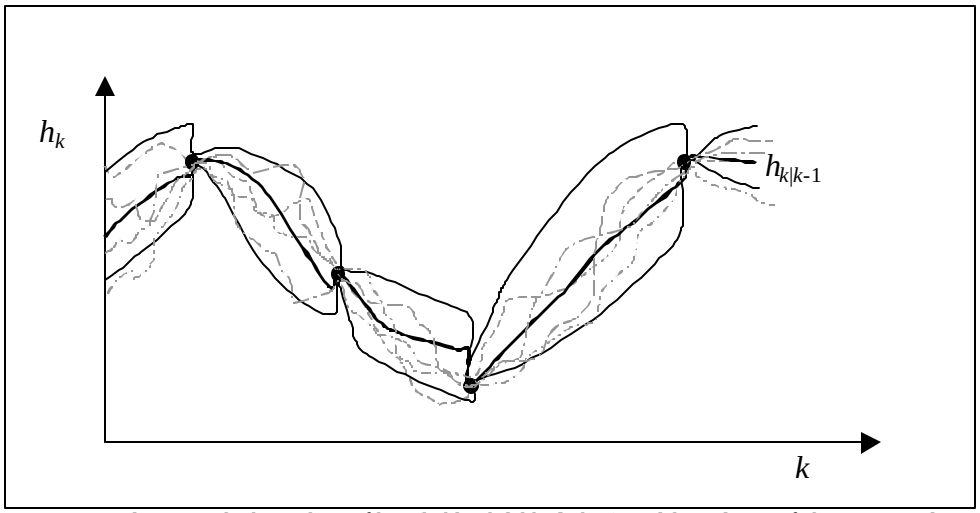


Figure 5 Prediction with the Kalman filter (bold solid black line) and boundaries of the 95% prediction interval (thin solid black lines); grey lines are a number of realisations, all passing through the observations; black dots are observations.

2.2 Stochastic simulation

Prediction of water table depth is important if we are interested in the actual, but non-observed water table depth. However, in many applications we are not particularly interested in the actual water table depth, but in some fluctuation quantities. For instance, we may be interested in the probability that at any day in the near future the level of 30 cm is exceeded, or we want to know the mean highest water table depth (MHW) or mean lowest water table depth (MLW) (Van der Sluijs and De Gruijter, 1985). These fluctuation quantities cannot be estimated from the predicted water table depth, because the predicted line typically overestimates the low values and underestimates the high values (see Figures 3 and 5). The reason for this is that it tries to minimise the squared prediction error (Equation 16).

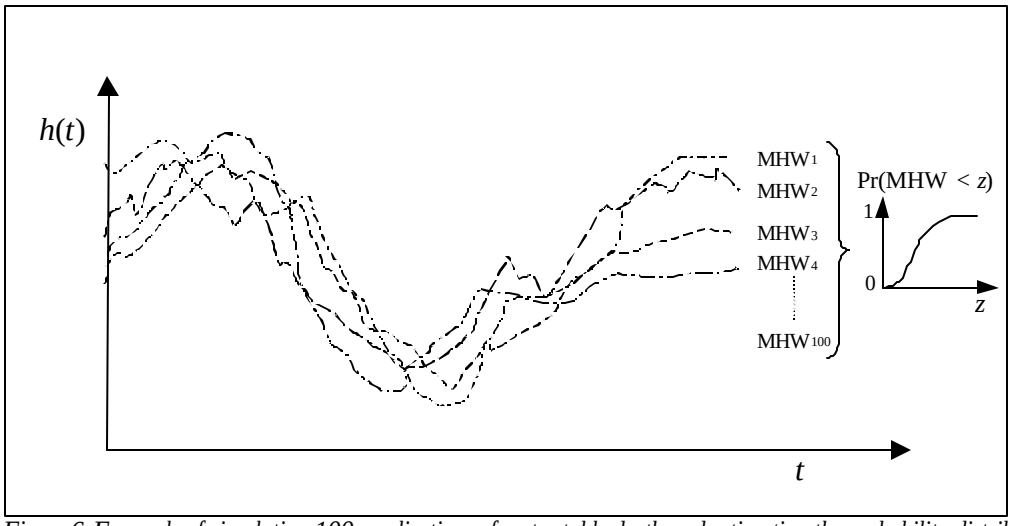


Figure 6 Example of simulating 100 realisations of water table depth and estimating the probability distribution of MHW from it.

In this case we should do the following (See Figure 6):

1. simulate a large number (at least 100) realisations of the stochastic process h_k using either the TFN model (1-3) or the ARX model (4) where the e_k are simulated by drawing from a Gaussian distribution with mean zero and variance s_e^2;
2. for each simulated realisation the appropriate fluctuation quantities are estimated. For 100 realisations this yields also 100 values of this property, e.g. 100 MHWs and 100 MLWs;
3. the cumulative frequency estimated from the replicas of the fluctuation quantity (e.g. the 100 MHWs and MLWs) represents an estimate of the probability distribution of the fluctuation quantity. This probability distribution expresses the uncertainty about the true value of the fluctuation quantity, uncertainty that arises from our model's inability to predict the unknown water table depth exactly. Usually the average of the replicas is used as an estimate of the true but unknown fluctuation quantity and as a measure of uncertainty the 95% confidence interval is calculated.

2.3 Calibration

To apply the ARX or TFN models at some location parameters $\{a,b,c,s_e^2\}$(ARX) or $\{a,b,c,f,s_e^2\}$(TFN) must be obtained by calibrating the models to a time series of water table depth. In case of a deterministic model, calibration only involves the minimisation of some least squares criterion between predictions and observations. However, in this case we also have to calibrate the noise parameters f and s_e^2. For linear stochastic models this parameter could be estimated separately from the residuals (differences between deterministic predictions and observations). However, even for linear models, such a two step approach generally leads to biased parameter

estimates (see Te Stroet, 1995). Also, if the time series is very irregular (not a constant frequency), estimating f and s_e^2 is quite problematic. Here, a method is used combining the Kalman filter and a maximum likelihood criterion that, given the assumptions (model prediction errors and measurement errors are Gaussian distributed), provides unbiased maximum likelihood estimates of the parameters. The method can be used for irregularly and sparsely observed time series and has the added advantage that it able to take account of measurement errors.
If we have M time steps with observations (not necessarily with regular intervals between them), the method proceeds as follows (we assume that the input time series, i.e. precipitation and potential evapo-transpiration have been collected for the period for which we have observations):

1. choose initial values for parameters $\{a, b, c, s_e^2\}$(ARX) or $\{a, b, c, f, s_e^2\}$(TFN);
2. use these parameter values to run the Kalman filter for the period that contains the observations. This yields for each observation occasion an "innovation" $n_l = y_l - \hat{h}_{l|l-1}$ and the associated innovation variance (calculated by the Kalman filter, see Bierkens, 1998) $s_{n_l}^2 = s_{l|l-1}^2 + s_{y_l}^2$;
3. From the M innovations and innovation variances the following maximum likelihood criterion is evaluated (Schweppe, 1973):

$$J = M \ln(2p) + \sum_{l=1}^{M} \ln\left[s_{n_l}^2\right] + \sum_{l=1}^{M} \left[\frac{n_l^2}{s_{n_l}^2}\right] \tag{27}$$

4. Choose a new set of parameters $\{a, b, c, s_e^2\}$(ARX) /$\{a, b, c, f, s_e^2\}$(TFN);
5. Repeat steps 2 to 4 until ML-criterion (27) is minimised. The resulting parameter set is a maximum likelihood estimate of the parameters.

Minimisation of (26), i.e. choosing the new set of parameters in step 3 such that the value of (26) will decrease, is done with a minimisation algorithm. In SSD the "downhill simplex method" is used (Press et al., 1986).

2.4 Systematic application of KALMAX and KALTFN: step by step

In the following example we assume that we have a location for which precipitation and reference evapo-transpiration are known for the years 1970-1999 and we have observations of water table depth available with an observation frequency of two times a month for the years 1995-1999. Application of the ARX/TFN models has two goals. First, we want to estimate fluctuation properties (MHW, MLW etc.) at the location. Second, we aim to monitor the water table depth on a daily basis in the future, while maintaining an observation frequency of two times a month. The following steps are taken:

2.4.1 Calibration

The parameters $\{a, b, c, s_e^2\}$ (KALMAX) or $\{a, b, c, f, s_e^2\}$ (KALTFN) are estimated through calibrating the models to the observations from 1995-1999 (section 2.3). Usually it is assumed that the observations are without error, as in time series analysis. However, an observation error variance of 1-4 cm^2 is appropriate for most piezometer data. Although in theory the parameters of the deterministic model part $\{a, b, c\}$ and the parameters of the stochastic part $\{f, s_e^2\}$ should be calibrated simultaneously, it is recommended to use a three-step procedure in KALMAX or KALTFN (in VIDENTE these three modes can be chosen from in the Calibration menu):

1. first, the parameters $\{a, b, c\}$ are calibrated (called "deterministic calibration" in VIDENTE). This achieved by fixing the parameter $s_e^2 = 0$ during the calibration (in case of KALTFN fix $f = 0.95$) and setting the observation error variance $s_{y_k}^2 = 1$ for all time steps;
2. in the second step, called the "stochastic calibration" the parameters $\{a, b, c\}$ are fixed at the values obtained in step 1 and the parameters s_e^2 (KALMAX) or $\{f, s_e^2\}$ (KALTFN) are calibrated while setting $s_{y_k}^2 = 0$ for all time steps;
3. finally, using the parameter values found in steps 1 and 2 as initial estimates, all parameters $\{a, b, c, s_e^2\}$ (KALMAX) or $\{a, b, c, f, s_e^2\}$ (KALTFN) are calibrated in the last step, thus making sure that unbiased estimates of these parameters are found ($s_{y_k}^2$ set at correct value). If the adjustment in step 3 is large, i.e. if very different values of the parameters are found, we must be suspicious and perhaps decide to use the parameters found in steps 1 and 2 and forget step 3.

The result of the calibration with KALMAX/KALTFN is a file with calibrated parameters and the mean error (*ME*), root mean squared error (*RMSE*) and mean absolute error (*MAE*) of Kalman filter time updates:

$$ME = \frac{1}{M}\sum_{k=1}^{M}\left(\hat{h}_{k|k-1} - y_k\right) \tag{28}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{k=1}^{M}\left(\hat{h}_{k|k-1} - y_k\right)^2} \tag{29}$$

$$MAE = \frac{1}{M}\sum_{k=1}^{M}\left|\hat{h}_{k|k-1} - y_k\right| \tag{30}$$

Furthermore, the file with Kalman filter predictions is given, which contains the Kalman filter time updates, measurement updates, their variances, observations, innovations (see 2.3) and the innovations 95% prediction interval. Figure 7 shows an example of the output from a calibration run.

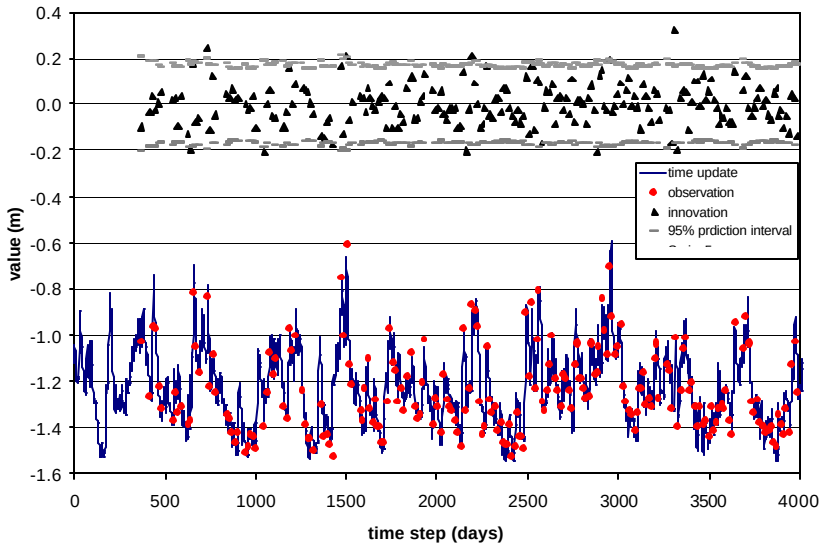


Figure 7 Example output from calibrating KALMAX or KALTFN

2.4.2 Verification: prediction without Kalman filter

After calibration, it is wise to check whether the deterministic part of the model is able to describe the dynamics of the water table depth sufficiently well. An indication that a deterministic fit is problematic would have been a strong adjustment of the parameter values in step 3 of the Calibration step. Figure 7 shows the time updates from the Kalman filter. A problem with using the Kalman filter could occur if the number of observations is large. In that case, the resulting time updates could be insensitive to the model dynamics, because updating is performed too Frequently. In that case a small value of criterion (27) could still result in poor estimates of the dynamic parameters *a* and *b*. Therefore, a prediction with the deterministic part of the model should be made, without the updating. This can be achieved by running KALMAX/KALTFN for the calibration period as a means of verification by setting $s_e^2 = 0$ and the observation error variance $s_{y_k}^2 = 1$, and fixing the values of g, e_0 and q_v at their calibrated values (in VIDENTE this is an option in the prediction menu). The result of the verification with KALMAX/KALTFN is a file with parameters used and the following verification statistics:

$$ME = \frac{1}{M}\sum_{k=1}^{M}\left(\hat{h}_k - y_k\right) \tag{31}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{k=1}^{M}\left(\hat{h}_k - y_k\right)^2} \tag{32}$$

$$MAE = \frac{1}{M}\sum_{k=1}^{M}\left|\hat{h}_k - y_k\right| \tag{33}$$

which are usually larger than the statistics from the Calibration run, because no updating is used. KALMAX/KALTFN also give the file with predictions, the observations and the residuals $\hat{h}_k - y_k$. Figure 8 shows and example of the output from a verification run, where in this case observations are present almost every day.

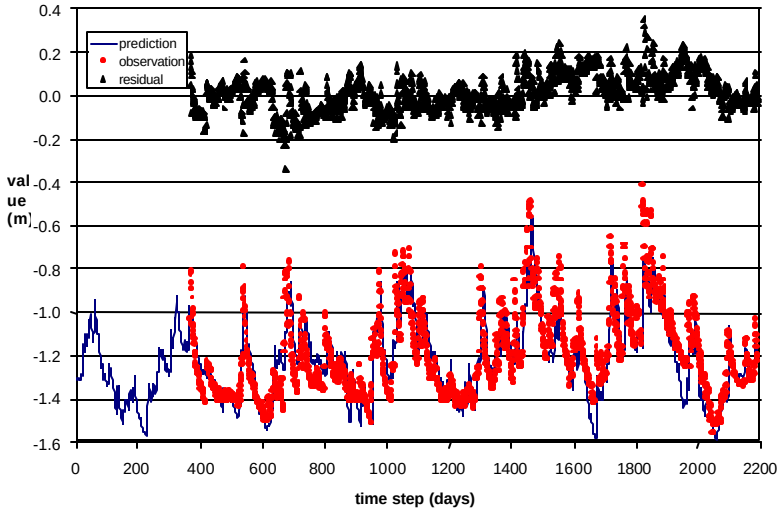


Figure 8 Example of verification of the deterministic part of the ARX or TFN model

2.4.3 Simulation

Using the calibrated parameters from the calibration step the programs KALMAX/KALTFN can be used to simulate realisations of h_k (If the local drainage level is known, these time series can be converted to simulated realisations of drainage discharge, using Equations 5 and 7). In case fluctuation quantities have to be estimated multiple realisations must be simulated. The output consists of a file containing the realisations of h_k. This file can be used as input for the program STATSIM in order to calculate the fluctuation properties. In our example we would typically simulate realisations of 30 years long, i.e. using precipitation and potential evapo-transpiration from 1970-1999, in order to obtain fluctuation quantities that are representative for the current climate, and not only for the weather conditions for the observation period 1995-1999 (see Knotters and Van Walsum, 1997).

2.4.4 Estimating fluctuation quantities

Using the program STATSIM fluctuation quantities can be calculated from the simulated realisations. The output consists of the following:

A file with statistics
An example of this output is shown hereafter in Figure 9. The mean highest water table, the mean lowest water table and the mean spring water table are given (expected values, median values and 5 and 95 percentiles). Here quantities are

calculated per realisation. So, the uncertainty here reflects only model uncertainty. The univariate statistics are statistics over all simulated water table depths, which means that they reflect both the within year variation and the year to year variation of the weather as well as the model uncertainty (variation between realisations).

Univariate Statistics

mean	-106.425819
variance	360.064484
stdev.	18.975365
3rd moment	-539.639771
P01	-148.944107
P05	-138.432495
P10	-132.007401
P25	-119.718979
P50	-105.809341
P75	-92.923340
P90	-82.159462
P95	-75.837044
P99	-65.247566

Mean Highest and Mean Lowest Water Table and Mean Spring Water Table

	mean	5%	50%	95%	std
GLG(t)	-135.358978	-138.189285	-135.264694	-132.976761	1.565006
GHG(t)	-78.744606	-81.293327	-78.789520	-76.163002	1.535088
GLG(z)	-135.058350	-138.189285	-134.869598	-132.487930	1.608089
GHG(w)	-79.471909	-82.413284	-79.414589	-76.819191	1.603910
GVG(t)	-99.525368	-103.282478	-99.257187	-96.835274	2.082639

Figure 9 File with fluctuation statistics as output of STATSIM; GHG: mean highest water table; GLG: mean lowest water table; GVG: mean spring water table; between brackets: denoting whether the statistic is determined for the whole year (t), the summer period (s) or the winter period (w).

A file with the frequency of exceeding graph (FOE-graph)

The FOE-graph gives for every level the number of days that the water table depth on any future day of any future year (at the same climate and hydrological regime as in the calibration period) exceeds that level. It is in fact the cumulative frequency distribution, reflecting both the within year variation and the year to year variation of the weather as well as the model uncertainty. Figure 10 shows an FOE-graph.

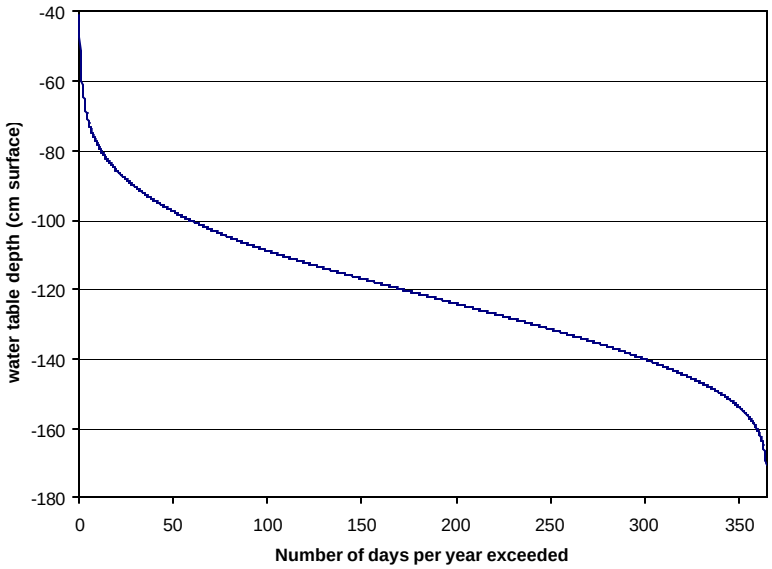


Figure 10 Frequency of exceeding graph

A file with the regime graph

The regime graph gives for every day number in a future year the expected water table depth, the median and the 5 and 95-percentiles. The variation per day number therefore reflects both year to year variation (i.e. our uncertainty about the future weather) as well as model uncertainty. Figure 11 gives an example of a regime graph.

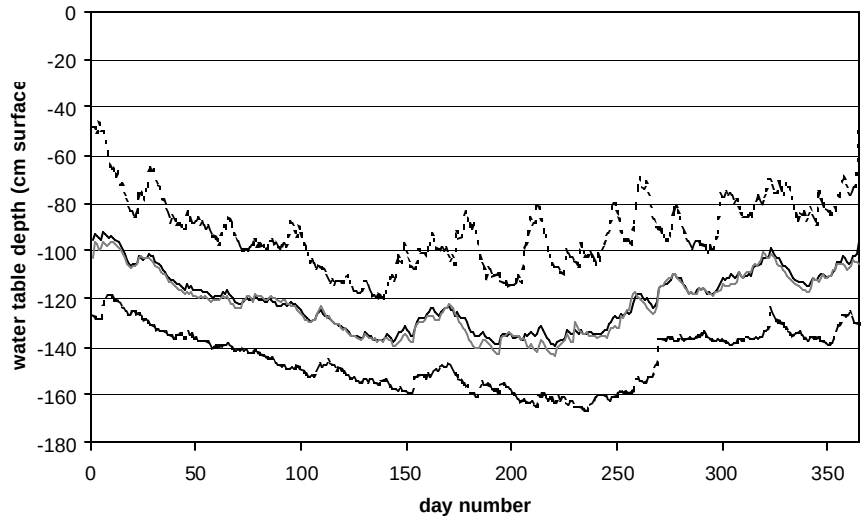


Figure 11 Regime graph; black solid line: mean; black dashed lines: 5- and 95-percentiles; grey solid line: median.

A file with the histogram

Like the FOE-graph, the histogram reflects both the within year variation and the year to year variation of the weather as well as the model uncertainty. From the histogram we can read the expected number of days that the water table depth will be within certain boundaries. Figure 12 gives an example histogram.

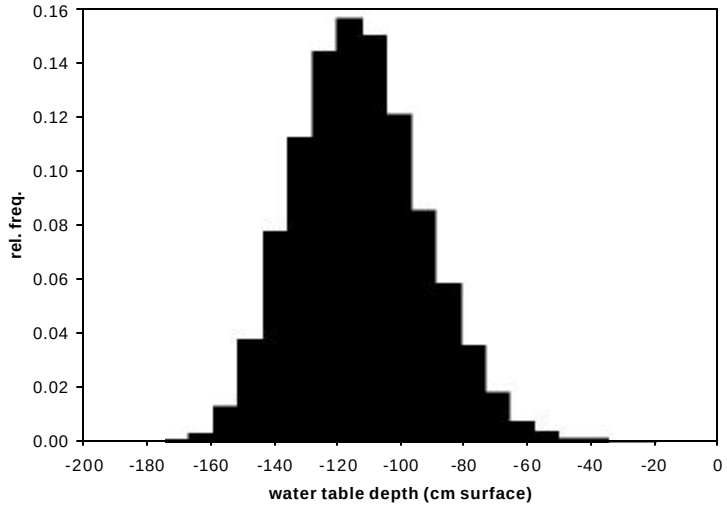


Figure 12 Example histogram

A file with the correlation function

The correlation function gives the correlation coefficient between day *k* and day *k* +1 (= lag 1), day *k* and day k+ 2 (lag 2), day *k* and day *k* + 3 etc. The correlation function reflects both the response time of the groundwater system, as well as the periodicity of the rainfall surplus. The longer it takes for the correlation function to cross the *x*-axis, the slower the response time of the groundwater system. However, because the periodicity of the rainfall surplus is also included, we cannot read the characteristic response time from the correlation function. The correlation function is the average of the correlation functions that are estimated for the realisations. An example correlation function is given below (Figure 13).

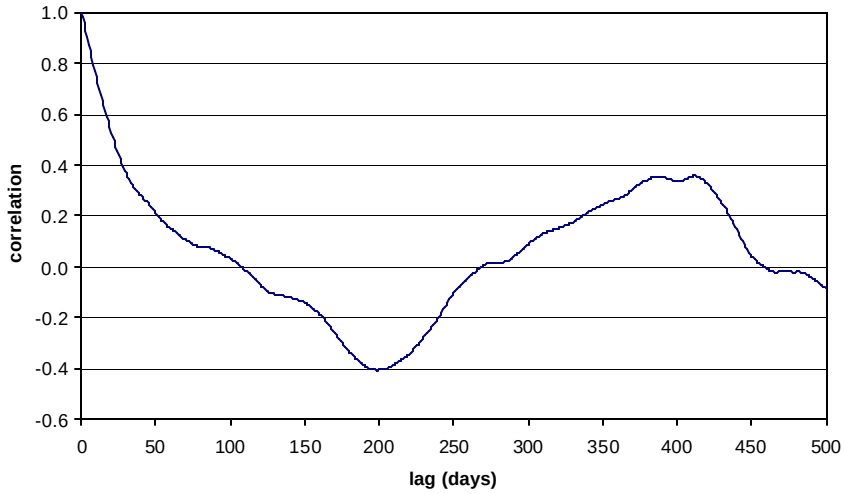


Figure 13 Example correlation function

The characteristic response time (in days) for the ARX model and the deterministic part of the TFN model is given by (Bierkens et al., 1999):

$$t_c = \frac{-3\Delta t}{\ln(a)} \tag{34}$$

The effective correlation length (time span over which n_k is correlated in days) of the noise process (3) is given by (Bierkens et al., 1999):

$$q_c = \frac{-3\Delta t}{\ln(f)} \tag{35}$$

Of course many more fluctuation quantities could be estimated from the simulated realisations; see for instance Bierkens (1998) and Knotters et al. (2000).

2.4.5 On line prediction and monitoring

Finally, using the calibrated model, the program SSDKOPT can be used for on line prediction, where in between dates that observations of water table depth are taken, optimal predictions of water table depth are obtained using the Kalman filter. The Kalman filter can also be used as a monitoring instrument by running it on line and checking whether not much more than 5% of the innovations fall outside the 95% error bounds (see Figure 7). If 5% falls outside, a change in the hydrological system may have occurred. An alternative way of monitoring such a change is running the deterministic predictions on line and plot the observations together with the prediction interval $\hat{h}_k \pm 1.96 s_e$ (with s_e calculated with Equations (17 or 18)). If the observations start to plot outside this interval, a significant change in the hydrological system may have occurred.

References

Bierkens, M.F.P., 1998. Modeling water table fluctuations by means of a stochastic differential equation. *Water Resources Research* 34(10), 2485-2499.

Bierkens, M.F.P., M. Knotters and F.C. Van Geer, 1999. Calibration of transfer-function noise models to sparsely or irregularly observed time series. *Water Resources Research* 35(6), 1741-1750.

Bierkens, M.F.P, M. Knotters and F.C. Van Geer, 1999. Tijdreeksanalyse nu ook toepasbaar bij onregelmatige meetfrequenties (Time series analysis also applicable when observation frequencies are irregular). *Stromingen* 5(2), 43-54.

Knotters, M. and P.E.V. Van Walsum, 1997. Estimating fluctuation quantities from time series of water-table depth using models with a stochastic component. *Journal of Hydrology* 197, 25-46.

Knotters, M. and M.F.P. Bierkens, 2000. Physical basis of time series models for water table depths. *Water Resources Research* 36(1), 181-188.

Knotters, M., M.F.P. Bierkens and C. Beets, 2000. Grondwaterdynamiek van vegetatiestandplaatsen; analyse van zesentwintig reeksen (Groundwater dynamics of vegetation stands; analysis of 26 time series). Rapport 095, Alterra, Wageningen.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1986. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, New York.

Schweppe, F.C., 1973. *Uncertain Dynamic Systems*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Te Stroet, C.B.M., 1995. *Calibration of stochastic groundwater flow models: Estimation of system noise statistics and model parameters.* Ph.D. thesis, Delft University of Technology.

Van der Sluijs, P. and J.J. De Gruijter, 1985. Water table classes: A method to describe seasonal fluctuation and duration of water tables on Dutch soil maps. *Agricultural Water Management* 10, 109-235.

Winter, T.C., D.O. Rosenberry, and A.M. Sturrock, 1995. Evaluation of 11 equations for determining evaporation for a small lake in the north central United States. *Water Resources Research* 31(4), 983-993.

Annex: Input instructions for KALMAX and KALTFN

Header of the program KALMAX with input instructions

```
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
C Copyright (C) 2000, Alterra, Green World Research
C
C This program is distributed in the hope that it will be useful,
C but WITHOUT ANY WARRANTY.  No author or distributor accepts
C responsibility to anyone for the consequences of using it or for
C whether it serves any particular purpose or work at all, unless he
C says so in writing. Everyone is granted permission to copy, modify
C and redistribute this program, but only under the condition that the
C the reference below is given and that this notice and the above
C copyright notice remain intact.
C
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c Program: KALMAX  -  KALman filtering of a simple arMAX model
c
c Goal:    Calibration of a simple ARMAX model of daily time steps
c          on measurements of water table depths taken at irregular
c          intervals. The program uses the output of the Kalman filter in
c          Schweppe's maximum likelihood criterium to obtain estimates of
c          the parameters of the ARMAX-model. The program can also be
c          used as stand alone Kalman filter (no optimization).
c
c Version: 1.0 (October 14th 1997)
c
c Order of ARMAX:  n=1, r=1, s=0
c
c Author:  Marc F.P. Bierkens
c Reference: Bierkens, M.F.P., M. Knotters and F.C. Van Geer, 1999.
c            Calibration of a transfer-function noise models to sparsely
c            or irregularly observed time series.
c            Water Resources Research 35(6), 1741-1750.
c**********************************************************************
c Input:
c interactive:
c          parfile - name of parameter file (input variables)
c
c**********************************************************************
c                    PARFILE
c**********************************************************************
c          record 1:  measfile - name of file with measurements of water
c                                table
c          record 2:  metfile  - name of file with precipitation and
c                                reference evapotranspiration
c          record 3:  kalfile  - name of file with Kalman filter results
c          record 4:  optfile  - name of file with results of
c                                optimization
c          record 5:  iflag    - 0: calibration mode
c                                1: - Kalman filter mode (no calibration)
c                                   - Prediction mode: set iflag to "1"
c                                     and set varw = 0 and smeas = 1
c                                2: Simulation mode
c                     (the following inputs are only used for simulation
c                      but values must be supplied even for modes 0 and
c                      1)
```

```
c                    iseed    - random seed, only used in simulation
c                               mode
c                    nsim     - number of realizations required
c                    nyears   - number of years per simulated
c                               realization
c                    year1    - first year of the simulated realizations
c          record 6:  initial values of parameters to be calibrated: if a
c                     parameter is actually calibrated it should be given
c                     a non-zero initial value!
c                     par0(1) - delta1 (-)
c                     par0(2) - omega0 (days)
c                     par0(3) - cpar (cm)
c                     par0(4) - varw (cm2)
c          record 7:  Indicators that are "1" if parameters are to be
c                     calibrated and "0" if they are not.
c                     do i = 1,4 {
c                        icalcode(i)
c                     }
c                     if all codes are zero, the Kalman filter is only
c                     run once with the parameters given in record 4.
c          record 8:  gwlev0  - initial value of groundwater level
c          record 9:
c                     smeas   - variance of measurement error (cm2)
c          record 10:
c                     dstep   - time step (days)
c                     simtim  - total simulation time (days)
c                               (in simulation mode: simtim also includes
c                                the startup time)
c                     ftol    - closure criterion used in calibration
c                               mode ftol typically has a value between
c                               0.000001 and 0.01. If the number of
c                               iterations exceeds 100 ftol should be
c                               increased.
c          record 12:
c                     startim - startup time (days)
c                               (only relevant in simulation mode; the
c                                time in days used for warming up the
c                                simulation ndays recorded in output =
c                                simtim-startim and should total the
c                                nyears per realization)
c
**********************  end parameter file*******************************
c
c*********************************************************************
c                  MEASFILE - file with measurements of heads
c*********************************************************************
c           record 1: nmeas - number of measurements
c           do i = 1,nmeas {
c              record(i+1): xmtime(i) - measurement time-step (days)
c                           xmeas(i)  - measurement (m)
c                         }
c********************** end file with measurements **********************
c
c*********************************************************************
c                  METEOFILE - file with meteo data
c*********************************************************************
c           record 1: ndays - number days to read
c                     ncols - number of columns present
c                     cfac  - multiplication factor (for unit conversion)
c                             if right unit: cfac = 1
c                     icolp - column number precipiation
c                     icolep - column number pot. evapotranspiration
```

```
c           do i = 1,ndays {
c              record(i+1): xmtime(i) - measurement time-step (days)
c                           ncols columns
c                          }
c******************** end meteofile ***********************************
c
c Output:
c*********************************************************************
c          1) kalfile
c          a) in kalman filter or calibration mode:
c                  time      - time (days)
c                  gw0       - time update (m below surface)
c                  gw1       - measurement update (m below surface)
c                  sgw0      - error variance time update (m2/day2)
c                  sgw1      - error variance measurement update
c                              (m2/day2)
c             when measurements are available:
c                  xmeas     - measured value (m below surface)
c                  (xmeas - gw0) - innovation (m below surface)
c          b) in simulation mode:
c             unformatted file (can be used as input for STATSIM)
c             record 1: year1 – first year of simulated realisation
c                       nsim – number of simulated realisations
c                      nyears – number of years per realisation
c             do k = 1,nsim
c                do i = 1,nsteps
c                   record 1 + (k-1)*nsim + i: gw1 – simulated water
c                                                    table depth
c                enddo
c             enddo
c note: nsteps = int(simtim/dstep)
c*********************************************************************
c          2) optfile:
c                  iter            - iteration number
c                  do i = 1,nwc+4 {
c                     par(i)       - parameters
c                {
c                funk(par)       - value of criterium
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

The following is an example parameter file for KALMAX

```
32cl3415.10
debilt.cal
armaxc15.k10
armaxc15.o10
0 -79359 1 5 1992
.954939 0.442178 -129.935593 2.0
1 1 1 1
-120
0.
1. 4017 0.00001
365
```

Header of the program KALTFN with input instructions

```
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
C Copyright (C) 2000, Alterra, Green World Research
C
C This program is distributed in the hope that it will be useful,
C but WITHOUT ANY WARRANTY.  No author or distributor accepts
C responsibility to anyone for the consequences of using it or for
C whether it serves any particular purpose or work at all, unless he
C says so in writing. Everyone is granted permission to copy, modify
C and redistribute this program, but only under the condition that the
C the reference below is given and that this notice and the above
C copyright notice remain intact.
C
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c Program: KALTFN  -  KALman filtering of Transfer Function Noise models
c
c Goal:    Calibration of a transfer function noise model of daily time
c          steps on measurements of water table depths taken at irregular
c          intervals. The program uses the output of the Kalman filter in
c          Schweppe's maximum likelihood criterium to obtain estimates of
c          the parameters of the TFN-model. The program can also be used
c          as stand alone Kalman filter (no optimization).
c
c Version: 0.0 (october 13th 1997)
c
c Order of TFN:  n=1, r=1, s=1, p=1, q=0, b=0
c
c Author:  Marc F.P. Bierkens
c Reference: Bierkens, M.F.P., M. Knotters and F.C. Van Geer, 1999.
c            Calibration of a transfer-function noise models to sparsely
c            or irregularly observed time series.
c            Water Resources Research 35(6), 1741-1750.
c***********************************************************************
c Input:
c interactive:
c          parfile - name of parameter file (input variables)
c
c***********************************************************************
c                    PARFILE
c***********************************************************************
c          record 1:  measfile - name of file with measurements of water
c                                table
c          record 2:  metfile  - name of file with precipitation and
c                                reference evapotranspiration
c          record 3:  kalfile  - name of file with Kalman filter results
c          record 4:  optfile  - name of file with results of
c                                optimization
c          record 5:  iflag    - 0: calibration mode
c                                1: - Kalman filter mode (no calibration)
c                                   - Prediction mode: set iflag to "1"
c                                     and set varw = 0 and smeas = 1
c                                2: Simulation mode
c                     (the following inputs are only used for simulation
c                      but values must be supplied even for modes 0 and
c                      1)
c                     iseed    - random seed, only used in simulation
c                                mode
c                     nsim     - number of realizations required
c                     nyears   - number of years per simulated
c                                realization
c                     year1    - first year of the simulated realizations
```

```
c          record 6:  initial values of parameters to be calibrated: if a
c                     parameter is actually calibrated it should be given
c                     a non-zero initial value!
c                     par0(1) - delta1
c                     par0(2) - omega0
c                     par0(3) - omega1 (set to zero at all times)
c                     par0(4) - phi1
c                     par0(5) - cpar
c                     par0(6) – varw
c          record 7:  Indicators that are "1" if parameters are to be
c                     calibrated and "0" if they are not (icalcode(3)=0
c                     at all times).
c                     do i = 1,6 {
c                        icalcode(i)
c                     }
c                     if all codes are zero, the Kalman filter is only
c                     run once with the parameters given in record 4.
c          record 8:  gwlev0  - initial value of groundwater level
c          record 9:
c                     smeas   - variance of measurement error (m2)
c          record 10:
c                     dstep   - time step (days)
c                     simtim  - total simulation time (days)
c                               (in simulation mode: simtim also includes
c                                the startup time)
c                     ftol    - closure criterion used in calibration
c                               mode ftol typically has a value between
c                               0.000001 and 0.01. If the number of
c                               iterations exceeds 100 ftol should be
c                               increased.
c          record 11:
c                     startim - startup time (days)
c                               (only relevant in simulation mode; the
c                                time in days used for warming up the
c                                simulation ndays recorded in output =
c                                simtim-startim and should total the
c                                nyears per realization)
c
**********************  end parameter file ******************************
c
c*********************************************************************
c                  MEASFILE - file with measurements of heads
c*********************************************************************
c           record 1: nmeas - number of measurements
c                     cfac - conversion factor
c           do i = 1,nmeas {
c              record(i+1): xmtime(i) - measurement time-step (days)
c                           xmeas(i)  - measurement (m)
c                          }
c********************** end file with measurements **********************
c
c*********************************************************************
c                  METEOFILE - file with meteo data
c*********************************************************************
c           record 1: ndays - number days to read
c                     ncols - number of columns present
c                     cfac  - multiplication factor (for unit conversion)
c                             if right unit: cfac = 1
c                     icolp - column number precipiation
c                     icolep - column number pot. evapotranspiration
```

```
c           do i = 1,ndays {
c              record(i+1): xmtime(i) - measurement time-step (days)
c                           ncols columns
c                          }
c********************* end meteofile ***********************************
c Output:
c*****************************************************************
c          1) kalfile
c          a) in kalman filter or calibration mode:
c                  time      - time (days)
c                  gw0       - time update (m below surface)
c                  gw1       - measurement update (m below surface)
c                  sgw0      - error variance time update (m2/day2)
c                  sgw1      - error variance measurement update
c                              (m2/day2)
c             when measurements are available:
c                  xmeas     - measured value (m below surface)
c                  (xmeas - gw0) - innovation (m below surface)
c          b) in simulation mode:
c             unformatted file (can be used as input for STATSIM)
c             record 1: year1 – first year of simulated realisation
c                       nsim – number of simulated realisations
c                      nyears – number of years per realisation
c             do k = 1,nsim
c                do i = 1,nsteps
c                   record 1 + (k-1)*nsim + i: gw1 – simulated water
c                                                    table depth
c                enddo
c             enddo
c note: nsteps =int(simtim/dstep)
c*****************************************************************
c          2) optfile:
c                  iter            - iteration number
c                  do i = 1,nwc+4 {
c                     par(i)       - parameters
c                {
c                funk(par)       - value of criterium
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

The following is an example parameter file for KALTFN

```
32cl34ra.10
debilt.cal
tfnc15.k10
tfnc15.o10
0 -79359 1 5 1992
0.954996 0.44010 0. 0.908164 -129.935883 15.492792
1 1 0 1 1 1
-120
0.
1. 4017 0.00001
365
```

Part 3: Documentation SSD

A program for modelling the water table depth using a stochastic differential equation of the soil water balance

1 Model, input variables and parameters

1.1 Model description

An extensive description of the model, its derivation and use can be found in:
English
Bierkens, M.F.P., 1998. Modeling water table fluctuations by means of a stochastic differential equation. *Water Resources Research* 34(11), 2485-2499.
Dutch
Bierkens, M.F.P., 1988. Eenvoudige stochastische modellen voor grondwaterstands-fluctuaties. Deel 1: Een stochastische differentiaalvergelijking. *Stromingen* 4(2), 5-26.

Here follows a brief description of the model and its parameters.
The water table depth is described with the following stochastic differential equation (SDE):

$$G(h)\frac{dh}{dt} = P(t) - E_a(S(h),t) + q_v(t) - q_d(h,t) + x(t) \qquad (1)$$

The SDE is derived by setting up the water balance for a soil column (Figure 1), where it is assumed that the soil moisture is at equilibrium at all times. The following variables and parameters are distinguished:

$h(t)$ water table height with respect to some reference level [L] representing the water storage in the groundwater zone;

$S(h)$ average soil saturation [-], which is zero if the soil is completely dry (all pores are filled with air) and 1 if the soil is saturated (all pores are filled with water). The average soil saturation represents water storage in the vadose zone;

It is assumed that the soil moisture profile is at equilibrium at all times. The equilibrium profile is modelled with the following Van Genuchten-type of relationship (Troch et al., 1993) (y is suction head [L]):

$$q(z,t) \equiv q_e(y) = q_r + (q_s - q_r)\left[\frac{1}{1+(a\,y)^n}\right]^{\frac{n+1}{n}} \qquad (2)$$

where a [L^{-1}] and n [-] are Van Genuchten-type parameters, q_s [-] is the saturated moisture content and q_r [-] the residual moisture content of the soil. Note that instead of the exponent $n/(n-1)$ that normally appears in the Van Genuchten's relationship (Van Genuchten, 1980) the exponent $n/(n+1)$ is used. Using (2) the average soil saturation can be written explicitly as a function of the water table height with respect to the surface elevation z_s):

$$S(h) = \frac{1}{a\,(z_s - h)} \left\{ 1 + \left[a\,(z_s - h) \right]^{-n} \right\}^{-1/n} \tag{3}$$

The assumption of an equilibrium soil moisture profile entails that the model is only suitable for shallow water table depths. In this case the redistribution of soil moisture as a result of precipitation or changing water table depth occurs rapidly, so that an equilibrium soil moisture profile can be expected to occur almost instantaneously. Moreover, the equivalent soil moisture profile, i.e. the soil moisture profile that will allow the time-averaged vertical flux to pass through the soil column, is approximately equal to the equilibrium profile for shallow water tables (Salvucci and Entekhabi, 1994).

$G(h)$: dynamic storage coefficient [-] given by:

$$G(h) = e_0 + (q_s - q_r) \left[1 - \left[1 + \left[a\,(z_s - h) \right]^{-n} \right]^{\frac{-1}{n}} \right] \tag{4}$$

The dynamic storage coefficient can be derived from (2) (Bierkens, 1998). The parameter e_0 the residual groundwater storage (due to air bubbles and surface depressions) that remains when the water table is close to the surface: $h(t) \rightarrow z_s$. From (4) it can be seen that the storage coefficient decreases when the water table is closer to the surface. This can be expected because a shallower water table has the effect that more soil pores are filled with water.

$P(t)$ net precipitation [LT^{-1}]: precipitation minus interception by foliage;

$E_a(S,t)$ actual evapo-transpiration [LT^{-1}], which depends on the so called Makkink reference crop evapo-transpiration $E_r(t)$ [LT^{-1}] (Winter et al., 1995), a crop factor F_c (1.0 for standard grassland (Feddes, 1987) and the soil saturation as:

$$E_a(S,t) = F_c E_r(t) \left[S(t) \right]^c \tag{5}$$

where c is a constant (≈0.5 for fully vegetated soils (Lowry, 1959));

$q_v(t)$ flux from/to the deeper groundwater [LT^{-1}];
$q_d(h,t)$ drainage to/from the surface waters [LT^{-1}].

Usually several types of surface water systems (including trenches and drains) are present. The total flux $q_d(h,t)$ is thought of as the sum of separate fluxes $q_{d,i}(h,t)$ to the m_d surface water systems present:

$$q_d(h,t) = \sum_{i=1}^{m_d} q_{d,i}(h,t) \tag{6}$$

Each of these fluxes depends on the water table height $h(t)$ relative to the surface water level $H_{s,i}(t)$ [L] and a drainage resistance g_i [T] as:

$$q_{d,i}(h,t) = \frac{h(t) - H_{s,i}(t)}{g_i} \quad (7)$$

If level $H_{s,i}(t) < h(t)$ the surface water drains the groundwater, if $H_{s,i}(t) > h(t)$ it supplements the groundwater. It is also possible that a trench or a drain is present, in which case $H_{s,i}(t)$ represents the trench bottom or drain elevation. Drains or trenches drain if $H_{s,i}(t) < h(t)$ and are inactive if $H_{s,i}(t) > h(t)$.

$x(t)$ system noise: a white noise process, which models all variations in $h(t)$ that cannot be explained by the model (including both the model error and the errors due to uncertainty about parameters and input variables).

The problem with a continuous white noise process is that it has infinite variance. This means that no real solution to (1) would be possible. To overcome this problem, not the white noise itself is modelled but the white noise process multiplied by a small time step dt, which is the same as an independent the increment of the Wiener-Levy process $b(t)$ (Brownian motion) over the interval dt:

$$x(t)dt = b(t+dt) - b(t) \equiv db_t \quad (8)$$

The symbol db_t is used to denote this increment, which has the following statistical properties ($E[]$ means expected value):

$$\begin{aligned} &E[db_t] = 0 \quad \forall t \\ &E[db_t^2] = s^2 dt \\ &E[b(t+dt)b(t)] = 0 \quad \forall dt \neq 0 \end{aligned} \quad (9)$$

From the last property it can be concluded that two increments of any two non-overlapping time periods dt_1 and dt_2 are independent, no matter how small dt.

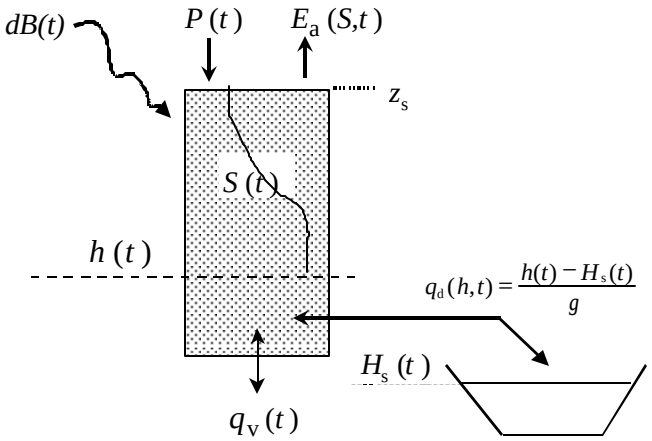


Figure 1. Schematic representation of the stochastic model for the soil water balance of a soil column

1.2 Input variables and parameters / invoervariabelen en parameters

Next follows the list of input variables and parameters used in the program SSD (units between square brackets), first in English, then in Dutch:

English:
<u>input variables</u>
$P(t)$ precipitation minus interception losses (mm/d). $P(t)$ can be calculated from the measured precipitation $R(t)$ (mm/d) using the interception fraction F_i:

$$P(t) = (1 - F_i)R(t) \tag{10}$$

Table 1 gives values of F_i for a number of different land use classes. Note that SSD does not take care of Equation (10), so that in the input file $P(t)$ and not $R(t)$ must be given;
$E_r(t)$ Makkink reference crop evapo-transpiration (mm/d);
<u>surface parameter and initial conditions</u>
z_s surface level (=reference level: usually set to 0) (cm). We advise to use the average water table depth, as is z_s is also used to give on the screen an indication of the characteristic response time and the prediction variance.
h_0 initial value water table depth (cm reference surface level e.g. -130);
<u>crop/land use parameter</u>
F_c crop factor [-] (see Equation 5). Table 1 gives the crop factor F_c for a number of different land use classes;
<u>soil parameters</u>
q_s saturated volumetric water content (-) (see Equation 2);
q_r residual volumetric water content (-) (see Equation 2);
a Van Genuchten-type parameter (cm^{-1}) (see Equation 2);
n Van Genuchten-type parameter (-) (see Equation 2).

Values of q_s, q_r , a and n for representative soils can be obtained from Table 2;

e_0 residual groundwater storage (-) (see Equation 3). Typical values range between 0.001 and 0.1;

hydrological parameters

q_v bottom flux: flux from/to deeper groundwater (mm/d) (see Equation 1). Typical values range between –1.0 and 3.0 mm/d;

m_d number of surface water systems (see Equation 6). These include also trenches and drains;

for each surface water level the following parameters i = 1,2,..:

$H_{s,i}$ drainage level (cm) with respect to surface level (e.g. –150). This can either be a surface water level or trench bottom or drainage depth. It is also possible that the drainage levels depend on time, i.e. $H_{s,i}(t)$, in which case they are input variables are read from a file (see Equation 7);

$dflag_i$ A flag which is "0" if system i is only draining (trenches and drains) and "1" if system i is both draining and infiltrating. The case that the surface water level depends on outflow from the water course and the inflow from the groundwater (as in a natural stream) is not modelled here;

g_i drainage resistance (days) of surface water system i (see Equation 7). Values may range between 5 and 500 days;

noise parameters

s^2 variance of the system noise process db_t (see Equations 8 and 9) (mm^2/d^2). Typical values range between 10 and 50 mm^2/d^2;

s_y^2 variance of measurement error (cm^2/d^2); only used in case the Kalman filter is used (see chapter 2). In most time series analysis set to zero.

In Table 2 the parameters for relationship (3) are given for the soils of the Staring Series (Wösten et al., 1994). Remember that relationship (3) is different from the original Staring Series so the parameters are also different. If the Staring Series is to be used together with SSD the correct parameters should be read from Table 2. The SDE has been derived for homogeneous soil profiles only. However, we added a number of additional codes for layered profiles. Table 3 lists the combination layers analysed and the Staring Series codes used for each of the layers. The effective parameters were obtained as follows. Suppose we have profile with 25 cm sand on 75 cm clay. First, for a large number of suction heads the soil moisture content is calculated for both sand and clay, using Equation (2). Next the total soil moisture content for each suction head is calculated by taking a weighted average of the soil moisture contents of sand (weight 0.25) and clay (weight 0.75). The result is an effective soil moisture equilibrium function for a soil of 25 cm sand on 75 cm clay or 75 cm clay on 25 cm sand. Finally the associated soil physical parameters can be found by fitting Equation (2) to the effective soil moisture equilibrium function.

Nederlands:

invoervariabelen

$P(t)$ Neerslag min interceptie (mm/d). $P(t)$ kan berekend worden uit de gemeten neerslag $R(t)$ (mm/d) via een interceptiefactor F_i:

$$P(t) = (1 - F_i)R(t) \qquad 10)$$

Tabel 1 geeft waarden van F_i voor verschillende landgebruiksklassen. SSD zelf past formule (10) niet toe. De gebruiker moet $P(t)$ dus buiten SSD berekenen en in de invoerfile meegeven en niet $R(t)$;

$E_r(t)$ Makkink referentieverdamping (mm/d);

referentie en beginvoorwaarde

z_s referentiehoogte (meestal het maaiveld, gelijk gezet aan 0) (cm);

h_0 beginwaarde grondwaterstand (cm referentie, bijv. -130). Neem hiervoor bij voorkeur de gemiddelde grondwaterstand, omdat deze parameter ook wordt gebruikt om op het scherm een indicatie te geven van de karakteristieke responstijd en de predictievariantie.

parameter landgebruik

F_c gewasfactor (-) (Vergelijking 5). Tabel 1 geeft gewasfactoren voor verschillende landgebruiksklassen;

bodemparameters

q_s verzadigd watergehalte (-) (zie Vergelijking 2);

q_r residueel watergehalte (-) (zie Vergelijking 2);

a Van Genuchten parameter (cm^{-1}) (zie Vergelijking 2);

n Van Genuchten parameter (-) (zie Vergelijking 2).

Tabel 2 geeft waarden van q_s ,q_r , a and n voor een aantal representatieve textuurklassen;

e_0 residuele grondwaterberging (-) (zie Vergelijking 2). Waarden van deze parameter variëren meestal tussen 0.001 and 0.1;

hydrological parameters

q_v kwel/infiltratie: flux van/naar dieper grondwater (mm/d) (zie Vergelijking 1). Waarden van deze parameter variëren meestal tussen –1.0 (infiltratie) en 3.0 (kwel) mm/d;

m_d aantal gemodelleerde oppervlaktewatersystemen (zie Vergelijking 6). Hieronder vallen ook drains en greppels;

Voor elk oppervlaktewatersysteem i = 1,2,..:

$H_{s,i}$ drainageniveau (cm maaiveld) (bijv. –150). Dit kan ofwel een oppervlakte-waterniveau zijn ofwel een draindiepte of greppeldiepte. Het is ook mogelijke dat de drainageniveaus afhangen van de tijd i.c. $H_{s,i}(t)$. In dat geval moeten ze van een file worden ingelezen (zie Vergelijking 7);

$dflag_i$ Als deze vlag op "0" staat geeft deze aan dat het oppervlaktewatersysteem i alleen draineert en dus kan droogvallen (greppels en drains) and als deze op "1" staat dat systeem i is zowel kan draineren als kan. Bedenk dat alleen opgegeven peilen worden gemodelleerd. Natuurlijke oppervlaktewateren waarbij het peil afhangt van het instromend grondwater en het uitstromend oppervlaktewater worden hier niet gemodelleerd;

g_i drainageweerstand (dagen) van oppervlaktewatersysteem i (zie Vergelijking 7). Waarden van deze parameter variëren meestal tussen 5 en 500 dagen;

Ruiscomponenten

s^2 variantie van de systeemruis db_t (zie Vergelijkingen 8 and 9) (mm^2/d^2). Waarden van deze parameter variëren meestal tussen 10 and 50 mm^2/d^2;

s_y^2 variantie van de meetfout (cm^2/d^2). Deze is alleen relevant als het Kalman filter wordt gebruikt (hoofdstuk 2). Bij tijdreeksanalyse wordt deze parameter meestal gelijk verondersteld aan 0.

Tabel 2 geeft de parameters voor Vergelijking (3) (de pf-curve) voor standaard textuurklassen van de Staring Reeks (Wösten et al., 1994). We herhalen hier nogmaals dat Vergelijking (3) afwijkt van de normale formulering van de Van Genuchten pf-curve, zodat de parameterwaarden ook afwijken van de waarden die vermeld staan in Wösten *et al.* (1994). Dus als SSD gebruikt wordt dan moeten de Van Genuchten parameters van de Staringreeks worden gelezen uit Tabel 2. Tabel 2 geeft ook effectieve Van Genuchten parameters voor bodems die uit verschillende verhoudingen van zand/klei, zand/veen en veen/klei bestaan. De verhoudingen van deze texturen per code staan in Tabel 3.

Table 1. Crop factors and interception fractions for various crops and land use (sources in footnotes)

Crop/land use	F_c	F_I
Grassland	0.96[1]	0.00
Potatoes	1.03[1]	0.00
Beets	0.98[1]	0.00
Grain	0.95[1]	0.00
Maize	0.92[1]	0.00
Other crops	0.88[1]	0.00
Fallow	0.70[2]	0.00
Deciduous wood	1.00[3]	0.20[3]
Coniferous wood	0.80[3]	0.40[3]
Tree nurseries	1.00[3]	0.20[3]
Other trees	0.90[3]	0.30[3]
Heath	0.70[2]	0.00
Wetland vegetation	1.04[4]	0.00
Dryland vegetation	0.70[2]	0.00
Other vegetation	0.87[5]	0.00
Orchard (soil not covered)	0.91[6]	0.06[7]
Orchard (grass strips)	0.97[6]	0.06[7]
Sports field	0.96[8]	0.00
Public garden	0.97[9]	0.06[9]
Horticulture under glass	1.30[10]	1.00
Horticulture not under glass	0.88[11]	0.00

[1] Crop factor of average growing season according to Feddes, R.A., 1997. Crop factors in relation to Makkink reference-crop evapo-transpiration. In: Verslagen en Mededelingen 39, pp. 33-45, CHO-TNO, The Hague.

[2] Jansen, P.C., 1986. De potentiële verdamping van (half-)natuurlijke vegetaties. ICW nota 1703, Wageningen (in Dutch).

[3] Moors, E.J., A.J. Dolman, W. Bouten en A.W.L. Veen, 1996. De verdamping van bossen H_2O 19(16), 462-466 (in Dutch). Furthermore, the parameters for "tree nurseries" have been taken the same as for deciduous wood, and those for "other wood" as the average of deciduous and coniferous wood.

[4] Seasonal average of Molinia from Moors, E.J., J.N.M. Stricker and G.D. van den Abeele, 1998. Evapo-transpiration of cut over bog covered by Molinea Caerulea. Agricultural University, Department of Environmental Sciences, report 73, Wageningen.

[5] Average of wetland and dryland vegetation.

[6] Assuming a tree coverage of 30% we calculate:

Orchard (soil not covered): $F_c = 0.7 \times 0.88$ (other crops) + 0.3×1.0(deciduous trees) = 0.91

Orchard (grass strips): $F_c = 0.7 \times 0.96$ (grassland) + 0.3×1.0(deciduous trees) = 0.97

[7] Assuming a tree coverage of 30% we also assume 30% of the interception of deciduous wood yielding an interception factor of $0.3 \times 0.20 = 0.06$.

[8] Taken the same as “grassland”.

[9] Taken the same as “orchard (grass strips)”

[10] All water is intercepted. The evapo-transpiration for horticulture under glass in the western part of the Netherlands is about 700 mm/year, which is about 1.3 times the average reference evapo-transpiration for that area (personal communication Philip Hamaker). The excess water needed is obtained from the intercepted precipitation water that is collected in tanks.

[11] Taken the same as “other crops”.

Table 2. Building blocks (texture classes) of the Staring Series with the soil physical parameters of Equation (2). Description of building blocks can be found in Wösten et al. (1994).

	q_s	q_r	a	N
Topsoils				
B1	0.446	0.0534	0.00203	0.803
B2	0.433	0.0611	0.00286	0.943
B3	0.465	0.0729	0.000785	0.701
B4	0.445	0.0467	0.00175	0.758
B7	0.430	0.0985	0.000359	0.540
B8	0.440	0.1330	0.000562	0.781
B9	0.437	0.1050	0.000396	0.783
B10	0.430	0.1250	0.00013	0.609
B11	0.626	0.2900	0.0000905	0.484
B12	0.548	0.2780	0.0000597	0.468
B14	0.425	0.1290	0.000345	0.773
B16	0.733	0.1210	0.000434	0.677
B17	0.749	0.3200	0.000182	0.541
B18	0.790	0.3210	0.00025	0.580
Subsoils				
O1	0.365	0.0201	0.00764	1.333
O2	0.383	0.0372	0.00696	1.316
O3	0.354	0.0416	0.00294	0.924
O4	0.364	0.0292	0.00179	0.873
O5	0.342	0.0125	0.0148	1.002
O6	0.459	0.1320	0.0000138	0.335
O8	0.470	0.0967	0.000736	0.795
O9	0.476	0.0892	0.000602	0.770
O10	0.500	0.1500	0.000433	0.751
O11	0.437	0.1580	0.0000857	0.464
O12	0.578	0.2630	0.000189	0.627
O13	0.579	0.2430	0.0000236	0.453
O14	0.381	0.0449	0.00066	1.181
O15	0.416	0.1280	0.000482	0.797
O16	0.906	0.1700	0.00076	0.816
O17	0.883	0.2230	0.000478	0.686
Two texture classes				
ZV1	0.773	0.1780	0.000583	0.647
ZV2	0.651	0.1390	0.00105	0.688
ZV3	0.521	0.0963	0.00274	0.857
ZK1	0.542	0.2140	0.000343	0.571
ZK2	0.496	0.1660	0.00125	0.683
ZK3	0.441	0.1090	0.00156	0.968
KV1	0.812	0.2290	0.000371	0.644
KV2	0.734	0.2410	0.000316	0.635
KV3	0.567	0.2520	0.000253	0.625

Table 3. Additional codes for two or more layer profiles consisting of two texture classes. Figures in table refer to the fraction of total soil depth consisting of associated texture class. For instance, 25% of the depth of a soil column with ZV1 consists of sand (building block O2) and 75% of the depth of peat (building block O17).

	Sand (O2)	Peat (O17)
ZV1	0.25	0.75
ZV2	0.50	0.50
ZV3	0.75	0.25
	Sand (O2)	Clay (O12)
ZK1	0.25	0.75
ZK2	0.50	0.50
ZK3	0.75	0.25
	Clay (O12)	Peat (O17)
KV1	0.25	0.75
KV2	0.50	0.50
KV3	0.75	0.25

2. Stochastic modelling: prediction, simulation and calibration

There are two different ways of using the stochastic differential equation: prediction and stochastic simulation. In the following sections these two different ways are further explored. Also, some of the parameters, namely g, e_0, q_v and s^2, can usually only be obtained through calibration. The section thereafter explains how calibration of the SDE is achieved. Finally, the last section recapitulates how the application of the SDE to a practical problem proceeds.

2.1 Prediction

Before we start with explaining prediction, we have to explain the nature of $h(t)$ when it is described with a stochastic model like Equation (1). It means that we are uncertain about the exact variation of h with time. We do know that it is likely to be lower in the summer time and higher during the winter, but there is still a lot of unknown variation left. This variation is due to errors in our inputs, model parameters and the fact that our model itself is only an approximation of reality. Therefore, we vision that $h(t)$ is not described by a single function of time, but as a collection of possible functions, each of which are equally probable of describing the real but unknown variation of h. Figure 2 illustrates this concept, showing four equally possible functions.

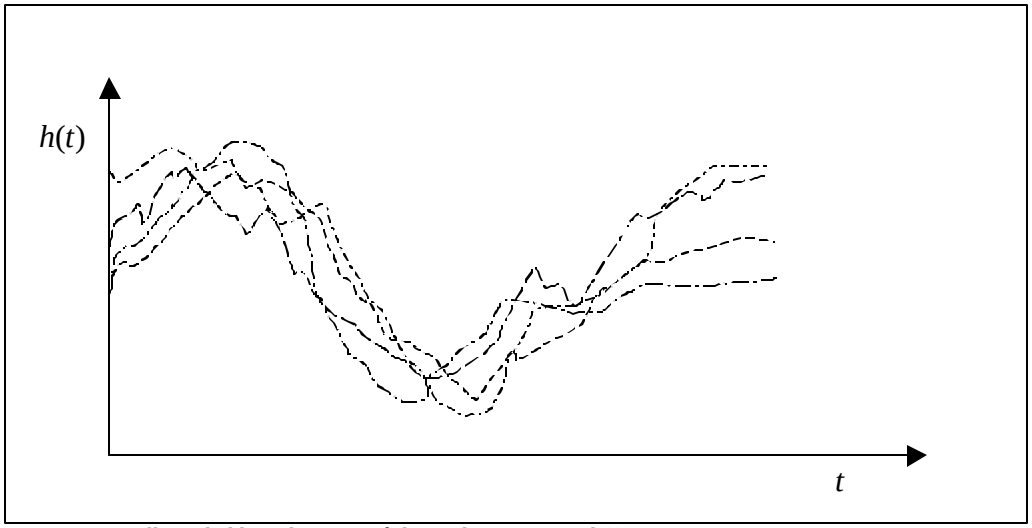


Figure 2 Equally probable realisations of the stochastic process h(t)

One particular function is called “a realisation” and the whole collection of equally probable realisations, usually an infinite number of them, is called “the ensemble” and $h(t)$ is said to be “a stochastic process”. Now reality is assumed to be on of these possible realisations, however which one exactly is unknown. We would want our prediction to be such that the prediction error is minimal. However, because we do

not know which of the realisations is reality it is not possible to evaluate the prediction error either. Instead, we are forced to look at every time step at the difference between our prediction and the values of all possible realisations. Suppose that $\hat{h}(t)$ is the prediction and $h_i(t)$ the value of realisation number *i*, then we seek a prediction for which the following properties hold:

1. The prediction error, i.e. the difference between our prediction and the realisations, is on average equal to zero:

$$\sum_{i=1}^{\infty}\left[\hat{h}(t)-h_i(t)\right]=0 \tag{11}$$

2. The average squared prediction error is minimal (as small as possible):

$$\sum_{i=1}^{\infty}\left[\hat{h}(t)-h_i(t)\right]^2 \Rightarrow \text{minimal} \tag{12}$$

It turns out that we achieve this if we take as prediction at every time the average of all realisations:

$$\hat{h}(t)=E[h(t)]=\sum_{i=1}^{\infty}h_i(t) \tag{13}$$

The average of all possible realisations is called "the expectation" or "expected value" of the stochastic process $h(t)$ and is usually denoted with the operator $E[]$. If a prediction is such that it has properties (11) and (12) it is said to be "unbiased" (Equation 11) "optimal in least squared sense" (Equation 12). If we describe the stochastic process $h(t)$ with a stochastic differential equation like equation (1) the expected value is (approximately) given by the solution of only the deterministic part of the differential equation (1):

$$G(E[h])\frac{dE[h]}{dt}=P(t)-E_a(S(E[h]),t)+q_v(t)-q_d(E[h],t) \tag{14}$$

Figure 3 shows the realisations with the optimal prediction. Also shown is the 95%-prediction interval which gives for every t the boundaries that contain 95% of the realisations. Because the noise db_t has a Gaussian (normal) distribution, the 95% prediction interval can be calculated from the standard deviation of the prediction error $s_e(t)$ as follows:

$$\left[E[h(t)]\pm 1.96 s_e\right]$$

with

$$s_{\mathrm{e}} = \sqrt{\sum_{i=1}^{\infty} \left[\hat{h}(t) - h_i(t)\right]^2} \tag{15}$$

As shown hereafter, the variance of the prediction error can be estimated using a Kalman filter. As it happens, for a non-linear model like (1) the prediction error depends on water table depth. A rough estimate of the average prediction error can be obtained with the following formula (based on Bierkens *et al.*, 1999 and Knotters and Bierkens, 2000):

$$s_e^2 = \frac{s^2}{1 - \left[\exp\left(\frac{-\Delta t}{G(\tilde{h})g_{\mathrm{eff}}}\right)\right]^2} \tag{16}$$

where $\tilde{h}$ is some nominal value of the water table depth (usually the mean value), $G(\tilde{h})$ the associate storage coefficient and g_{eff} the effective drainage resistance of the number $m_{\mathrm{d}}(\tilde{h})$ of active drainage levels for the nominal water table depth $\tilde{h}$. The effective drainage resistance is calculated as:

$$g_{\mathrm{eff}} = \left\{\frac{1}{m_{\mathrm{d}}(\tilde{h})} \sum_{i=1}^{m_{\mathrm{d}}(\tilde{h})} \frac{1}{g_i}\right\}^{-1} \tag{17}$$

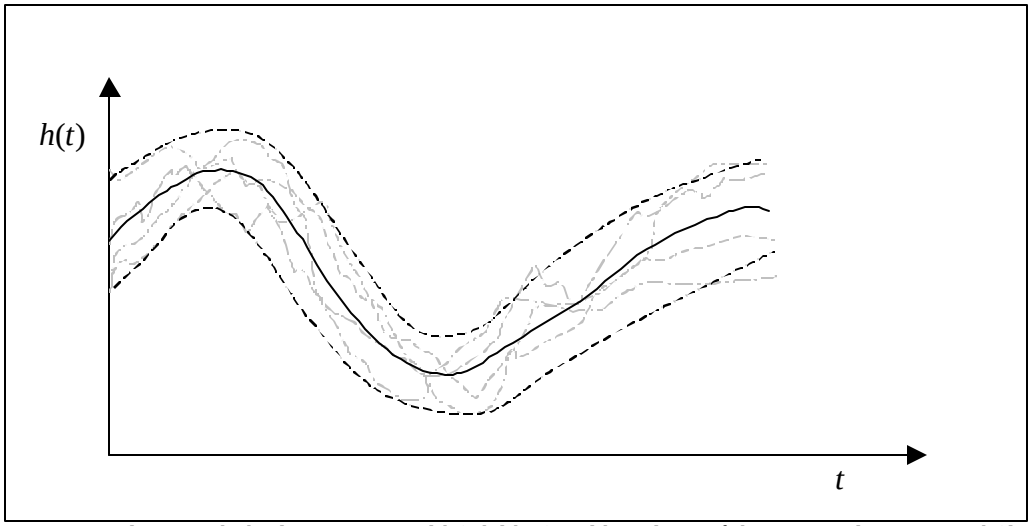


Figure 3 Prediction with the deterministic model (solid line) and boundaries of the 95% prediction interval (dashed lines); grey lines are a number of realisations.

In case we have observations of $h(t)$ we can use these to further improve the predictions. This is shown in Figure 4. Here we have a clue about the realisation that stands for reality because it has been observed a number of times. If the realisations were without error, than the best strategy would be to simply set the predicted value

equal to the observation as soon as the observation comes in (Figure 4a) and proceed from there. At that time the prediction error would be zero. If there is an observation error, the truth would be somewhere in between and an improved prediction would be some weighted average of the prediction and observation. The prediction error would than be somewhat smaller than the observation error (Figure 4b).

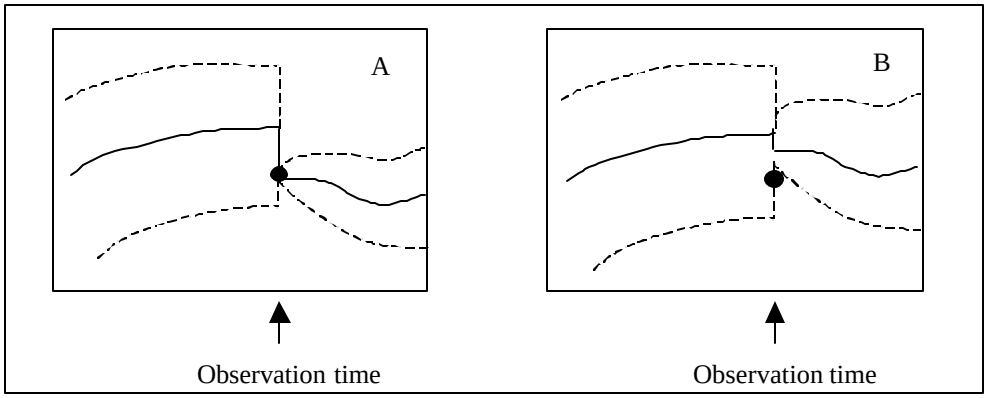


Figure 4 Correcting a prediction (solid line) with an observation without an observation error (Figure 4a) and with an observation error (Figure 4b). The dashed lines show the 95% prediction interval, which will be of zero width at the measurement time in case 4a and of limited width in case 4b.

An algorithm that performs such a correction is called the "Kalman filter". Although the Kalman filter can be used in continuous time, it is much easier to use it for discrete time steps $t = k\Delta t$, $k = 0,1,2,\ldots$ In our case, this could for instance be time steps of one day, if the input variables such as precipitation and evapo-transpiration have been observed daily. Suppose that $\hat{h}(k \mid k-1)$ is the prediction at time step k that has been obtained using (14) and corrections for all observations up to and including time step k-1 and at time step k an observation y_k is available. The corrected prediction is then obtained as follows:

$$\hat{h}(k \mid k) = \left(1 - \frac{s^2_{k|k-1}}{s^2_{k|k-1} + s^2_{y_k}}\right)\hat{h}(k \mid k-1) + \left(\frac{s^2_{k|k-1}}{s^2_{k|k-1} + s^2_{y_k}}\right) y_k \tag{18}$$

where

$\hat{h}(k \mid k-1)$ the "time update", which is the optimal prediction at time step k given all observations up to including time step k-1. The time update is such that the difference between the time update and all realisations is on average zero and that the variance of this difference is minimal:

$$\sum_{i=1}^{\infty}\left[\hat{h}(k \mid k-1) - h_i(k\Delta t)\right] = 0 \tag{19}$$

$$s^{2}_{k|k-1} = \sum_{i=1}^{\infty}\left[\hat{h}(k \mid k-1) - h_i(k\Delta t)\right]^2 \Rightarrow \text{minimal} \tag{20}$$

$s^{2}_{k|k-1}$ the variance of the error in the time update as defined in (20), which is calculated by the Kalman filter;

$\hat{h}(k \mid k)$ the "measurement update", which is the optimal prediction at time step k given all observations up to including time step k. The measurement update is such that the difference between the measurement update and all realisations is on average zero and that the variance of this difference is minimal:

$$\sum_{i=1}^{\infty}\left[\hat{h}(k \mid k) - h_i(k\Delta t)\right] = 0 \tag{21}$$

$$s^{2}_{k|k} = \sum_{i=1}^{\infty}\left[\hat{h}(k \mid k) - h_i(k\Delta t)\right]^2 \Rightarrow \text{minimal} \tag{22}$$

$s^{2}_{k|k}$ the variance of the error in the measurement update as defined in (22) which is also calculated by the Kalman filter;

$s^{2}_{y_k}$ the variance of the error in the observations.

The time update is calculated by application of the deterministic model (14) between time steps k-1 and k with measurement update $\hat{h}(k-1 \mid k-1)$ as initial condition:

$$\hat{h}(k \mid k-1) = \hat{h}(k-1 \mid k-1) + \int_{(k-1)\Delta t}^{k\Delta t} \frac{P(t) - E_a(S(E[h]),t) + q_v(t) - q_d(E[h],t)}{G(h)}dt \tag{23}$$

From Equation (18) the workings of the Kalman filter become clear. If the observation error is small compared to the model prediction error (i.e. the time update), the updating in (18) will be such that the measurement update is close to the value of the observation. On the other hand, if the error in the time update is small compared to the observation error, i.e. the model is very precise when compared to the observations, the measurement update resembles very much the time update. The weight used in the weighted average (18), i.e. $[s^{2}_{k|k-1}/(s^{2}_{k|k-1} + s^{2}_{y_k})]$ is called the "Kalman gain".

If at time step k there are no observations available, we simply have:

$$\hat{h}(k \mid k) = \hat{h}(k \mid k-1) \tag{24}$$

$$s^{2}_{k|k} = s^{2}_{k|k-1} \tag{25}$$

If for a long time no observations are taken we have that $s^2_{k|k} = s^2_{k|k-1} \rightarrow s^2_e$, i.e. the Kalman filter prediction variances become equal to the prediction variance of the deterministic models, as given in Equation (15), and approximated with Equation (16).

Figure 5 is similar to Figure 3, but now we have a number of observations present and predictions are performed with the Kalman filter. For convenience, we have shown the case with no observation errors. At a number of time steps observations are present. Again, our unknown reality is modelled as a large number of equally probable realisations. However, because we have observations, only those realisations are eligible that pass through the observations (in case of a small observation error, the realisations do not have to pass through the observations exactly, but should pass close by). The time update is shown, as well as the 95% prediction interval. Because the predicted water table depth is updated, the variance of the prediction error is smaller than without updating (Figure 3); to put it in another way: the averaged squared difference between the prediction (time update) and the realisations is smaller than without the use of the Kalman filter.

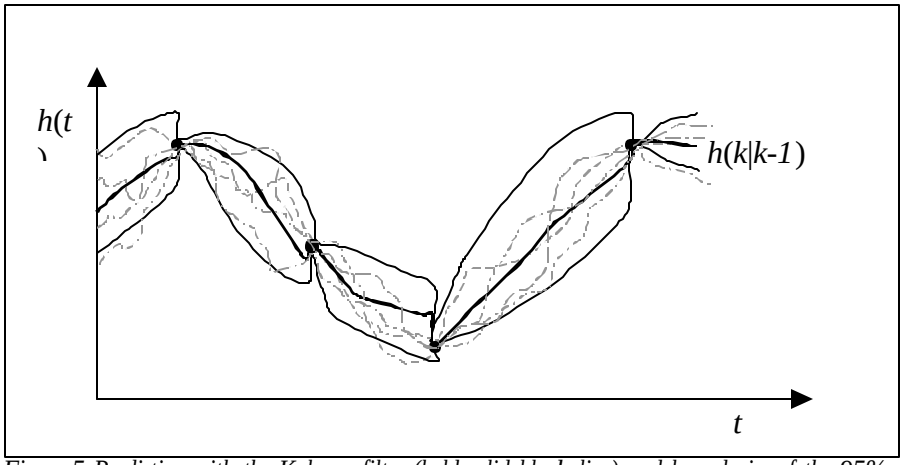


Figure 5 Prediction with the Kalman filter (bold solid black line) and boundaries of the 95% prediction interval (thin solid black lines); grey lines are a number of realisations, all passing through the observations; black dots are observations.

2.2 Stochastic simulation

Prediction of water table depth is important if we are interested in the actual, but non-observed water table depth. However, in many applications we are not particularly interested in the actual water table depth, but in some fluctuation quantities. For instance, we may be interested in the probability that at any day in the near future the level of 30 cm is exceeded, or we want to know the mean highest water table depth (MHW) or mean lowest water table depth (MLW) (Van der Sluijs and De Gruijter, 1985). These fluctuation quantities cannot be estimated from the predicted water table depth, because the predicted line typically overestimates the low values and underestimates the high values (see Figures 3 and 5). The reason for this

is that it tries to minimise the squared prediction error (Equation 15). In this case we should do the following (See Figure 6):

1. simulate a large number (at least 100) realisations of the stochastic process $h(t)$. A realisation is simulated by numerical evaluation of the following integrals (based on Equation (1) with Equation (8) for the noise component):

$$h(t) = h(t_0) + \int_{t_0}^{t} \frac{P(t) - E_a(S(E[h]),t) + q_v(t) - q_d(E[h],t)}{G(h)}\,dt + \int_{t_0}^{t} \frac{s}{G(h)}\,db(t) \quad (26)$$

   where the increments of the Gaussian process $b(t)$ are simulated for an integration step of size dt by drawing from a Gaussian distribution with mean zero and variance dt. The numerical evaluation of the second integral is not particularly straightforward, because of the special properties of the process $b(t)$. Special Runge-Kutta integration can be used as explained by Newton (1991);
2. for each simulated realisation the appropriate fluctuation quantities are estimated. For 100 realisations this yields also 100 values of this property, e.g. 100 MHWs and 100 MLWs;
3. the cumulative frequency estimated from the replicas of the fluctuation quantity (e.g. the 100 MHWs and MLWs) represents an estimate of the probability distribution of the fluctuation quantity. This probability distribution expresses the uncertainty about the true value of the fluctuation quantity, uncertainty that arises from our model's inability to predict the unknown water table depth exactly. Usually the average of the replicas is used as an estimate of the true but unknown fluctuation quantity, and as a measure of uncertainty the 95% confidence interval is calculated.

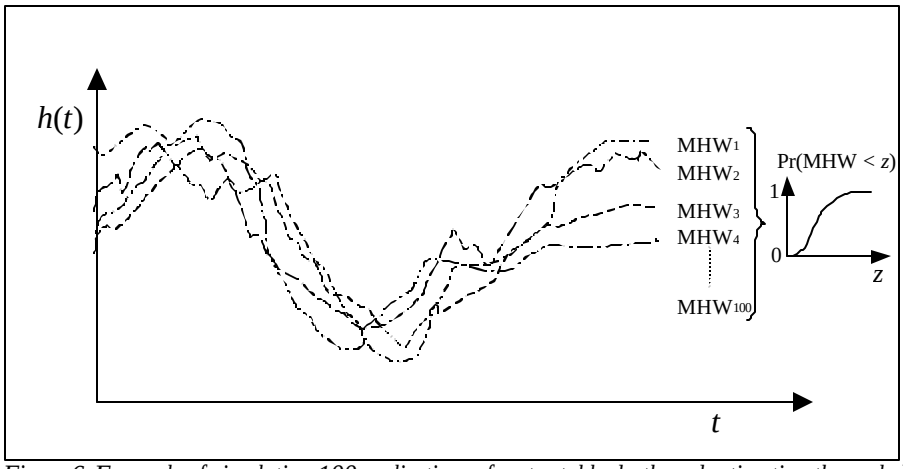


Figure 6 Example of simulating 100 realisations of water table depth and estimating the probability distribution of MHW from it.

2.3 Calibration

To apply the SDE at some location the following parameters must be given a value: z_s, q_s, q_r, a, n, F_i, F_c, H_s, g, $e_{0,}$ q_v and s^2. Elevation z_s is usually known from surveying. The parameters q_s, q_r, a, n can either be obtained from laboratory analysis on sample cores, or they can be taken from a reference soil moisture retention curve belonging to the texture class of the soil (e.g. Wösten and Van Genuchten, 1988) (Table 2). The drainage water levels H_s and the dominant land use near the location are determined from field observations. From the dominant land use follow the interception factor F_i and the crop factor F_c (Table 1). The resulting parameters g, $e_{0,}$ q_v and s^2 must be obtained by calibrating the SDE to the time series of water table depth. In case of a deterministic model, calibration only involves the minimisation of some least squares criterion between predictions and observations. However, in this case we also have to calibrate the noise parameter s^2. For linear stochastic models this parameter could be estimated separately from the residuals (differences between deterministic predictions and observations). However, even for linear models, such a two step approach generally leads to biased parameter estimates (see Te Stroet, 1995). Also, if the time series is very irregular (not a constant frequency), estimating s^2 is quite problematic. Here, a method is used combining the Kalman filter and a maximum likelihood criterion that, given the assumptions (model prediction errors and measurement errors are Gaussian distributed), provides unbiased maximum likelihood estimates of the parameters g, $e_{0,}$ q_v and s^2. The method can be used for irregularly and sparsely observed time series and has the added advantage that it is able to take account of measurement errors.

If we have M time steps with observations (not necessarily with regular intervals between them), the method proceeds as follows (we assume that the input time series, i.e. precipitation, potential evapo-transpiration and surface water levels have been collected for the period for which we have observations):

1. choose initial values for parameters g, $e_{0,}$ q_v and s^2;
2. use these parameter values to run the Kalman filter for the period that contains the observations. This yields for each observation occasion an "innovation" $n_l = y_l - \hat{h}(l \mid l-1)$ and the associated innovation variance (calculated by the Kalman filter, see Bierkens, 1998) $s_{n_l}^2 = s_{l|l-1}^2 + s_{y_l}^2$;
3. from the M innovations and innovation variances the following maximum likelihood criterion is evaluated (Schweppe, 1973):

$$J = M \ln(2p) + \sum_{l=1}^{M} \ln\left[s_{n_l}^2\right] + \sum_{l=1}^{M} \left[\frac{n_l^2}{s_{n_l}^2}\right] \tag{27}$$

4. choose a new set of parameters g, $e_{0,}$ q_v and s^2;
5. repeat steps 2 to 4 until ML-criterion (27) is minimised. The resulting parameter set is a maximum likelihood estimate of the parameters.

Minimisation of (27), i.e. choosing the new set of parameters in step 3 such that the value of (27) will decrease, is done with a minimisation algorithm. In SSD the "downhill simplex method" is used (Press et al., 1986).

2.4 Systematic application of SSD: step by step

In the following example we assume that we have a location for which the values of z_s, q_s, q_r, a, n, F_i, F_c have been determined. Surface water levels, precipitation and reference evapo-transpiration are known for the years 1970-1999 and we have observations of water table depth available with an observation frequency of two times a month for the years 1995-1999. Application of the SSD model has two goals. First, we want to estimate fluctuation properties (MHW, MLW etc.) at the location. Second, we aim to monitor the water table depth on a daily basis in the future, while maintaining an observation frequency of two times a month. The following steps are taken:

2.4.1 Calibration

The parameters g, e_0, q_v and s^2 are estimated through calibrating the SDE to the observations from 1995-1999 (section 2.3). For this purpose the program SSDKOPT is used. Usually it is assumed that the observations are without error, as in time series analysis. However, an observation error variance of 1-4 cm^2 is appropriate for most piezometer data. Although in theory the parameters of the deterministic model part g, e_0, q_v and the parameter of the stochastic part s^2 should be calibrated simultaneously, it is recommended to use a three-step procedure in SSDKOPT (in VIDENTE these three modes can be chosen from in the Calibration menu):

1. first, the parameters g, e_0, q_v are calibrated (called "deterministic calibration" in VIDENTE). This achieved by fixing the parameter $s^2 = 0$ during the calibration and setting the observation error variance $s_{y_k}^2 = 1$ for all time steps.
2. in the second step, called the "stochastic calibration" the parameters g, e_0, q_v are fixed at the values obtained in step 1 and the parameter s^2 is calibrated while setting $s_{y_k}^2 = 0$ for all time steps.
3. finally, using the parameter values found in steps 1 and 2 as initial estimates, all parameters g, e_0, q_v and s^2 are calibrated in the last step, thus making sure that unbiased estimates of these parameters are found ($s_{y_k}^2$ set at correct value). If the adjustment in step 3 is large, i.e. if very different values of the parameters are found, we must be suspicious and perhaps decide to use the parameters found in steps 1 and 2 and forget step 3.

The result of the calibration with SSDKOPT is a file with calibrated parameters and the mean error (*ME*), root mean squared error (*RMSE*) and mean absolute error (*MAE*) of Kalman filter time updates:

$$ME = \frac{1}{M}\sum_{k=1}^{M}\left(\hat{h}(k \mid k-1) - y_k\right) \tag{28}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{k=1}^{M}\left(\hat{h}(k \mid k-1) - y_k\right)^2} \tag{29}$$

$$MAE = \frac{1}{M}\sum_{k=1}^{M}\left|\hat{h}(k \mid k-1) - y_k\right| \tag{30}$$

Furthermore, the file with Kalman filter predictions is given, which contains the Kalman filter time updates, measurement updates, their variances, observations, innovations (see 2.3) and the innovations 95% prediction interval. Also given is a file with the other water balance terms, such as soil saturation, specific discharge and actual evapo-transpiration. Figure 7 shows an example of the output from a calibration run.

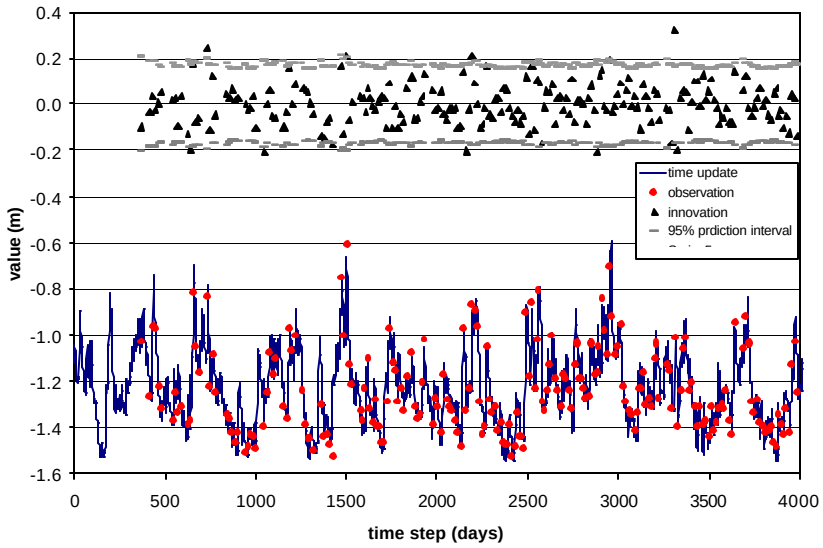


Figure 7 Example output from calibrating the SDE

2.4.2 Verification: prediction without Kalman filter

After calibration, it is wise to check whether the deterministic part of the model is able to describe the dynamics of the water table depth sufficiently well. An indication that a deterministic fit is problematic would have been the a strong adjustment of the parameter values in step 3 of the Calibration step. Figure 7 shows the time updates from the Kalman filter. A problem with using the Kalman filter could occur if the number of observations is large. In that case, the resulting time updates could be insensitive to the model dynamics, because updating is performed too frequently. In that case a small value of criterion (25) could still result in poor estimates of the

dynamic parameters g, e_0 and q_v. Therefore, a prediction with the deterministic part of the model should be made, without the updating. This can be achieved by running SSDKOPT for the calibration period as a means of verification by setting $s^2 = 0$ and the observation error variance $s^2_{y_k} = 1$, and fixing the values of g, e_0 and q_v at their calibrated values (in VIDENTE this is an option in the prediction menu). The result of the verification with SSDKOPT is a file with parameters used and the following verification statistics:

$$ME = \frac{1}{M}\sum_{k=1}^{M}\left(\hat{h}(k\Delta t) - y_k\right) \tag{31}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{k=1}^{M}\left(\hat{h}(k\Delta t) - y_k\right)^2} \tag{32}$$

$$MAE = \frac{1}{M}\sum_{k=1}^{M}\left|\hat{h}(k\Delta t) - y_k\right| \tag{33}$$

which are usually larger than the statistics from the calibration run, because no updating is used. SSDKOPT also gives the file with predictions, the observations and the residuals $\hat{h}(k\Delta t) - y_k$. Figure 8 shows an example of the output from a verification run, where in this case observations are present almost every day.

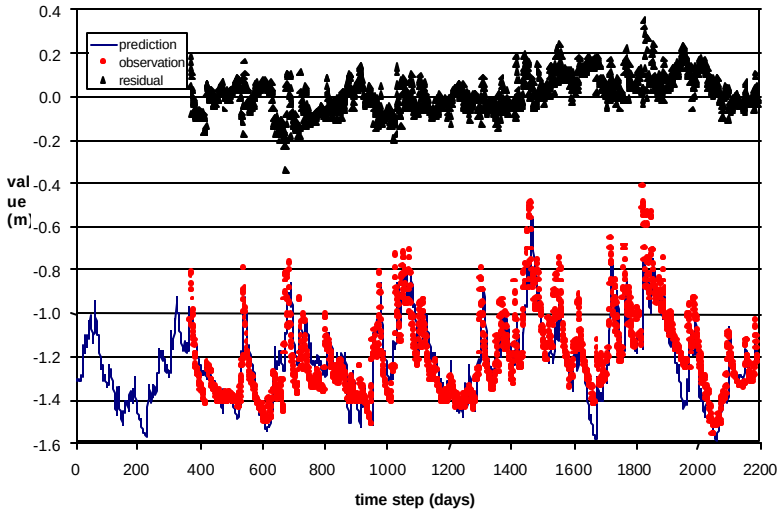


Figure 8 Example of verification of the deterministic part of the SDE model by prediction without updating

2.4.3 Simulation

Using the calibrated parameters from the calibration step the program SSDPATH can be used to simulate single realisations of $h(t)$ and all related properties such as soil

saturation, drainage discharge and evapo-transpiration. In case fluctuation quantities have to be estimated, multiple realisations are required which can be simulated using the program SSDSIM. The output of SSDSIM consists of three files, containing respectively the realisations of $h(t)$, $S(t)$ and $q_d(t)$. These files can be used as input for the program STATSIM, in order to calculate the fluctuation properties. In our example we would typically simulate realisations of 30 years long, i.e. using precipitation and evapo-transpiration (and possibly surface water levels) from 1970-1999, in order to obtain fluctuation quantities that are representative for the current climate, and not only for the weather conditions for the observation period 1995-1999 (see Knotters and Van Walsum, 1997).

2.4.4 Estimating fluctuation quantities

Using the program STATSIM fluctuation quantities can be calculated from the simulated realisations. The output consists of the following (here we describe the output for $h(t)$; similar output can be obtained for $S(t)$ and $q_d(t)$:

A file with statistics

An example of this output is shown hereafter in Figure 9. The mean highest water table, the mean lowest water table and the mean spring water table are given (expected values, median values and 5 and 95 percentiles). Here quantities are calculated per realisation. So, the uncertainty here reflects only model uncertainty. The univariate statistics are statistics over all simulated water table depths, which means that they reflect both the within year variation and the year to year variation of the weather as well as the model uncertainty (variation between realisations).

```
---------------------------------------
      Univariate Statistics
---------------------------------------
        mean   -106.425819
    variance    360.064484
       stdev.    18.975365
 3rd moment   -539.639771
---------------------------------------
         P01   -148.944107
         P05   -138.432495
         P10   -132.007401
         P25   -119.718979
         P50   -105.809341
         P75    -92.923340
         P90    -82.159462
         P95    -75.837044
         P99    -65.247566
---------------------------------------


-------------------------------------------------------------------------------
 Mean Highest and Mean Lowest Water Table and Mean Spring Water Table
-------------------------------------------------------------------------------
                 mean          5%          50%          95%         std
-------------------------------------------------------------------------------
      GLG(t) -135.358978 -138.189285 -135.264694 -132.976761    1.565006
      GHG(t)  -78.744606  -81.293327  -78.789520  -76.163002    1.535088
      GLG(z) -135.058350 -138.189285 -134.869598 -132.487930    1.608089
      GHG(w)  -79.471909  -82.413284  -79.414589  -76.819191    1.603910
      GVG(t)  -99.525368 -103.282478  -99.257187  -96.835274    2.082639
-------------------------------------------------------------------------------
```

Figure 9 File with fluctuation statistics as output of STATSIM; GHG: mean highest water table; GLG: mean lowest water table; GVG: mean spring water table; between brackets: denoting whether the statistic is determined for the whole year (t), the summer period (s) or the winter period (w).

A file with the frequency of exceeding graph (FOE-graph)

The FOE-graph gives for every level the number of days that the water table depth on any future day of any future year (at the same climate and hydrological regime as in the calibration period) exceeds that level. It is in fact the cumulative frequency distribution, reflecting both the within year variation and the year to year variation of the weather as well as the model uncertainty. Figure 10 shows an FOE-graph.

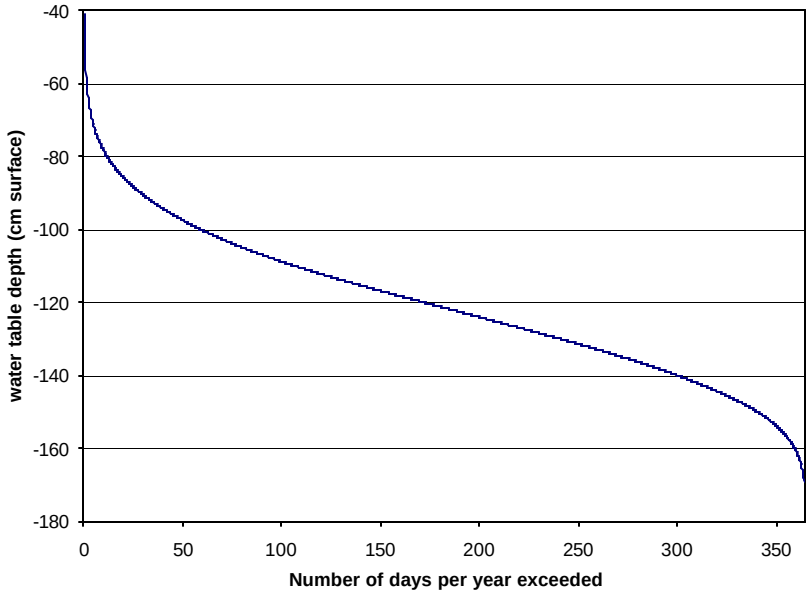


Figure 10 Frequency of exceeding graph

A file with the regime graph

The regime graph gives for every day number in a future year the expected water table depth, the median and the 5 and 95-percentiles. The variation per day number therefore reflects both year to year variation (i.e. our uncertainty about the future weather) as well as model uncertainty. Figure 11 gives an example of a regime graph.

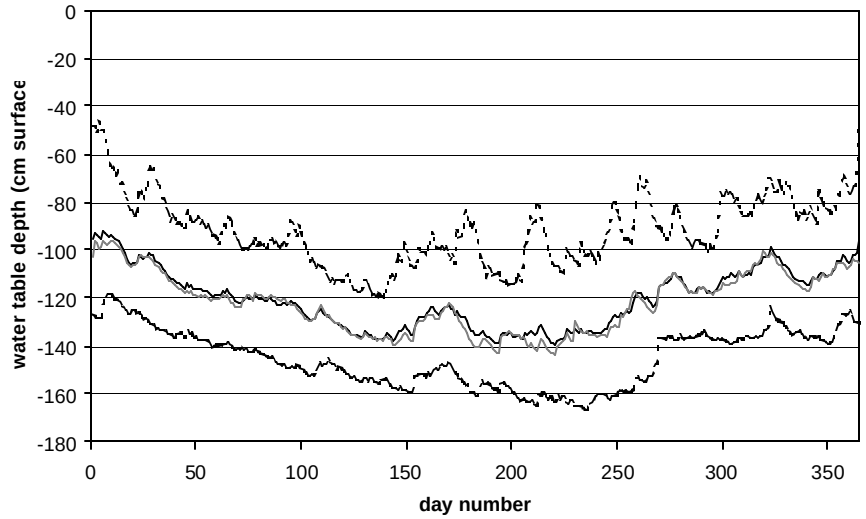


Figure 11 Regime graph; black solid line: mean; black dashed lines: 5- and 95-percentiles; grey solid line: median.

A file with the histogram

Like the FOE-graph, the histogram reflects both the within year variation and the year to year variation of the weather as well as the model uncertainty. From the histogram we can read the expected number of days that the water table depth will be within certain boundaries. Figure 12 gives an example histogram.

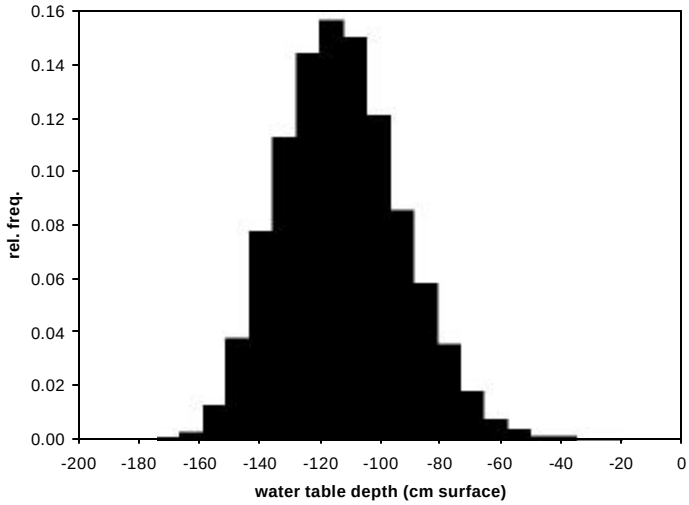


Figure 12 Example histogram

A file with the correlation function

The correlation function gives the correlation coefficient between day *k* and day *k* +1 (= lag 1), day *k* and day k+ 2 (lag 2), day *k* and day *k* + 3 etc. The correlation function reflects both the response time of the groundwater system, as well as the periodicity of the rainfall surplus. The longer it takes for the correlation function to cross the *x*-axis, the slower the response time of the groundwater system. However, because the periodicity of the rainfall surplus is also included, we cannot read the characteristic response time from the correlation function. The correlation function is the average of the correlation functions that are estimated for the realisations. An example correlation function is given below (Figure 13).

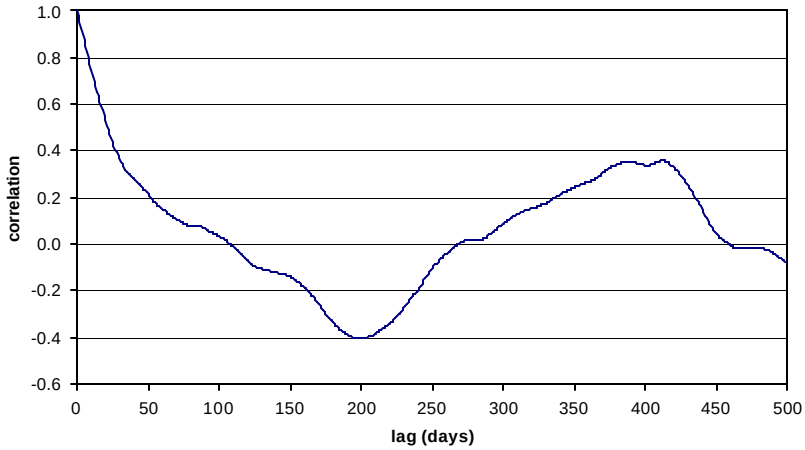


Figure 13 Example correlation function

Strictly speaking, the characteristic response time of a non-linear system like Equation (1) actually does not exist, because it depends on the water table depth itself. However, an average value of the characteristic response time (in days) can be approximated as (in analogy with Knotters and Bierkens, 2000):

$$t_c = 3G(\tilde{h})g_{\text{eff}} \tag{34}$$

where $\tilde{h}$ is some nominal value of the water table depth (as in Equation (16)). The response time gives the maximum time interval that is allowed between observations.

Of course many more fluctuation quantities could be estimated from the simulated realisations; see for instance Bierkens (1998) and Knotters *et al*. (2000).

2.4.5 On line prediction and monitoring

Finally, using the calibrated model, the program SSDKOPT can be used for on line prediction, where in between dates that observations of water table depth are taken, optimal predictions of water table depth are obtained using the Kalman filter. Also, non-observed properties can be predicted with the model, such as soil saturation, drainage discharge and actual evapo-transpiration. The Kalman filter can also be used as a monitoring instrument by running it on line and checking whether not much more than 5% of the innovations fall outside the 95% error bounds (see Figure 7). If more than 5% falls outside, a change in the hydrological system may have occurred. An alternative way of monitoring such a change is running the deterministic predictions on line and plot the observations together with the prediction interval $\hat{h}_k \pm 1.96 s_e$ (with s_e approximated with Equation (16) or calculated exactly by running SSDKOPT without the observations). If the observations start to plot outside this interval, a significant change in the hydrological system may have occurred.

References

Bierkens, M.F.P., 1998. Modeling water table fluctuations by means of a stochastic differential equation. *Water Resources Research* 34(10), 2485-2499.

Bierkens, M.F.P., 1988. Eenvoudige stochastische modellen voor grondwaterstandsfluctuaties. Deel 1: Een stochastische differentiaalvergelijking (Simple stochastic models of water table fluctuations. Part 1: a stochastic differential equation). *Stromingen* 4(2), 5-26.

Bierkens, M.F.P., M. Knotters and F.C. Van Geer, 1999. Calibration of transfer-function noise models to sparsely or irregularly observed time series. *Water Resources Research* 35(6), 1741-1750.

Feddes, R.A., 1987. Crop factors in relation to Makkink reference-crop evapotranspiration. In: *Comissie Hydrologisch Onderzoek TNO, Verslagen en Mededelingen* 39. 's-Gravenhage, 33-45.

Knotters, M. and P.E.V. Van Walsum, 1997. Estimating fluctuation quantities from time series of water-table depth using models with a stochastic component. *Journal of Hydrology* 197, 25-46.

Knotters, M. and M.F.P. Bierkens, 2000. Physical basis of time series models for water table depths. *Water Resources Research* 36(1), 181-188.

Knotters, M., M.F.P. Bierkens and C. Beets, 2000. *Grondwaterdynamiek van vegetatiestandplaatsen; analyse van zesentwintig reeksen (Groundwater dynamics of vegetation stands; analysis of 26 time series).* Rapport 095, Alterra, Wageningen.

Lowry, W.P., 1959. The falling rate phase of evaporative soil moisture loss: A critical evaluation. *Bulletin of the American Meteorological Society* 40, 605.

Moors, E.J.. A.J. Dolman, W.Bouten and A.W.J. Veen et al., 1996. Verdamping van Bossen (Evaporation of Forests). H_2O 29(16), 462-466.

Newton, N.J., 1991. Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations. *SIAM Journal of Applied Mathematics* 51(2), 542-567.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1986. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, New York.

Salvucci, G.D. and D. Entekhabi, 1994. Equivalent steady soil moisture profile and the time compression approximation in water balance modeling. *Water Resources Research* 30(10), 2737-2749.

Schweppe, F.C., 1973. *Uncertain Dynamic Systems*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Te Stroet, C.B.M., 1995. *Calibration of stochastic groundwater flow models: Estimation of system noise statistics and model parameters.* Ph.D. thesis, Delft University of Technology.

Troch, P.A., F.P. De Troch and W. Brutsaert, 1993. Effective water table depth to describe initial conditions prior to storm rainfall in humid regions. *Water Resources Research* 29(2), 427-434.

Van der Sluijs, P. and J.J. De Gruijter, 1985. Water table classes: A method to describe seasonal fluctuation and duration of water tables on Dutch soil maps. *Agricultural Water Management* 10, 109-235.

Van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal* 44, 892-898.

Winter, T.C., D.O. Rosenberry, and A.M. Sturrock, 1995. Evaluation of 11 equations for determining evaporation for a small lake in the north central United States. *Water Resources Research* 31(4), 983-993.

Wösten, J.H.M., Van Genuchten, M.Th, 1988. Using texture and other soil properties to predict the unsaturated hydraulic conductivity. *Soil Science Society of America Journal* 52(6), 1762-1770.

Annex: Input instructions for SSD

Note: For all stand alone SSD programs (SSDKOPT, SSDPATH, SSDSIM) the input of levels (initial water table depth, surface level, surface water levels) is in meters (m) and the variance of the observation error in meters squared (m^2). This is different from the input screens of VIDENTE where all these inputs for SSD (in accordance to the other programs) are in cm and cm^2.

The program SSDKOPT is used for calibration (deterministic, stochastic, all parameters) and prediction (with and without the use of a Kalman filter)

Header of the program SSDKOPT with input instructions

```
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
C Copyright (C) 2000, Alterra, Green World Research
C
C This program is distributed in the hope that it will be useful,
C but WITHOUT ANY WARRANTY.  No author or distributor accepts
C responsibility to anyone for the consequences of using it or for
C whether it serves any particular purpose or work at all, unless he
C says so in writing. Everyone is granted permission to copy, modify
C and redistribute this program, but only under the condition that the
C the reference below is given and that this notice and the above
C copyright notice remain intact.
C
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c Program: SSDKOPT -  Soil Stochastic Differential equation Kalman filter
c                     and OPTimization
c
c Goal:    Calibration of a stochastic model for the simulation of
c          groundwater heads and drainage discharge, using a Kalman
c          filter form of the Gaussian maximum likelihood criterium.
c          The model is postulated in the form of a Stratonovitch
c          stochastic differential equation which is numerically solved
c          with Runge kutta integration.
c
c Version: 1.2 (december 8th 1997)
c          * Equilibrium profile modelled with a modified van Genuchten
c            model
c          * Maximum three drainage levels
c          * bouncing boundaries in optimization
c          * closure criterium optimization (ftol) read from file
c
c Author:  Marc F.P. Bierkens
c Reference: Bierkens, M.F.P., 1998. Modelling water table
c             fluctuations by means of a stochastic differential
c             equation. Water Resources Research 34(10), 2485-2499.
c
c**********************************************************************
c Input:
c interactive:
c          parfile1 - name of parameter file for the stochastic model
c          parfile2 - name of parameter file for the calibration part
c
```

```
c***********************************************************************
c                  PARFILE1
c***********************************************************************
c The values should be put in the parameter file in the following order:
c
c          general input:
c          record 1) : outfile  - name output file of perdiction routine
c          record 2) :
c                      irandom  - integer flag which is set to 0 if the
c                                 mean trajectory is required. All other
c                                 values lead to the simulation of random
c                                 paths.
c                      iseed    - seed for the random generator; value
c                                 must be negative; a value is always
c                                 needed
c
c          meteo parameters:
c          record 3) : ipflag  - if zero, precipitation is constant,
c                                if not zero it is the unit number under
c                               which the precipitation is read.
c          record 4) : if ipflag .eq. 0 then
c                         xmp   - mean precipitation (mm/day)
c                         vp    - variance precipitation (mm2/day2)
c                      else
c                         vp       - variance of noise on precipitation
c                                    (mm2/day2)
c                                    must be at first 10 positions
c                         icolnr   - column nummer that must be read from
c                                    file must be at positions 11-20
c                         precname - name of precipitation input file
c                      endif
c          record 5) : iepflag  - if zero, potential evapotranspiration
c                                 is constant, if not zero it is the unit
c                                 number under which the potential
c                                 evapotransipration is read.
c          record 6) : if iepflag .eq. 0 then
c                         xmep  - mean potential evapotranspiration
c                                 (mm/day)
c                         vep   - variance potential evapotranspiration
c                                 (mm2/day2)
c                         exps  - exponent of the relation average soil
c                                 saturation and actual
c                                 evapotranspiration (-)
c                         cropf - crop factor (1.0 for grass)
c                      else
c                         input file
c                         vep  -  variance of noise potential evapo-
c                                 transpiration (m2/day2)
c                                 must be at positions 1-10
c                         exps -  exponent of the relation average soil
c                                 saturation and actual
c                                 evapotranspiration (-) must be at
c                                 positions 11-20
c                         cropf - crop factor (1.0 for grass)
c                                 must be at positions 21-30
c                         icolnr - column number to read from file
c                                  must be at positions 31-40
c                         evapname - name of potential evapotranspiration
c                                    input file
c                      endif
c
```

```
c          soil physical parameters (modified van Genuchten model)
c          record 7):  thesat  - saturated volumetric water content (-)
c                      theres  - residual volumetric water content (-)
c                      alpha   - alpha value of model (positive) (1/cm)
c                      expn    - exponent of the modified van Genuchten
c                                relation (-)
c                      eps0    - elastic (residual) groundwater storage
c                                (-)
c
c          geohydrological parameters and inputs
c          record 8):
c                      xsurf   - surface elevation level (m reference
c                                level)
c          record 9):
c                      gwstart - groundwater level at start simulation
c                                (m reference level)
c          record 10)  iwflag  - if zero, water levels are constant,
c                                if not zero it is the unit number under
c                                which the water levels of nearby water
c                                courses are read.
c                      if iwflag .eq. 0 then
c          record 11):    nwc       - number of different drainage levels
c                                     (maximum of 3)
c          record 12):
c                          do i = 1,nwc {
c                             wl(i)    - water level of nearby water
c                                        course i (m reference) level
c                             gamma(i) - drainage resistance of ith water
c                                        course (days)
c                          }
c                      else
c          record 11):    nwc       - number of different drainage levels
c                                     must be placed in first ten
c                                     positions
c                         wlname    - name of file with water levels
c          record 12):
c                         do i = 1,nwc {
c                            gamma(i) - drainage resistance of level i
c                                       (days)
c                         }
c                      endif
c          record 13)  idrcode (i=1,nwc)
c                               - if one, the water course is both
c                                 draining and infiltrating; if zero the
c                                 water course is only draining, i.e. when
c                                 the groundwater table is above the
c                                 drainage level of this water course.
c          record 14)  ivflag  - if zero, vertical fluxes are constant,
c                                if not zero it is the unit number under
c                                which the vertical fluxes are read.
c          record 15)  if ivflag .eq. 0 then
c                         vflux  - flux to/from deeper groundwater
c                                  (mm/day);
c                                  positive value for exfiltration
c                                  (upwards)
c                      else
c                         vfname - name of file with vertical fluxes
c                      endif
c          parameters for numerical integration:
```

```
c        record 16):
c                      dstep   - time step (days)
c                      simtim  - total integration time (days)
c                      iwrite  - frequency (time steps) of writing the
c                                results
c                      intflag - if equal to 1 a first order Runge-Kutta
c                                integration will be performed, otherwise
c                                a fourth order Runge Kutta will be used.
c                      iint    - number of integration steps per time
c                                step not exceeding MAXINT
c
c******************** end parameter file 1 *****************************
c
c***********************************************************************
c                    PARFILE2
c***********************************************************************
c          record 1:  measfile - name of file with measurements
c          record 2:  kalfile  - name of file with Kalman filter results
c          record 3:  optfile  - name of file with results of
c                                optimization
c          record 4:  initial values of parameters to be calibrated: if a
c                     parameter is actually calibrated it should be given
c                     a non-zero initial value!
c                     do i = 1,nwc {
c                         par0(i) - initial drainage resistance
c                                   [gamma(i)] (days)
c                     }
c                     par0(nwc+1) - initial value of eps0 [eps0] [-]
c                     par0(nwc+2) - initial vertical flux [vflux]
c                                   (mm/day)
c                     par0(nwc+3) - initial additive noise variance [vp]
c                                   (mm2/day2)
c                     par0(nwc+4) - initial multiplicative noise variance
c                                   [vep] (mm2/day2); If vp is calibrated
c                                   vep must be set to "0." and icalcode
c                                   (nwc+4) set to "0".
c          record 5:  integer code to determine whether to calibrate
c                     a parameter: (1) or not: (0)
c                     do i = 1,nwc+4 {
c                        icalcode(i)
c                     }
c          record 6:  gw1    - initial value of groundwater level for
c                              filter algorithm (m reference level)
c                     sgw1   - initial value of measurement update
c                              variance (m2)
c          record 7:
c                     smeas  - variance of measurement error (m2)
c          record 8:
c                     ftol   - closure criterium for optimization,
c                              typically having a value between 0.0001
c                              and 0.001. If no minimum value is reached
c                              within 100 iteration steps, the closure
c                              criterium should be increased.
c
********************** end parameter file 2 *****************************
c
c***********************************************************************
c                  MEASFILE - file with measurements of heads
c***********************************************************************
c           record 1: nmeas - number of measurements
c                     convf - conversion factor to convert the numbers
c                             in the file to the numbers used in the
c                             program.
```

```
c           do i = 1,nmeas {
c              record(i+1): xmtime(i) - measurement time-step (days)
c                           xmeas(i)  - measurement (m reference level)
c                           }
c******************** end file with measurements **********************
c
c***********************************************************************
c    FILES WITH METEODATA - precipitation or potential evapotranspiration
c***********************************************************************
c           record 1: ndata - number data to read
c                     ncols - number of columns present
c                     cfac  - multiplication factor (for unit conversion)
c                             if right unit: cfac = 1
c           do i = 1,ndata {
c              record(i+1): time(i) - time-step (days)
c                           ncols columns (one of which is precipition or
c                                          potential evapotranspiration)
c                           }
c
c note : input will be assumed constant for time(i-1) < timestep <=
c        time(i)
c******************** end meteofile ***********************************
c
c***********************************************************************
c                          FILE WITH WATER LEVELS
c***********************************************************************
c           record 1: ndata - number data to read
c           do i = 1,ndata {
c              record(i+1): time(i) - time-step (days)
c                           nwc columns with water levels (m with respect
c                               to reference level; i.e. one column for
c                               every level of water course)
c                           }
c
c note : input will be assumed constant for time(i-1) < time step <=
c        time(i)
c********************* end file with water levels **********************
c
c***********************************************************************
c                          FILE WITH BOTTOM FLUXES
c***********************************************************************
c           record 1: ndata - number data to read
c           do i = 1,ndata {
c              record(i+1): time(i) - time-step (days)
c                           column with bottom fluxes (mm/day)
c                           }
c
c note : input will be assumed constant for time(i-1) < time step <=
c        time(i)
c********************* end file bottom fluxes **************************
c
c Output:
c***********************************************************************
c          1) outfile:
c             file with a time series of the following record:
c             record i = 1 to int(simtim/dstep):
c                   time    - time (days)
c                   pnet    - net precipitation P (mm/day)
c                   gwlev1  - groundwater level (m reference level)
c                   ss1     - mean soil saturation (-)
c                   dflux   - discharge flux (mm/day)
c                             (positive when out of system)
c                   plas    - overland flow (mm/day)
c                   rest    - rest term of water balance (Eq. 1) (mm/day)
```

```
c*********************************************************************
c          2) balance.out:
c             file with a time series of cumulative water balance
c             components
c             record i = 1 to int(simtim/dstep):
c                   time      - time (days)
c                   stgw      - cumalitive storage in groundwater (mm)
c                   stsoil    - cumalitive storage in soil moisture (mm)
c                   cumpnet   - cumlative amount of net precipitation
c                               (mm)
c                   cumdflux  - cumalitive amount drained to water
                                courses (mm)
c                   cumvflux  - cumalitive amount from/to deeper
c                               groundwater (mm)
c                   cumplas   - cumulative amount of generated surface
c                               water (mm)
c                   cumrest1  - rest term of cumulative water balance
c                               (mm)
c*********************************************************************
c          3) kalfile:
c                   time      - time (days)
c                   gw0       - time update (m below surface)
c                   gw1       - measurement update (m below surface)
c                   sgw0      - error variance time update (m2/day2)
c                   sgw1      - error variance measurement update
c                               (m2/day2)
c             when measurements are available:
c                   xmeas     - measured value (m below surface)
c                   (xmeas - gw0) - innovation (m below surface)
c*********************************************************************
c          4) optfile:
c                   iter          - iteration number
c                   do i = 1,nwc+4 {
c                      par(i)     - parameters
c                {
c                   funk(par)        - value of criterium
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Following are parameter files (parfile1 and parfile2) for SSDKOPT

```
sdec1510.out
0 -79359
14
      20.0         4   debilt.cal
15
       0.0       0.5       1.0         5  debilt.cal
0.433  0.0611  0.00286  0.943  0.05
0.0
-1.30
0
1
-1.60 100.
1
0
2.0
1. 4017. 1 4 100
```

parfile1

```
32cl3415.10
sdec15.k10
sdec15.o10
114.466400 0.000640 2.331859 28.383366 0.
1 1 1 1 0
-1.30  0.0
0.
0.0001
```

parfile2

The program SSDPATH is used for simulating a single realisation, not only of the water table depth but of all water balance terms.

Header of the program SSDPATH with input instructions

```
C%%%
```

```
c          record 4) : if ipflag .eq. 0 then
c                         xmp    - mean precipitation (mm/day)
c                         vp     - variance precipitation (mm2/day2)
c                      else
c                         vp     - variance of noise on precipitation
c                                  (mm2/day2)
c                                  must be at first 10 positions
c                         icolnr - column nummer that must be read from
c                                  file must be at positions 11-20
c                          precname - name of precipitation input file
c                      endif
c          record 5) : iepflag  - if zero, potential evapotranspiration
c                                 is constant, if not zero it is the unit
c                                 number under which the potential
c                                 evapotransipration is read.
c          record 6) : if iepflag .eq. 0 then
c                         xmep  - mean potential evapotranspiration
c                                 (mm/day)
c                         vep   - variance potential evapotranspiration
c                                 (mm2/day2)
c                         exps  - exponent of the relation average soil
c                                 saturation and actual
c                                 evapotranspiration (-)
c                         cropf - crop factor (1.0 for grass)
c                      else
c                         input file
c                         vep  -  variance of noise potential evapo-
c                                 transpiration (m2/day2)
c                                 must be at positions 1-10
c                         exps -  exponent of the relation average soil
c                                 saturation and actual
c                                 evapotranspiration (-) must be at
c                                 positions 11-20
c                         cropf - crop factor (1.0 for grass)
c                                 must be at positions 21-30
c                         icolnr - column number to read from file
c                                  must be at positions 31-40
c                         evapname - name of potential evapotranspiration
c                                    input file
c                      endif
c
c          soil physical parameters (modified van Genuchten model)
c          record 7):
c                      thesat  - saturated volumetric water content (-)
c                      theres  - residual volumetric water content (-)
c                      alpha   - alpha value of model (positive) (1/cm)
c                      expn    - exponent of the modified van genuchten
c                                relation (-)
c                      eps0    - elastic (residual) groundwater storage
c                                (-)
c
c          geohydrological parameters and inputs
c          record 8):
c                      xsurf   - surface elevation level (m reference
c                                level)
c          record 9):
c                      gwstart - groundwater level at start simulation
c                                (m reference level)
c          record 10)  iwflag  - if zero, water levels are constant,
c                                if not zero it is the unit number under
c                                which the water levels of nearby water
c                                courses are read.
```

```
c                     if iwflag .eq. 0 then
c          record 11):    nwc       - number of different drainage levels
c                                      (maximum of 3)
c          record 12):
c                         do i = 1,nwc {
c                           wl(i)     - water level of nearby water
c                                       course I (m reference level)
c                             gamma(i)  - drainage resistance of ith water
c                                         course (days)
c                          }
c                     else
c          record 11):    nwc       - number of different drainage levels
c                                      must be placed in first ten
c                                      positions
c                         wlname    - name of file with water levels
c          record 12):
c                         do i = 1,nwc {
c                            gamma(i) - drainage resistance of level i
c                                       (days)
c                         }
c                     endif
c          record 13)  idrcode (i=1,nwc)
c                                - if one, the water course is both
c                                  draining and infiltrating; if zero the
c                                  water course is only draining, i.e. when
c                                  the groundwater table is above the
c                                  drainage level of this water course.
c          record 14)  ivflag  - if zero, vertical fluxes are constant,
c                              if not zero it is the unit number under
c                              which the vertical fluxes are read.
c          record 15)  if ivflag .eq. 0 then
c                         vflux  - flux to/from deeper groundwater
c                                  (mm/day); positive value for
c                                  exfiltration (upwards)
c                      else
c                          vfname - neaem of file with vertical fluxes
c                      endif
c          parameters for numerical integration:
c         record 16):
c                      dstep   - time step (days)
c                      simtim  - total integration time (days)
c                      iwrite  - frequency (time steps) of writing the
c                                results
c                      intflag - if equal to 1 a first order Runge-Kutta
c                                integration will be performed, otherwise
c                                a fourth order Runge Kutta will be used.
c                      iint    - number of integration steps per time
c                                step not exceeding MAXINT
c
c******************** end parameter file ********************************
c
c************************************************************************
c    FILES WITH METEODATA - precipitation or potential evapotranspiration
c************************************************************************
c           record 1: ndata - number data to read
c                     ncols - number of columns present
c                     cfac  - multiplication factor (for unit conversion)
c                               if right unit: cfac = 1
c           do i = 1,ndata {
c              record(i+1): time(i) - time-step (days)
c                           ncols columns (one of which is precipition or
c                                          potential evapotranspiration)
c                         }
```

```
c note : input will be assumed constant for time(i-1) < timestep <=
c        time(i)
c*************************************************************************
c                           FILE WITH WATER LEVELS
c*************************************************************************
c           record 1: ndata - number data to read
c           do i = 1,ndata {
c              record(i+1): time(i) - time-step (days)
c                           nwc columns with water levels (m with respect
c                               to reference level; i.e. one column for
c                               every level of water course)
c                           }
c note : input will be assumed constant for time(i-1) < time step <=
c        time(i)
c*************************************************************************
c                           FILE WITH BOTTOM FLUXES
c*************************************************************************
c           record 1: ndata - number data to read
c           do i = 1,ndata {
c              record(i+1): time(i) - time-step (days)
c                           column with bottom fluxes (mm/day)
c                           }
c note : input will be assumed constant for time(i-1) < time step <=
c        time(i)
c*************************************************************************
c Output:
c          1) outfile:
c             file with a time series of the following record:
c             record i = 1 to int(simtim/dstep):
c                   time    - time (days)
c                   pnet    - net precipitation P (mm/day)
c                   gwlev1  - groundwater level (m)
c                   ss1     - mean soil saturation (-)
c                   dflux   - discharge flux (mm/day)
c                            (positive when out of system)
c                   plas    - overland flow (mm/day)
c                   rest    - rest term of water balance (Eq. 1) (mm/day)
c          2) balance.out:
c             file with a time series of cumulative water balance
c             components
c             record i = 1 to int(simtim/dstep):
c                   time      - time (days)
c                   stgw      - cumalitive storage in groundwater (mm)
c                   stsoil    - cumalitive storage in soil moisture (mm)
c                   cumpnet     - cumlative amount of net precipitation
c                                 (mm)
c                   cumdflux  - cumalitive amount drained to water
c                               courses (mm)
c                   cumvflux  - cumalitive amount from/to deeper
c                               groundwater (mm)
c                   cumplas   - cumulative amount of generated surface
c                               water (mm)
c                   cumrest1  - rest term of cumulative water balance
c                               (mm)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Example parameter file for SSDPATH

```
sdep15.k10
1 -79359 1000
14
28.437038          4   debilt.val
15
       0.0       0.5       1.0         5  debilt.val
0.433  0.0611  0.00286  0.943  0.000103
0.0
-1.30
0
1
-1.60 111.006554
1
0
2.409005
1. 2191. 1 4 100
```

The program SSDSIM is used for simulating multiple realisations of water table depth (m), soil saturation (-) and drainage discharge (mm/d).

Header of the program SSDSIM with input instructions

```
C%%%
```

```
c
c          general input:
c          record 1) : outfile1  - name binary output file gws
c          record 2) : outfile2  - name binary output file sat
c          record 3) : outfile3  - name binary output file dis
c
c          record 4) :
c                      irandom  - integer flag which is set to 0 if the
c                                 mean trajectory is required. All other
c                                 values lead to the simulation of random
c                                 paths.
c                      iseed    - seed for the random generator; value
c                                 must be negative; a value is always
c                                 needed
c                      nsim     - number of realisations to be simulated
c
c          meteo parameters:
c          record 4) : ipflag  - if zero, precipitation is constant,
c                                if not zero it is the unit number under
c                               which the precipitation is read.
c          record 5) : if ipflag .eq. 0 then
c                         xmp   - mean precipitation (mm/day)
c                         vp    - variance precipitation (mm2/day2)
c                      else
c                         vp       - variance of noise on precipitation
c                                    (mm2/day2)
c                                    must be at first 10 positions
c                         icolnr   - column nummer that must be read from
c                                    file must be at positions 11-20
c                         precname - name of precipitation input file
c                      endif
c          record 6) : iepflag  - if zero, potential evapotranspiration
c                                 is constant, if not zero it is the unit
c                                 number under which the potential
c                                 evapotransipration is read.
c          record 7) : if iepflag .eq. 0 then
c                         xmep  - mean potential evapotranspiration
c                                 (mm/day)
c                         vep   - variance potential evapotranspiration
c                                 (mm2/day2)
c                         exps  - exponent of the relation average soil
c                                 saturation and actual
c                                 evapotranspiration (-)
c                         cropf - crop factor (1.0 for grass)
c                      else
c                         input file
c                         vep  -  variance of noise potential evapo-
c                                 transpiration (m2/day2)
c                                 must be at positions 1-10
c                         exps -  exponent of the relation average soil
c                                 saturation and actual
c                                 evapotranspiration (-) must be at
c                                 positions 11-20
c                         cropf - crop factor (1.0 for grass)
c                                 must be at positions 21-30
c                         icolnr - column number to read from file
c                                  must be at positions 31-40
c                         evapname - name of potential evapotranspiration
c                                    input file
c                      endif
c
```

```
c          soil physical parameters (modified van Genuchten model)
c          record 8):
c                      thesat  - saturated volumetric water content (-)
c                      theres  - residual volumetric water content (-)
c                      alpha   - alpha value of model (positive) (1/cm)
c                      expn    - exponent of the modified van genuchten
c                                relation (-)
c                      eps0    - elastic (residual) groundwater storage
c                                (-)
c
c          geohydrological parameters and inputs:
c          record 9):
c                      xsurf   - surface elevation level (m reference
c                                level)
c          record 10):
c                      gwstart - groundwater level at start simulation (m
c                                reference level)
c          record 11)  iwflag  - if zero, water levels are constant,
c                                if not zero it is the unit number under
c                                which the water levels of nearby water
c                                courses are read.
c                      if iwflag .eq. 0 then
c          record 12):    nwc       - number of different drainage levels
c                                     (maximum of 3)
c          record 13):
c                         do i = 1,nwc {
c                            wl(i)    - water level of nearby water
c                                       course I (m reference level)
c                            gamma(i) - drainage resistance of ith water
c                                       course (days)
c                          }
c                      else
c          record 14):    nwc       - number of different drainage levels
c                                     must be placed in first ten
c                                     positions
c                         wlname    - name of file with water levels
c          record 15):
c                         do i = 1,nwc {
c                            gamma(i) - drainage resistance of level i
c                                       (days)
c                         }
c                      endif
c          record 16)  idrcode (i=1,nwc)
c                               - if one, the water course is both
c                                 draining and infiltrating; if zero the
c                                 water course is only draining, i.e. when
c                                 the groundwater table is above the
c                                 drainage level of this water course.
c          record 17)  ivflag  - if zero, vertical fluxes are constant,
c                                if not zero it is the unit number under
c                                which the vertical fluxes are read.
c          record 18)  if ivflag .eq. 0 then
c                         vflux  - flux to/from deeper groundwater
c                                  (mm/day); positive value for
c                                  exfiltration (upwards)
c                      else
c                          vfname - neaem of file with vertical fluxes
c                      endif
c
c         parameters for numerical integration:
```

```
c        record 19):
c                       dstep   - time step (days)
c                       simtim  - total integration time (days)
c                       iwrite  - frequency (time steps) of writing the
c                                 results
c                       intflag - if equal to 1 a first order Runge-Kutta
c                                 integration will be performed, otherwise
c                                 a fourth order Runge Kutta will be used.
c                       iint    - number of integration steps per time
c                                 step not exceeding MAXINT
c          record 20):
c                       year1   - first year to be simulated (start at
c                                 januari 1)
c          record 21):
c                       startim - startup time (days)
c**********************************************************************
c    FILES WITH METEODATA - precipitation or potential evapotranspiration
c**********************************************************************
c           record 1: ndata - number data to read
c                     ncols - number of columns present
c                     cfac  - multiplication factor (for unit conversion)
c                             if right unit: cfac = 1
c           do i = 1,ndata {
c              record(i+1): time(i) - time-step (days)
c                           ncols columns (one of which is precipition or
c                                          potential evapotranspiration)
c                           }
c note : input will be assumed constant for time(i-1) < timestep <=
c        time(i)

c**********************************************************************
c                           FILE WITH WATER LEVELS
c**********************************************************************
c           record 1: ndata - number data to read
c           do i = 1,ndata {
c              record(i+1): time(i) - time-step (days)
c                           nwc columns with water levels (m with respect
c                               to reference level; i.e. one column for
c                               every level of water course)
c                           }
c note : input will be assumed constant for time(i-1) < time step <=
c        time(i)
c**********************************************************************
c                           FILE WITH BOTTOM FLUXES
c**********************************************************************
c           record 1: ndata - number data to read
c           do i = 1,ndata {
c              record(i+1): time(i) - time-step (days)
c                           column with bottom fluxes (mm/day)
c                           }
c
c note : input will be assumed constant for time(i-1) < time step <=
c        time(i)
c**********************************************************************
c
c Output:
c
c unformatted files (can be used as input for STATSIM)
c outfile 1:
c             record 1: year1 - first year of simulated realisation
c                       nsim - number of simulated realisations
c                       nyears - number of years per realisation
```

```
c             do k = 1,nsim
c                do i = 1,nsteps
c                   record 1 + (k-1)*nsim + i:
c                            gw1 - simulated water table depth (m)
c                enddo
c             enddo
c note: nsteps = int(simtim/dstep)
c The same format for
c outfile2:        ss1 - soil saturation
c outfile3:        dflux - drainage flux (mm/d)
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Example parameter file for SSDSIM

```
sdes1510.hea
sdes1510.sat
sdes1510.dis
1 -79359 1000
14
28.383366          4   debilt.val
15
       0.0       0.5       1.0         5  debilt.val
0.433  0.0611  0.00286  0.943  0.000103
0.0
-1.30
0
1
-1.60 111.006554
1
0
2.409005
1. 2191. 1 4 100
1991
365
```

Part 4: Documentation EMERALD

A program for modelling the water table depth using a physically based stochastic model

1. Model, input variables and parameters

1.1 Model description

An extensive description of the model, its derivation and use can be found in:
English
Walvoort, D.J.J. and M.F.P. Bierkens, 1999. *A stochastic modelling approach for rapid assessment of groundwater dynamics.* Report 171, DLO Staring Centrum, Wageningen.
Dutch
Bierkens, M.F.P. en D.J.J. Walvoort, 1998. Eenvoudige stochastische modellen voor grondwaterstandsfluctuaties. Deel 2: Gecombineerd bodem-grondwatermodel met stochastische invoer. *Stromingen* 4(3), 5-20.

Here follows a brief description of the model and its parameters.

The basic model used in EMERALD (as applied in VIDENTE) is the following physically based transfer function-noise model:

$$h_k = h_k^* + n_k \tag{1}$$

$$h_k^* = h_s + a(h_{k-1}^* - h_s) + b_1 q_n(k) + b_2 q_n(k-1) \tag{2}$$

$$n_k = f\, n_{k-1} + e_k \tag{3}$$

with

k	discrete time steps of size Δt, i.e. $t = k\Delta t, k = 0,1,2,\ldots$;
h_k	water table depth [L] at time step k;
h_k^*	deterministic part of the water table depth [L] at time step k;
n_k	auto-regressive noise process [L] at time step k;
e_k	zero mean discrete Gaussian white noise process at time step k;
$q_n(k)$	average net input to the groundwater system between $(k-1)\Delta t$ and $k\Delta t$ [LT^{-1}] (= groundwater recharge + seepage – capillary rise);
h_s	surface water level [L] (assumed constant);
a	auto-regressive parameter of the deterministic model [-];
b_1, b_2	moving average parameters of the deterministic model [T];
f	auto-regressive parameter of the noise model [-];
s_e^2	variance of the white noise process e_k [L^2].

Equation (1) shows that the water table depth consists of a “deterministic” part (Equation (2)), which is explained by variations of net groundwater input, and a “stochastic” part (Equation (3)) which describes the part that is not explained by precipitation surplus. Model (1)-(3) model is therefore a stochastic model and h_k a stochastic process. In chapter 2 the nature of stochastic processes is described.

The deterministic model (2) that describes the water table fluctuation as a function of net input is based on a solution to the linearised Boussinesq Equation as provided by Kraijenhoff van de Leur (1958), Maasland (1959) and a further analysis of their solutions by de Zeeuw (1966). For a derivation of Equation (2) we refer to Walvoort and Bierkens (1999). Figure 1 shows the hydrological system under study.

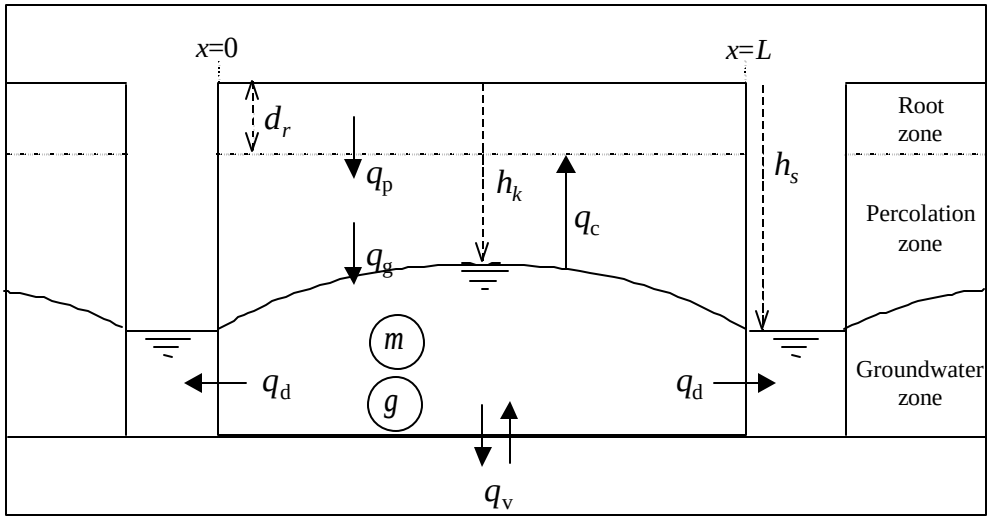


Figure 1. Schematic representation of hydrological system modelled with Equation (2)

The following additional variables and parameters are shown in Figure 1:

d_r thickness of the root zone [L];

$q_p(k)$ average percolation flux [LT^{-1}] between k-1 and k, which is the water that percolates from the root zone when it is above field capacity;

$q_g(k)$ average groundwater recharge [LT^{-1}] between k-1 and k;

$q_c(k)$ average capillary rise [LT^{-1}] between k-1 and k;

q_v flux from/to the deeper groundwater system. [LT^{-1}]. This flux, which is assumed constant, can both be positive (seepage or exfiltration) and negative (downward seepage or infiltration);

$q_d(k)$ average drainage flux to surface water [LT^{-1}] between k-1 and k. This flux can also be negative if the water table falls below the surface water level;

m specific yield (effective porosity) [-];

g drainage resistance [T] (here $g = (L^2/8kD)$, where kD is the transmissivity);

x position along the strip of land with respect to the surface waters varying from 0 (left) to L (right), where L is the distance between the surface waters.

Notice that in this scheme both the water table as the surface water level are measured with respect to surface elevation. In the following we will describe the three zones depicted in Figure 1 in more detail. The models of the root zone and the percolation zone are almost entirely based on the work of Zwamborn (1995).

Root zone

The root zone is modelled as a reservoir that overflows when it is full (see Figure 2).

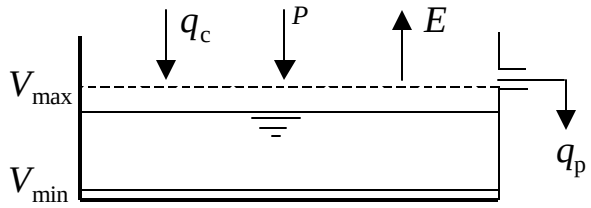


Figure 2 Schematic representation of the root zone

For each time step the volume of moisture is calculated according to the following balance equation:

$$V_k = V_{k-1} + [P(k) - E_a(k) + q_c(k) - q_p(k)]\Delta t \quad (4)$$

where

V_k volume of soil moisture stored in root zone [L];
$P(k)$ average precipitation minus interception losses [LT^{-1}] between k-1 and k;
$E_a(k)$ average actual evapo-transpiration [LT^{-1}] between k-1 and k.

The upper and lower boundaries V are governed by the soil moisture content at field capacity q_{fc} [-] and wilting point q_{wp} [-] respectively:

$$V_{max} = d_r q_{fc} \quad (5)$$
$$V_{min} = d_r q_{wp} \quad (6)$$

If $R(k)$ is the observed average precipitation between k and k-1 without interception losses is , $P(k)$ is calculated as:

$$P(k) = (1 - F_i)R(k) \quad (7)$$

where F_i is the interception fraction. The actual evapo-transpiration is calculated as:

$$E_a(k) = r(h_p) F_c E_r(k) \quad (8)$$

where

$E_r(k)$ average "Makkink reference crop evapo-transpiration" between k and k-1 [LT^{-1}], which gives the transpiration of a full grass cover under optimal conditions of water supply;
F_c crop factor [-];
$r(h_p)$ reduction function giving the reduction in potential evapo-transpiration as a function of average pressure head h_p [L] in the root zone.

Table 1 list for a number of different land use classes the crop factor F_c as well as the interception fraction F_i. Figure 3 shows the reduction function that has two parameters:

h_{lp} pressure head at limiting point [L]. At this pressure head the evapo-transpiration becomes smaller than potential (usually -500 cm);

h_{wp} pressure head at wilting point [L]. At this pressure head the evapo-transpiration becomes zero (usually -8000 cm);

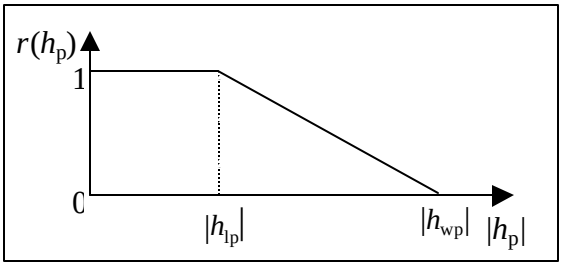


Figure 3 Evapo-transpiration reduction function

Pressure head is dependent on the moisture content of the root zone and is calculated according to Van Genuchten (1980):

$$h_p = \frac{1}{a}\left(\left(\frac{q - q_r}{q_s - q_r}\right)^{\frac{n}{n-1}} - 1\right)^{\frac{1}{n}} \qquad (9)$$

where a [L^{-1}], n [-], q_s [-] and q_r [-] are Van Genuchten parameters, in particular q_s is the saturated moisture content and q_r the residual moisture content of the soil. The Van Genuchten parameters for several representative soils have been tabulated by Wösten *et al*. (1994) and are used by EMERALD by reading a file "staring.dat". From Van Genuchten's relationship we can also obtain the soil moisture at field capacity q_{fc} (for Equation (5)), i.e. the soil moisture content at h_p = -100 cm. This file can be edited to add Van Genuchten parameters of other soil types or samples. For the moisture content needed in (9) the average moisture content of the root zone is taken as calculated from the storage at the previous time step:

$$q = \frac{V_{k-1}}{d_r} \qquad (10)$$

Finally, to complete the water balance the capillary rise has to be accounted for. The capillary rise depends on the water table depth. As the water table depth is the object of our calculation, the capillary rise is calculated from the water table depth at the previous time step. First the maximum possible capillary rise is calculated as a function g_c of h_{k-1}:

$$q_c^{max}(k) = g_c(h_{k-1}) \qquad (11)$$

The function g_c is a logistic growth function of h whose form also depends on the soil moisture retention curve and the unsaturated conductivity curve. These are both parameterised with Van Genuchten (1980) parameters and can thus be obtained for the representative soils tabulated by Wösten *et al*. (1994) and the "staring.dat". In Walvoort and Bierkens (1999) the function g_c is given. If h_{k-1} is too deep below the root zone, the capillary rise becomes zero.
Of course capillary rise can only occur if no percolation takes place, and if it occurs it can never be so large that it exceeds the storage capacity of the root zone. Therefore, the actual amount of capillary rise is given by:

$$q_c(k) = \begin{cases} 0 & \text{if } V_{k-1} + [P(k) - E_a(k)]\Delta t > V_{max} \\ \dfrac{V_{max} - \left(V_{k-1} + [P(k) - E_a(k)]\Delta t\right)}{\Delta t} & \text{if } V_{k-1} + [P(k) - E_a(k) + q_c^{max}]\Delta t > V_{max} \\ q_c^{max} & \text{if } V_{k-1} + [P(k) - E_a(k) + q_c^{max}]\Delta t \le V \end{cases} \quad (12)$$

If capillary rise occurs no percolation occurs and vice versa. Percolation is therefore given by:

$$q_p(k) = \begin{cases} \dfrac{\left(V_{k-1} + [P(k) - E_a(k)]\Delta t\right) - V_{max}}{\Delta t} & \text{if } V_{k-1} + [P(k) - E_a(k)]\Delta t > V_{max} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

Percolation zone
In case of shallow water tables, the percolated water is added directly to the groundwater as groundwater recharge: $q_r(k) = q_p(k)$. However, if the unsaturated zone is large, the percolated water will be significantly attenuated and delayed before it reaches the groundwater. Flow through the unsaturated zone is described with Richards' Equation. To avoid solving this equation Zwamborn (1995) described the attenuation and delay using a convolution equation:

$$q_g(k) = \sum_{i=1}^{k} q_p(i) U(\Delta t, k - i + 1) \quad (14)$$

where $U(\Delta t, k)$ is the pulse response of the percolation zone, i.e. the outflow of a linear system due to an input of unit volume occurring at a uniform rate for a period Δt. The pulse response can be found by linearising Richards' equation around the effective soil moisture content q_{eff} (see Walvoort and Bierkens (1999) for a derivation). If the water table is deep enough a fair assumption is that the downward flux in the percolation zone is equal to the unsaturated conductivity (gravity flow). So the effective soil moisture content is that soil moisture content for which the unsaturated conductivity is equal to the long year averaged groundwater recharge $<q_g>$ (which is also equal to the long year average percolation):

$$q_{eff} \Leftarrow k(q_{eff}) = < q_g > \quad (15)$$

Wesseling (1991) gives the long year average soil water balance for various combinations of land use and soil type. From these water balances the long year average groundwater recharge can be obtained. The effective soil moisture content follows from the unsaturated conductivity curve and the soil moisture retention curve, which again can be obtained from the Van Genuchten parameters and a representative soil type (Wösten *et al.*, 1994).
In case more layers are present in the percolation zone, for each layer the effective soil moisture content can be found and from this the pulse response. So, Equation (14) is applied to each layer separately, while the outflow from one layer is used as inflow from the layer below:

$$q_j(k) = \sum_{i=1}^{k} q_{j-1}(i) U_j(\Delta t, k-i+1) \qquad j = 1,2,...,n_L \tag{16}$$

where n_L is the number of layers, q_j the outflow from layer j and $U_j(\Delta t, k)$ its pulse response. Note that $q_0 = q_p$ and $q_{n_L} = q_g$.
It is also possible to assign for each layer i a fraction f_b as bypass flow. This is flow that is not delayed or attenuated but transferred to the next layer within the same time step. This fraction can be seen as preferential flow through macropores or along roots.

Groundwater zone
The net input to the groundwater system at any time step k is given by:

$$q_n(k) = q_g(k) - q_c(k) + q_v \tag{17}$$

The varying water table due to this input is given by Equation (2). The parameters have a physical meaning as Equation (2) is an approximation to a solution of the linearised Boussinesq Equation provided by Kraijenhoff van de Leur (1958), Maasland (1959) and a further analysis of their solutions by de Zeeuw (1966) (see Walvoort and Bierkens (1999) for its derivation). The resulting expressions for the parameters of (2) are:

$$a = \exp\left(-\frac{p^2 \Delta t}{8mg}\right) \tag{18}$$

$$b_1 = \frac{32}{p^3} g (1-a) \sin\left(\frac{px}{L}\right) + \Delta U_h \tag{19}$$

$$b_2 = -a \Delta U_h \tag{20}$$

where

$$\Delta U_h = \frac{32}{p^3} g \sum_{n=3,5,7,..}^{\infty} n^{-3} (1-a^{n^2}) \sin\left(\frac{npx}{L}\right)(\tag{21}$$

As can be seen, the dynamics of the water table depth depend on the distance of the location to the surface waters. In most sandy soils, the largest component in the drainage resistance is the resistance to flow close to the surface waters (including entrance resistance), instead of the resistance to horizontal flow. In this case, the form of the water table is more like a mesa (steep gradients close to the water courses, almost no gradients elsewhere) rather than a sinus. It is best not to take account of the distance in this case and use $x=L/2$, which means that the sinus terms disappear from above equations. This is the standard setting in VIDENTE. The distance to the water course is important in the following cases only: clayey soils and unconfined sandy aquifers with deep water tables and large distances between water courses.
Because we are working with a physically based model of groundwater flow it is also possible to obtain an expression like equation (2) for the drainage discharge to the water courses [LT^{-1}] (Walvoort and Bierkens, 1999):

$$q_d^*(k) = a q_d^*(k-1) + b_1 q_n(k) + b_2 q_n(k-1) \tag{22}$$

with

$$a = \exp\left(-\frac{p^2 \Delta t}{8mg}\right) \tag{23}$$

$$b_1 = \frac{8}{p^2}(1-a) + \Delta U_q \tag{24}$$

$$b_2 = -a\Delta U_q \tag{25}$$

where

$$\Delta U_q = \frac{8}{p^2} \sum_{n=3,5,7,..}^{\infty} n^{-2}(1-a^{n^2}) \tag{26}$$

1.2 Input variables and parameters / invoervariabelen en parameters

Next we list the input variables and parameters used in the program EMERALD (units between square brackets). These will be given in English and in Dutch.

English:
Input variables

$P(k)$ average precipitation minus interception (mm/d) between $(k-1)$ and k, calculated from the measured precipitation with Equation (7). If interception is accounted for, the calculation should be performed outside of EMERALD en $P(k)$ offered to EMERALD in the input file. Interception fractions for various types of land use are given in Table 1;

$E_r(k)$ average Makkink reference crop evapo-transpiration between k and k-1, used to calculate actual evapo-transpiration using Equation (8).

soil and land use parameters

For top soil (root zone) and each subsoil layer (each layer in percolation zone):

Code code of reference soil type (texture class) used in the file "staring.dat" (see Wösten et al., 1994 for an explanation of these codes). For each code a set of van Genuchten parameters is read that is used to characterise the soil physical properties of the layer (e.g. Equations (9) and (15)). Staring.dat can be edited to include the Van Genuchten parameters of additional texture classes of those obtained from individual samples;

Thickness thickness of layer (cm)

f_b bypass fraction (between 0 and 1). Gives for each layer the fraction of percolation water that is preferential flow. This fraction is transferred without delay and attenuation to the next layer. Not used for the root zone layer. For shallow water tables (< 150 cm below surface) it is advised to set this parameter at 1 for all layers;

F_c crop factor [-] (see Equation 8). Table 1 gives for a number of different land use classes the crop factor F_c;

h_{fc} pressure head at field capacity (cm) (see Equation 5). Can be given as positive (pressure head) or negative value (suction head). A typical value is 100 cm;

h_{lp} ET reduction limiting point (cm): pressure head at which the evapo-transpiration becomes limited due to moisture shortage (Equation 8 and Figure 3). Can be given as positive or negative value. A typical value is 500 cm;

h_{wp} ET wilting point (cm): pressure head at which evapo-transpiration comes to a stop (see Equations 6 and 8 and Figure 3). Can be given as positive or negative value. A typical value is 8000 cm;

Hydrological parameters

h_s drainage level (cm) with respect to surface elevation (e.g. –150) (see Figure 1). This can either be a surface water level or trench bottom or drainage depth;

$<q_g>$ Long year average groundwater recharge (mm/year). This parameter is used to calculate the effective soil moisture q_{eff} of each layer of the percolation zone (Equation 15; used in Equations (14) and (16)). Wesseling (1991) gives typical values for various combinations of soil type and land use in the Netherlands (e.g. grassland on loamy medium sand: 261 mm/year). In case the bypass parameters for all layers are set to 1 this parameter is of no consequence;

g drainage resistance (days) of surface water system (see Figure 1). Typical values range between 30 and 500 days;

m specific yield (effective porosity) (-) (See Figure 1). Typical values range between 0.15 and 0.25;

q_v seepage/infiltration: bottom flux from/to deeper groundwater (mm/d) (see Figure 1). Typical values range between –1.0 (infiltration) and 3.0 mm/d (seepage or exfiltration);

noise parameters (see Equation 3)

f auto-regressive parameter of the noise model (between 0 and 1); Typical values range between 0.5 and 0.99;

s_e^2 variance of the white noise process e_k (cm^2). Typical values range between 5 and 50 cm^2;

Nederlands:

Input variables

$P(k)$ Gemiddelde neerslag min interceptie (mm/d) tussen tijdstappen $(k-1)$ and k. Deze kan berekend worden uit de gemeten neerslag via Vergelijking (7). Als rekening gehouden moet worden met interceptie dan moet de evaluatie van Vergelijking (7) buiten EMERALD plaatsvinden $P(k)$ in de invoerfile van EMERALD worden gezet. Tabel 1 geeft interceptiefracties voor verschillende landgebruiksvormen;

$E_r(k)$ gemiddelde Makkink referentieverdamping tussen k and k-1. Hieruit wordt door EMERALD de werkelijke verdamping berekend via Vergelijking (8);

Bodem- en landgebruiksparameters

Voor de bovengrond (wortelzone) en voor elke ondergrondlaag (percolatiezone):

Code code van de textuurklasse (bouwsteen) van de Staringreeks in "staring.dat" (codes worden uitgelegd in Wösten et al., 1994). Voor elke code wordt uit staring.dat een set Van Genuchten parameters ingelezen waarmee de hydraulische eigenschappen van de bodemlagen kunnen worden gekarakteriseerd (bijv. in Vergelijkingen (9) and (15)). Staring.dat kan ook worden aangepast door er eigen codes en bijbehorende Van Genuchten parameters aan toe te voegen (bijv. individuele monsters);

Dikte laagdikte (cm);

f_b *bypass* fractie (tussen 0 and 1). Deze geeft per laag de fractie van het percolatiewater dat preferent stroomt (bijv. langs scheuren en wortels). Dit water wordt zonder vertraging en demping binnen dezelfde tijdstap toegevoegd aan de onderliggende laag. Voor grondwaterstanden ondieper dan 150 cm – maaiveld wordt aangeraden om voor alle lagen een *bypass* fractie van 1 te hanteren;

F_c gewasfactor [-] (zie Vergelijking 8). Tabel 1 geeft gewasfactoren voor verschillende landgebruikstypes

h_{fc} drukhoogte bij veldcapaciteit (cm) (zie Vergelijking 5). Kan zowel als een negatieve waarde (drukhoogte) als positieve waarde (zuigspanning) worden opgegeven. Een typische waarde is 100 cm;

h_{lp} Punt van verminderde verdamping: drukhoogte waarbij de actuele verdamping minder wordt dan de potentiële verdamping. Kan zowel als een negatieve waarde (drukhoogte) als positieve waarde (zuigspanning) worden opgegeven. Een typische waarde is 500 cm;

h_{wp} Verwelkingspunt: drukhoogte waarbij de verdamping gelijk aan 0 wordt. Kan zowel als een negatieve waarde (drukhoogte) als positieve waarde (zuigspanning) worden opgegeven. Een typische waarde is 8000 cm;

Hydrologische parameters

h_s drainageniveau (cm referentieniveau) (bijv. –150) (zie Figuur 1). Dit kan zowel een oppervlaktewaterstand of peil zijn als de bodemhoogte van een drain of een greppel;

$<q_g>$ Langjarig gemiddelde grondwateraanvulling (mm/jaar). Deze parameter is nodig om het effectieve watergehalte q_{eff} van de lagen van de percolatiezone te bepalen (Vergelijking (15); gebruikt om de pulsrespons in Vergelijkingen (16) en (17) te bepalen). Wesseling (1991) geeft typische waarden van deze parameter voor Nederlandse omstandigheden voor verschillende combinaties van bodem en landgebruik (bijv. grasland op lemig matig zand: 261 mm/jaar). Als de *bypass* parameter van alle bodemlagen gelijk is aan 1, dan doet deze parameter niet ter zake;

g drainageweerstand (dagen) van het (zie Figuur 1). Waarden variëren meestal tussen de 30 en 500 dagen;

m effectieve porositeit grondwaterzone (*specific yield* in Engels) (-) (zie Figuur 1). Waarden variëren meestal tussen 0.15 en 0.25;

q_v kwel/infiltratie: flux van/naar dieper grondwater (mm/d) (zie Figuur 1). Waarden van deze parameter variëren meestal tussen –1.0 (infiltratie) en 3.0 (kwel) mm/d;

ruisparameters (zie Vergelijking 3)

f autoregressieve parameter van het ruismodel (kan tussen 0 en 1 liggen, maar ligt meestal tussen 0.9 en 0.99);

s_e^2 variantie witte ruis proces e_k (cm^2). Waarden variëren meestal tussen 5 en 50 cm^2;

Table 1. Crop factors and interception fractions for various crops and land use (sources in footnotes)

Crop / land use	F_c	F_I
Grassland	0.96[1]	0.00
Potatoes	1.03[1]	0.00
Beets	0.98[1]	0.00
Grain	0.95[1]	0.00
Maize	0.92[1]	0.00
Other crops	0.88[1]	0.00
Fallow	0.70[2]	0.00
Deciduous wood	1.00[3]	0.20[3]
Coniferous wood	0.80[3]	0.40[3]
Tree nurseries	1.00[3]	0.20[3]
Other trees	0.90[3]	0.30[3]
Heath	0.70[2]	0.00
Wetland vegetation	1.04[4]	0.00
Dryland vegetation	0.70[2]	0.00
Other vegetation	0.87[5]	0.00
Orchard (soil not covered)	0.91[6]	0.06[7]
Orchard (grass strips)	0.97[6]	0.06[7]
Sports field	0.96[8]	0.00
Public garden	0.97[9]	0.06[9]
Horticulture under glass	1.30[10]	1.00
Horticulture not under glass	0.88[11]	0.00

[1] Crop factor of average growing season according to Feddes, R.A., 1997. Crop factors in relation to Makkink reference-crop evapo-transpiration. In: Verslagen en Mededelingen 39, pp. 33-45, CHO-TNO, The Hague.

[2] Jansen, P.C., 1986. De potentiële verdamping van (half-)natuurlijke vegetaties. ICW nota 1703, Wageningen (in Dutch).

[3] Moors, E.J., A.J. Dolman, W. Bouten en A.W.L. Veen, 1996. De verdamping van bossen H_2O 19(16), 462-466 (in Dutch). Furthermore, the parameters for "tree nurseries" have been taken the same as for deciduous wood, and those for "other wood" as the average of deciduous and coniferous wood.

[4] Seasonal average of Molinia from Moors, E.J., J.N.M. Stricker and G.D. van den Abeele, 1998. Evapo-transpiration of cut over bog covered by Molinea Caerulea. Agricultural University, Department of Environmental Sciences, report 73, Wageningen.

[5] Average of wetland and dryland vegetation.

[6] Assuming a tree coverage of 30% we calculate:

Orchard (soil not covered): $F_c = 0.7 \times 0.88$ (other crops) + 0.3×1.0(deciduous trees) = 0.91

Orchard (grass strips): $F_c = 0.7 \times 0.96$ (grassland) + 0.3×1.0(deciduous trees) = 0.97

[7] Assuming a tree coverage of 30% we also assume 30% of the interception of deciduous wood yielding an interception factor of $0.3 \times 0.20 = 0.06$.

[8] Taken the same as "grassland".

[9] Taken the same as "orchard (grass strips)"

[10] All water is intercepted. The evapo-transpiration for horticulture under glass in the western part of the Netherlands is about 700 mm / year, which is about 1.3 times the average reference evapo-transpiration for that area (personal communication Philip Hamaker). The excess water needed is obtained from the intercepted precipitation water that is collected in tanks.

[11] Taken the same as "other crops".

2. Stochastic modelling: prediction, simulation and calibration

There are two different ways of using a stochastic model: prediction and stochastic simulation. In the following sections these two different ways are further explored. The section thereafter explains how some of the parameters of EMERALD can be estimated from time series of water table depth through calibration. Finally, the last section recapitulates how the application of EMERALD to a practical problem proceeds.

2.1 Prediction

Before we start with explaining prediction, we have to explain the nature of h_k when it is described with stochastic models. It means that we are uncertain about the exact variation of h with time. We do know that it is likely to be lower in the summer time and higher during the winter, but there is still a lot of unknown variation left. This variation is due to errors in our inputs, model parameters and the fact that our model itself is only an approximation of reality. Therefore, we vision that h_k is not described by a single function of time, but as a collection of possible functions, each of which is equally probable of describing the real but unknown variation of h. Figure 4 illustrates this concept, showing four equally possible functions (note that for convenience we have drawn them as continuous functions)

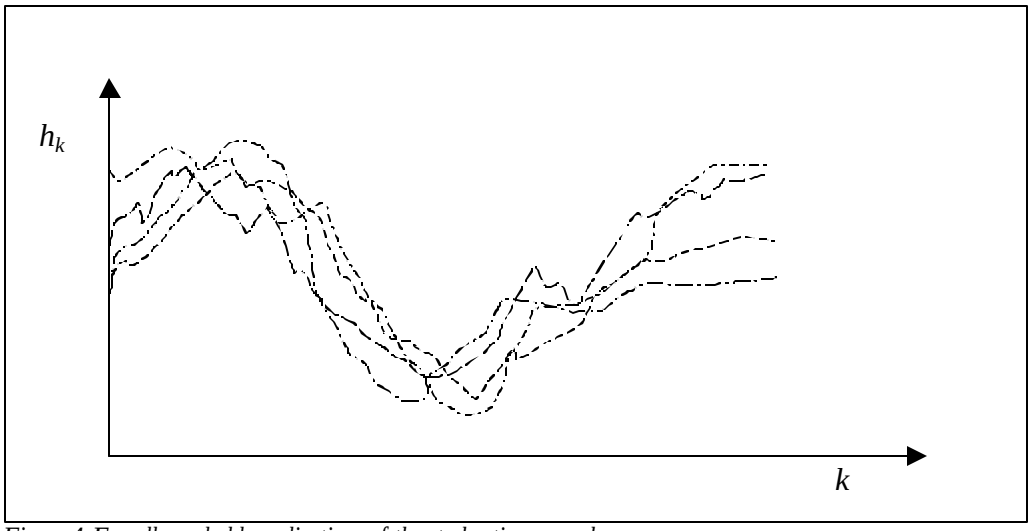


Figure 4 Equally probable realisations of the stochastic process h_k

One particular function is called “a realisation” and the whole collection of equally probable realisations, usually an infinite number of them, is called “the ensemble” and h_k is said to be “a stochastic process”. Now reality is assumed to be on of these possible realisations, however which one exactly is unknown. We would want our prediction to be such that the prediction error is minimal. However, because we do not know which of the realisations is reality it is not possible to evaluate the

prediction error either. Instead, we are forced to look at every time step at the difference between our prediction and the values of all possible realisations. Suppose that $\hat{h}_k$ is the prediction and $h_k^{(i)}$ the value of realisation number i, then we seek a prediction for which the following properties hold:

1. The prediction error, i.e. the difference between our prediction and the realisations, is on average equal to zero:

$$\sum_{i=1}^{\infty}\left[\hat{h}_k - h_k^{(i)}\right] = 0 \tag{27}$$

2. The average squared prediction error is minimal (as small as possible):

$$\sum_{i=1}^{\infty}\left[\hat{h}_k - h_k^{(i)}\right]^2 \Rightarrow \text{minimal} \tag{28}$$

It turns out that we achieve this if we take as prediction at every time the average of all realisations:

$$\hat{h}_k = E[h_k] = \sum_{i=1}^{\infty} h_k^{(i)} \tag{29}$$

The average of all possible realisations is called "the expectation" or "expected value" of the stochastic process h_k and is usually denoted with the operator $E[]$. If a prediction is such that it has properties (27) and (28), it is said to be "unbiased" (Equation (27)) and "optimal in least squared sense" (Equation (28). If we describe the stochastic process h_k with Equations (1)-(3) the expected value is the deterministic part of these models:

$$\hat{h}_k = c + a(\hat{h}_{k-1} - c) + b_1 q_n(k) + b_2 q_n(k-1) \tag{30}$$

Figure 5 shows the realisations with the optimal prediction. Also shown is the 95%-prediction interval which gives for every time step the boundaries that contain 95% of the realisations. Because the noise e_k has a Gaussian (normal) distribution, the 95% prediction interval can be calculated from the standard deviation of the prediction error $s_e(t)$ as follows:

$$\left[\hat{h}_k \pm 1.96 s_e\right]$$

with

$$s_e = \sqrt{\sum_{i=1}^{\infty}\left[\hat{h}_k - h_k^{(i)}\right]^2} \tag{31}$$

For the model (1)-(3) the variance of the prediction error s_e^2 is equal to the variance of the noise process n_k (s_n^2) and can be calculated as (Bierkens *et al.*, 1999):

$$s_e^2 = s_n^2 = \frac{s_e^2}{1 - f^2} \tag{32}$$

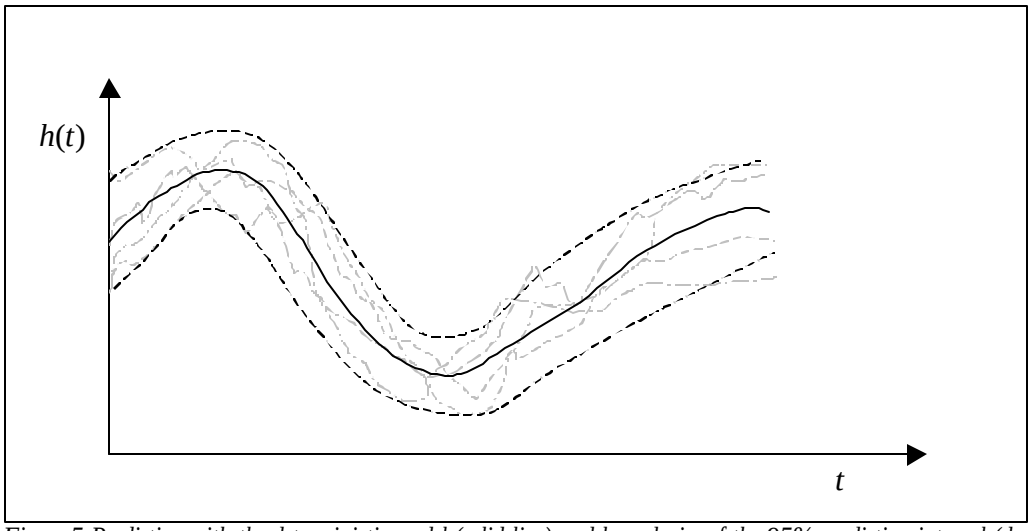


Figure 5 Prediction with the deterministic model (solid line) and boundaries of the 95% prediction interval (dashed lines); grey lines are a number of realisations.

2.2 Stochastic simulation

Prediction of water table depth is important if we are interested in the actual, but non-observed water table depth. However, in many applications we are not particularly interested in the actual water table depth, but in some fluctuation quantities. For instance, we may be interested in the probability that at any day in the near future the level of 30 cm is exceeded, or we want to know the mean highest water table depth (MHW) or mean lowest water table depth (MLW) (Van der Sluijs and De Gruijter, 1985). These fluctuation quantities cannot be estimated from the predicted water table depth, because the predicted line typically overestimates the low values and underestimates the high values (see Figures). The reason for this is that it tries to minimise the squared prediction error (Equation 31). In this case we should do the following (See Figure 6):

1. simulate a large number (at least 100) realisations of the stochastic process h_k using model (1-3) where the e_k are simulated by drawing from a Gaussian distribution with mean zero and variance s_e^2;
2. or each simulated realisation the appropriate fluctuation quantities are estimated. For 100 realisations this yields also 100 values of this property, e.g. 100 MHWs and 100 MLWs;
3. the cumulative frequency estimated from the replicas of the fluctuation quantity (e.g. the 100 MHWs and MLWs) represents an estimate of the probability

distribution of the fluctuation quantity. This probability distribution expresses the uncertainty about the true value of the fluctuation quantity, uncertainty that arises from our model's inability to predict the unknown water table depth exactly. Usually the average of the replicas is used as an estimate of the true but unknown fluctuation quantity, and as a measure of uncertainty the 95% confidence interval is calculated.

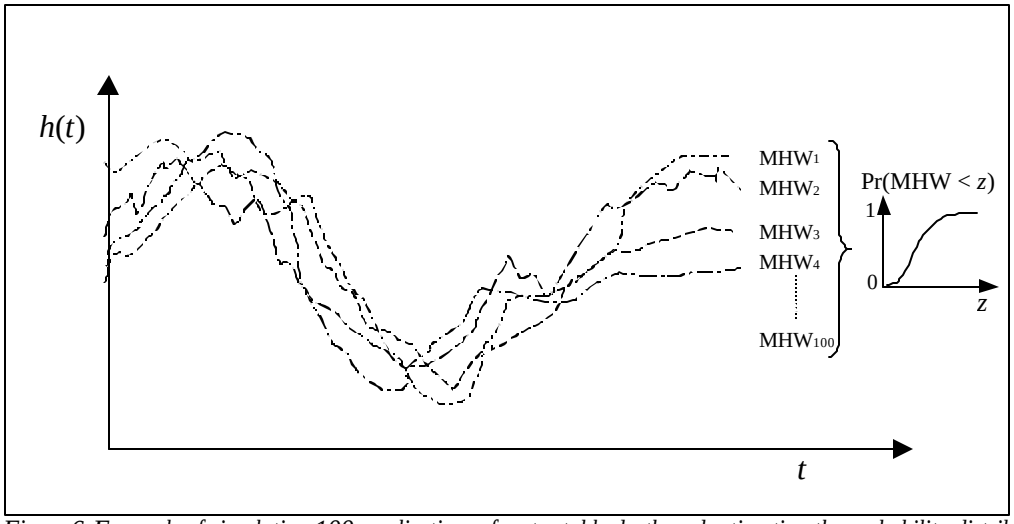


Figure 6 Example of simulating 100 realisations of water table depth and estimating the probability distribution of MHW from it.

2.3 Calibration

Most parameters can be obtained from the combinations of soil profile descriptions and land use classification in combination with standard values from the literature (e.g. Table 1 and Wösten *et al.*, 1994). However, the hydrological parameters $\{g, m, q_v\}$ and the noise parameters $\{f, s_e^2\}$ are difficult to estimate directly from physical knowledge and must generally be obtained from calibration. To obtain unbiased estimates of the parameters of a stochastic model like equation (1)-(3), they should be estimated simultaneously, e.g. using a state estimator such as the Kalman filter (see Te Stroet, 1995). Unfortunately, a Kalman filter as used in SSD (Bierkens, 1998) or KALMAX/KALTFN (Bierkens *et al*., 1999) has not been implemented yet for EMERALD. Therefore, we use a two step procedure that, in theory, may lead to unbiased estimates.

Step 1: calibration of the deterministic model (Equation (2))

The parameters $\{g, m, q_v\}$ that are part of the deterministic model (2) are found by running the deterministic model (2) only (prediction) and minimising the following criterion with respect to $\{g, m, q_v\}$ using L observations y_j, j=1,..,L of water table depth:

$$J(L;g,m,q_v) = \sum_{j=1}^{L} \left[h_j^*(g,m,q_v) - y_j \right]^2 \qquad (33)$$

where $h_j^*(g,m,q_v)$ is the prediction with the deterministic model with parameters $\{g,m,q_v\}$ at time step j where observation y_j has been taken. Minimisation of (33) is done with a minimisation algorithm. In EMERALD the "downhill simplex method" is used (Press *et al.*, 1986).

Step 2: calibration of the stochastic model (Equation (3))

With the estimated parameters $\{\hat{g},\hat{m},\hat{q}_v\}$ obtained from step 1 the parameters $\{f, s_e^2\}$ are estimated using the following steps:

1. calculate the L residuals from the L observations:

$$e_j = h_j^*(\hat{g},\hat{m},\hat{q}_v) - y_j \qquad j = 1,..,L \qquad (34)$$

2. calculate the auto-covariance function of the residuals. Assuming that observation intervals are more or less equidistant, this auto-covariance function is estimated as:

$$\hat{C}(k) = \frac{1}{L-k} \sum_{j=1}^{L-k} e_j e_{j+k} \qquad k = 0,1,...,k_{max} \qquad (35)$$

where k_{max} is the maximum lag calculated. From Equations (31), (32) and (35) it can be seen that $C(0) = s_e^2 = s_n^2$;

3. fit the theoretical auto-covariance function of the noise model (3) given by (Chatfield, 1989; p. 36):

$$C(k) = \frac{s_e^2}{1-f^2} f^{|k|} \qquad k = 0,1,..... \qquad (36)$$

to the experimental auto-covariance (35) by minimising the following weighted least squares criterion:

$$J(k_{max};f,s_e^2) = \sum_{k=0}^{k_{max}} w(k) \left[C(f,s_e^2;k) - \hat{C}(k) \right]^2 \qquad (37)$$

where the weights $w(k)$ are proportional to the number of data used in (35) to obtain estimate $\hat{C}(k)$. Minimisation of (37) is obtained with the downhill-simplex method (Press *et al.*, 1986). In VIDENTE the default value of k_{max} = 9 is used.

Clearly, there is an inconsistency in these steps. In step 1 a criterion is used (Equation 33) that assumed that the residuals e_j are independent, whereas in step 2 a covariance function is fitted that entails that the e_j are dependent in time. This inconsistency is

the cause of the bias in the parameter estimates. Criterion (33) yields biased estimates of $\{g, m, q_v\}$ if the e_j are dependent in time. These biased estimates are then used in step (2), yielding biased estimates of $\{f, s_e^2\}$ also. How serious this bias is depends on the length of the series of observations. If the time span over which observations have been taken is long compared to the effective correlation length, bias will be small. The effective correlation length, i.e. the time span over which the e_j are correlated (in days), is given by (Bierkens *et al.*, 1999):

$$q_c = \frac{-3\Delta t}{\ln(f)} \tag{38}$$

The correct way to estimate the parameters is to estimate $\{g, m, q_v\}$ with generalised least squares assuming dependent e_j (as opposed to ordinary least squares used in Equation (33)). However, for this we would need the parameters $\{f, s_e^2\}$ which are also unknown. Therefore some iterative scheme must be used where a generalised least square estimate of $\{g, m, q_v\}$ is used in step 1 that is alternated with improved estimates of $\{f, s_e^2\}$ as in step 2. The Kalman filter algorithm and a filter criterion according to Schweppe (1979) as used in Bierkens *et al*. (1999) also provides such unbiased estimates.

2.4 Systematic application of EMERALD: step by step

In the following example we assume that we have a location for which precipitation and reference evapo-transpiration are known for the years 1970-1999 and we have observations of water table depth available with an observation frequency of two times a month for the years 1995-1999. Application of EMERALD has two goals. First, we want to estimate fluctuation properties (MHW, MLW etc.) at the location. Second, we aim to monitor the water table depth on a daily basis in the future, while maintaining an observation frequency of two times a month. The following steps are taken:

2.4.1 Calibration

The parameters $\{g, m, q_v, f, s_e^2\}$ are estimated through calibrating the model to the observations from 1995-1999 (section 2.3). EMERALD calibrates the parameters of the deterministic and the stochastic model consecutively, where it possible to fix each of the parameters, if necessary. The result of the calibration with EMERALD is a file with calibrated parameters and the mean error (*ME*), root mean squared error (*RMSE*) and mean absolute error (*MAE*) of Kalman filter time updates:

$$ME = \frac{1}{L}\sum_{j=1}^{L}\left(\hat{h}_j - y_j\right) \tag{39}$$

$$RMSE = \sqrt{\frac{1}{L}\sum_{k=1}^{L}\left(\hat{h}_j - y_j\right)^2} \tag{40}$$

$$MAE = \frac{1}{L}\sum_{j=1}^{L}\left|\hat{h}_j - y_j\right| \tag{41}$$

Furthermore, if asked for in the EMERALD parameter file (standard in VIDENTE), the output file contains for the calibration period the predictions of water table depth (cm surface) and the following components of the water balance: $P(k)$ (mm/d), $E_r(k)$ (mm/d), $E_p(k)$ (mm/d), $E_a(k)$ (mm/d), V_k (mm), $q_p(k)$ (mm/d), $q_g(k)$ (mm/d), $q_c(k)$ (mm/d) and $q_n(k)$ (mm/d) (all for the optimised parameter set $\{\hat{g}, \hat{m}, \hat{q}_v\}$). For the observation times, the observations of water table depth y_k and the residuals y_k-h_k are also given. With the calibrated parameters $\{\hat{f}, \hat{s}_e^2\}$ the prediction variance can be estimated with Equation (32). Finally, the file contains the theoretical auto-covariance function calculated with $\{\hat{f}, \hat{s}_e^2\}$ as fitted to the experimental one. Figure 7 shows the prediction with the calibrated Equation (2) together with the observations. Also shown are residuals with the theoretical prediction interval (Equation 32). Figure 8 shows the theoretical and experimental auto-covariance function as obtained with the estimated parameters $\{\hat{f}, \hat{s}_e^2\}$.

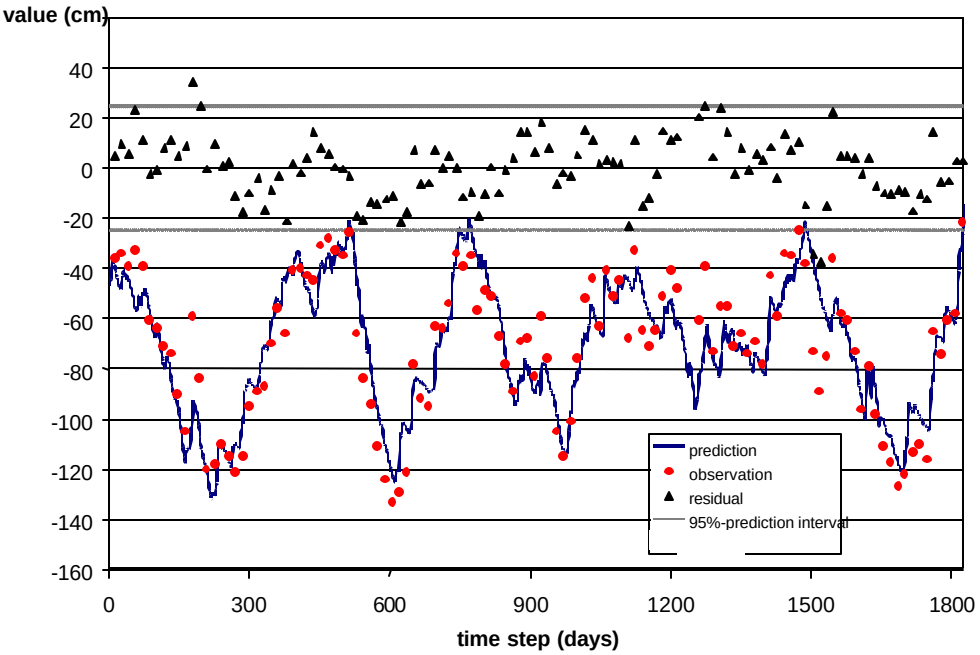


Figure 7 Example output from calibrating EMERALD

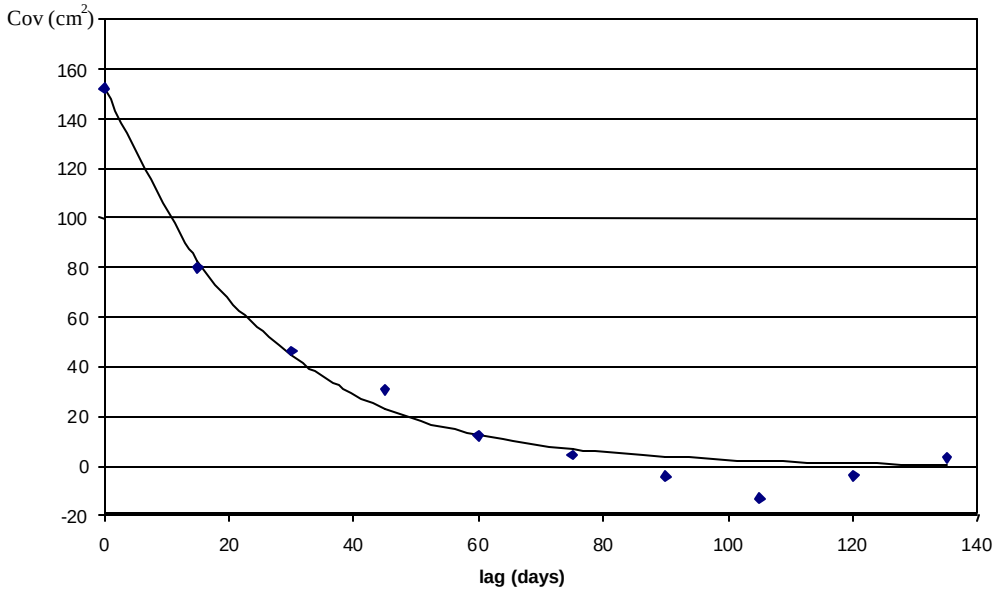


Figure 8 Example output from calibrating EMERALD; fitted auto-covariance (solid line) and experimental auto-covariance (diamonds).

2.4.2 Prediction

When calibrating with the stand alone version of EMERALD, one should indicate explicitly whether prediction has to be performed for the calibration period. In VIDENTE this is standard. It is advisable to always predict the actual water table depth for the calibration period also, because it enables you to plot a figure like Figure 7. This way you can see whether the calibrated model captures the dynamics satisfactorily. Of course, prediction is mostly used to obtain the water table depth and the terms of the water balance for periods without observations (we must of course have precipitation and potential evapo-transpiration for this period) . The output is the same as described under 2.4.1. If in the prediction period observations have been taken, EMERALD will use these to calculate statistics (39) to (41). Also, these observations and the residuals are put in the output file together with the predicted water table depths. This way a figure like Figure 7 can be made and the calibrated model be validated for a different period than the calibration period.

2.4.3 Simulation

Using the calibrated parameters from the calibration step EMERALD can be used to simulate realisations of h_k. It is not possible to simulate realisations of the other water balance terms in EMERALD as implemented in VIDENTE. Using a different stochastic model than described in Equations (1)-(3), it is possible to simulate such realisations with the stand alone version of EMERALD. We refer to Walvoort and Bierkens (1999) for this option.

In case fluctuation quantities have to be estimated, multiple realisations must be simulated. The output consists of a file containing the realisations of h_k. This file can be used as input for the program STATSIM in order to calculate the fluctuation properties. In our example we would typically simulate realisations of 30 years long, i.e. using precipitation and potential evapo-transpiration from 1970-1999, in order to obtain fluctuation quantities that are representative for the current climate, and not only for the weather conditions of the observation period 1995-1999 (see Knotters and Van Walsum, 1997).

2.4.4 Estimating fluctuation quantities

Using the program STATSIM fluctuation quantities can be calculated from the simulated realisations. The output consists of the following:

A file with statistics

An example of this output is shown hereafter in Figure 9. The mean highest water table, the mean lowest water table and the mean spring water table are given (expected values, median values and 5 and 95 percentiles). Here quantities are calculated per realisation. So, the uncertainty here reflects only model uncertainty. The univariate statistics are statistics over all simulated water table depths, which means that they reflect both the within year variation and the year to year variation of the weather as well as the model uncertainty (variation between realisations).

```
----------------------------------------
Univariate Statistics
----------------------------------------
mean   -106.425819
variance     360.064484
stdev.      18.975365
3rd moment    -539.639771
----------------------------------------
P01    -148.944107
P05    -138.432495
P10    -132.007401
P25    -119.718979
P50    -105.809341
P75     -92.923340
P90     -82.159462
P95     -75.837044
P99     -65.247566
----------------------------------------


-------------------------------------------------------------------------------------------
Mean Highest and Mean Lowest Water Table and Mean Spring Water Table
-------------------------------------------------------------------------------------------
mean             5%          50%             95%         std
-------------------------------------------------------------------------------------------
GLG(t)    -135.358978    -138.189285    -135.264694    -132.976761       1.565006
GHG(t)     -78.744606     -81.293327     -78.789520     -76.163002       1.535088
GLG(z)    -135.058350    -138.189285    -134.869598    -132.487930       1.608089
GHG(w)     -79.471909     -82.413284     -79.414589     -76.819191       1.603910
GVG(t)     -99.525368    -103.282478     -99.257187     -96.835274       2.082639
-------------------------------------------------------------------------------------------
```

Figure 9 File with fluctuation statistics as output of STATSIM; GHG: mean highest water table; GLG: mean lowest water table; GVG: mean spring water table; between brackets: denoting whether the statistic is determined for the whole year (t), the summer period (s) or the winter period (w).

A file with the frequency of exceeding graph (FOE-graph)

The FOE-graph gives for every level the number of days that the water table depth on any future day of any future year (at the same climate and hydrological regime as in the calibration period) exceeds that level. It is in fact the cumulative frequency distribution, reflecting both the within year variation and the year to year variation of the weather as well as the model uncertainty. Figure 10 shows a FOE-graph.

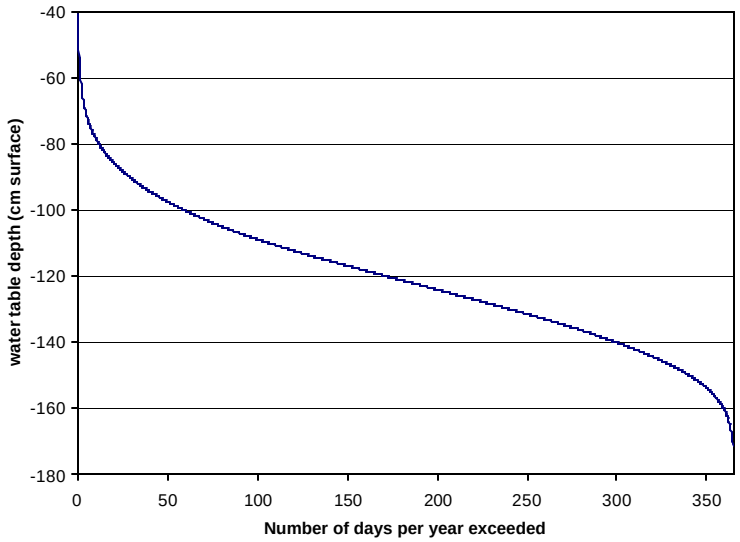


Figure 10 Frequency of exceeding graph

A file with the regime graph

The regime graph gives for every day number in a future year the expected water table depth, the median and the 5 and 95-percentiles. The variation per day number therefore reflects both year to year variation (i.e. our uncertainty about the future weather) as well as model uncertainty. Figure 11 gives an example of a regime graph.

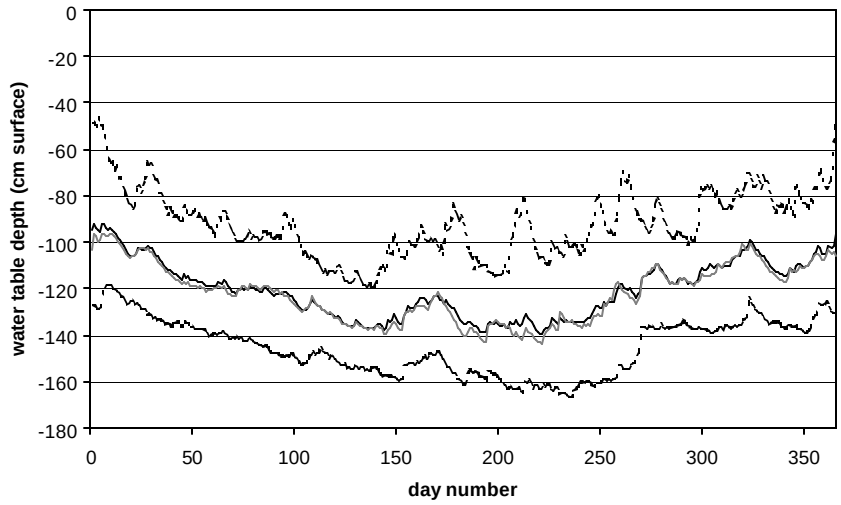


Figure 11 Regime graph; black solid line: mean; black dashed lines: 5- and 95-percentiles; grey solid line: median.

A file with the histogram

Like the FOE-graph, the histogram reflects both the within year variation and the year to year variation of the weather as well as the model uncertainty. From the histogram we can read the expected number of days that the water table depth will be within certain boundaries. Figure 12 gives an example histogram.

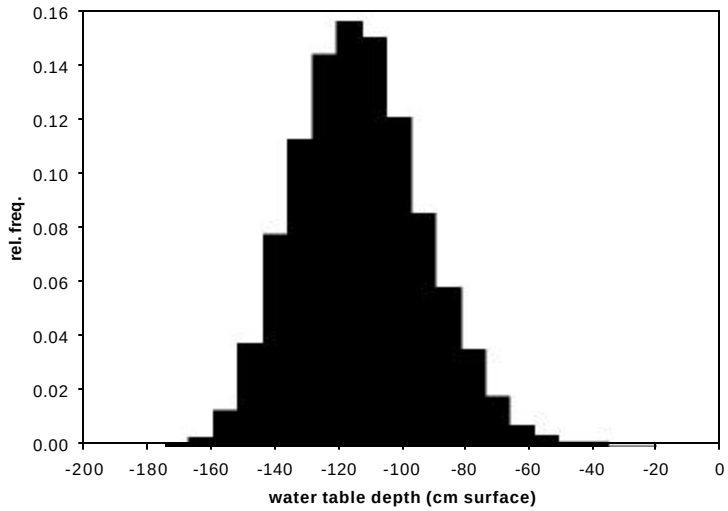


Figure 12 Example histogram

A file with the correlation function

The correlation function gives the correlation coefficient between day *k* and day *k* +1 (= lag 1), day *k* and day k+ 2 (lag 2), day *k* and day *k* + 3 etc. The correlation function reflects both the response time of the groundwater system, as well as the periodicity of the rainfall surplus. The longer it takes for the correlation function to cross the *x*-axis, the slower the response time of the groundwater system. However, because the periodicity of the rainfall surplus is also included, we cannot read the characteristic response time from the correlation function. The correlation function is the average of the correlation functions that are estimated for the realisations. An example correlation function is given below (Figure 13).

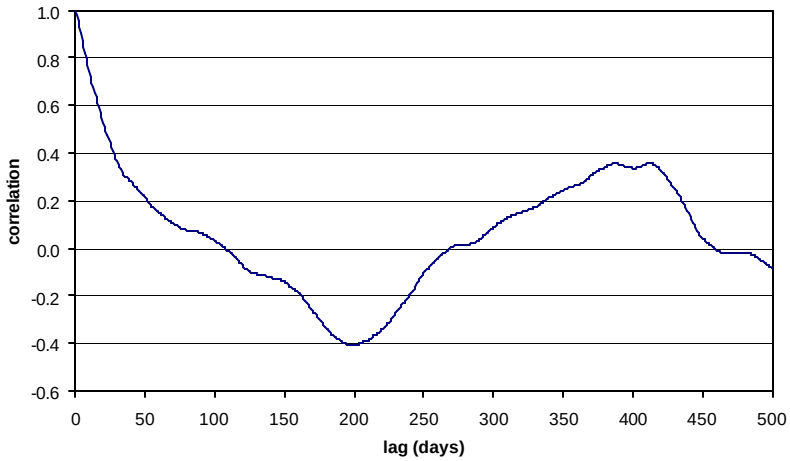


Figure 13 Example correlation function

The characteristic response time (in days) for the deterministic part of the EMERALD (Equation 2) is given by (Bierkens et al., 1999):

$$t_c = \frac{-3\Delta t}{\ln(a)} \tag{42}$$

As already given in Equation (38), the effective correlation length (time span over which n_k is correlated in days) of the noise process (3) is given by (Bierkens et al., 1999):

$$q_c = \frac{-3\Delta t}{\ln(f)} \tag{43}$$

Of course many more fluctuation quantities could be estimated from the simulated realisations; see for instance Bierkens (1998) and Knotters et al. (2000).

2.4.5 On line prediction and monitoring

Finally, using the calibrated model, EMERALD can be used for on line prediction, where in between dates that observations of water table depth are taken, optimal predictions of water table depth are obtained. EMERALD can also be used as a monitoring instrument by running it on line and checking whether not much more than 5% of the residuals fall outside the 95% error bounds (see Figure 7). If this is the case, a change in the hydrological system may have occurred. An alternative way of monitoring such a change is to plot the observations together with the prediction interval $\hat{h}_k \pm 1.96 s_e$ (with s_e calculated with Equation (32)). If the observations start to plot outside this interval, a significant change in the hydrological system may have occurred.

References

Bierkens, M.F.P., 1998. Modeling water table fluctuations by means of a stochastic differential equation. *Water Resources Research* 34(10), 2485-2499.

Bierkens, M.F.P., M. Knotters and F.C. Van Geer, 1999. Calibration of transfer-function noise models to sparsely or irregularly observed time series. *Water Resources Research* 35(6), 1741-1750.

Bierkens, M.F.P. en D.J.J. Walvoort, 1998. Eenvoudige stochastische modellen voor grondwaterstandsfluctuaties. Deel 2: Gecombineerd bodem-grondwatermodel met stochastische invoer (Simple stochastic models of water table fluctuations. Part 2: a combined soil-groundwater model with stochastic input). *Stromingen* 4(3), 5-20.

Chatfield, C., 1989. *The Analysis of Time Series: an Introduction. Fourth Edition*. Chapman and Hall, London.

De Zeeuw, J.W., 1966. *Analyse van het afvoerverloop van gebieden met hoofdzakelijk grondwaterafvoer. (Hydrograph analysis of areas with prevailing groundwater discharge).* Ph.D. Thesis Wageningen Agricultural University, Wageningen.

Knotters, M. and P.E.V. Van Walsum, 1997. Estimating fluctuation quantities from time series of water-table depth using models with a stochastic component. *Journal of Hydrology* 197, 25-46.

Knotters, M., M.F.P. Bierkens and C. Beets, 2000. *Grondwaterdynamiek van vegetatiestandplaatsen; analyse van zesentwintig reeksen (Groundwater dynamics of vegetation stands; analysis of 26 time series).* Rapport 095, Alterra, Wageningen.

Kraijenhoff van de Leur, D.A., 1958. *A study of non-steady groundwater flow with special reference to a reservoir coefficient*. De Ingenieur 70 (19), 87-94.

Maasland, M, 1959. *Water table flluctuations induced by intermittent recharge.* Journal of Geophysical Research 64(5), 549-559.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1986. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, New York.

Schweppe, F.C., 1973. *Uncertain Dynamic Systems*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Te Stroet, C.B.M., 1995. *Calibration of stochastic groundwater flow models: Estimation of system noise statistics and model parameters.* Ph.D. thesis, Delft University of Technology.

Van der Sluijs, P. and J.J. De Gruijter, 1985. Water table classes: A method to describe seasonal fluctuation and duration of water tables on Dutch soil maps. *Agricultural Water Management* 10, 109-235.

Van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal* 44, 892-898.

Walvoort, D.J.J. and M.F.P. Bierkens, 1999. *A stochastic modelling approach for rapid assessment of groundwater dynamics.* Report 171, DLO Staring Centrum, Wageningen.

Wesseling, J.W., 1991. *Meerjarige simulatie van grondwaterstroming voor verschillende bodemprofielen, grondwatertrappen en gewassen met het model SWATRE (Long year simulations of groundwater flow for various soil types, water table classes and crops, using the model SWATRE).* Rapport 152, DLO-Staring Centrum, Wageningen.

Wösten, J.H.M., G.J. Veerman and J. Stolte, 1994. *Waterretentie – en doorlatendheidskarakteristieken van boven- en ondergronden in Nederland: de Staringreeks (characteristics of soil water retention and unsaturated hydraulic conductivity of topsoils and subsoils of the Netherlands: the Staring Series).* Technisch Document 18, DLO-Staring Centrum, Wageningen.

Zwamborn, M.H., 1995. *Modellering van de onverzadigde zone ten behoeve van grondwatermodellen (Modelling the vadose zone for groundwater models).* Rapport SWI 95.142, Kiwa, Nieuwegein.

Annex: Input instructions for EMERALD

Following is the user's manual of EMERALD of Walvoort and Bierkens (1999). The stand alone version of EMERALD has some extra features that have not been included in VIDENTE, mainly because they are rarely used. First, in VIDENTE the ratio of the distance to the water course x and the distance between the water courses L is always taken as ½. The stand alone version allows for other ratios. Furthermore, the stand alone version allows for an additional noise model, called "internal noise", as opposed to the external noise model of Equations (1)-(3). In this internal noise model, the coloured noise process (Equation 3) is added to the net input to the groundwater system resulting in the following equations:

$$h_k = h_s + a(h_{k-1} - h_s) + b_1 q_n(k) + b_2 q_n(k-1) + b_1 n_k + b_2 n_{k-1} \quad \text{(A1)}$$

$$n_k = f\, n_{k-1} + e_k \quad \text{(A2)}$$

As can be seen, the properties of the error process $h_k - \hat{h}_k$ (i.e. its variance and correlation in time) are not only dependent on the noise parameter f but also on the (physical) parameters a, b1 and b2. EMERALD provides two different ways of calibrating the noise process. Because the noise is now an intricate part of the system, i.e. it is seen as the error in the net input qn(k), stochastic simulation is possible for all other water balance terms that depend on hk, and not only for the water table depth as in VIDENTE. It is not implemented in VIDENTE, because calibration of the internal noise case is rather cumbersome and therefore not much used.

User's manual emerald (version: July 1998)

Introduction

EMERALD is implemented on a personal computer. It is written in Borland's Turbo Pascal (version 7.0) for ms-dos (assumed screen resolution 600×800). A data flow diagram (DFD) of emerald is given in Figure A1. In this diagram, processes are denoted by circles, temporary files by two parallel horizontal lines, user accessible files by boxes, and data flows by arrows.

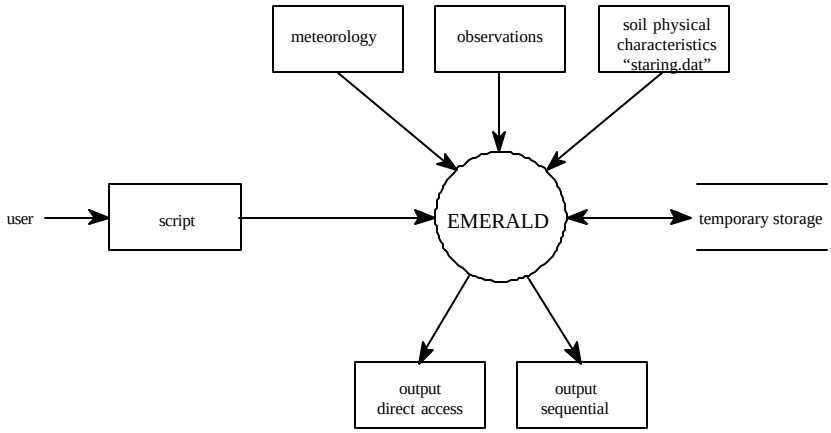


Figure A1 Data flow diagram of EMERALD

EMERALD is driven by a script file, which takes a central position in the DFD. A script file contains a high-level computer language, i.e. the script, which prescribes which actions to perform, which files to read, and where to store results. At run-time, emerald reads the script, deciphers it by means of its built-in script interpreter, and carries out the appointed tasks. Each line in the script file consists of a script directive and associated parameters. In general, the syntax reads:

DIRECTIVE <mandatory_parameter> [optional_parameter] <option_1 | option_2>

Throughout this manual, mandatory parameters are given in angular brackets and optional parameters in square brackets. Furthermore, piping symbols, i.e. |, are used to separate two or more options. Only one of these options should be selected. Annotations should be preceded by a "+" on the first position of each commentary line in the script file. Parameters printed in italics and Greek symbols should be replaced by appropriate numeric values, parameters printed upright should simply be copied. emerald's script language is order invariant. This means that the user can put the script directives in any order (s)he prefers. However, if emerald encounters a directive more than once, only the first is processed.

Script directives

In this section the script directives are addressed in alphabetical order.

BINARY	*output*
syntax:	binary <* | [Ea] [Ep] [V] [qp] [qg] [qc] [qn] [h] [qd]>
purpose:	Stores output in direct access files. In case of simulation, storage of all realisations may take up a substantial amount of disk space. In order to suppress this demand, emerald offers the opportunity to store the realisations in direct access format. This can be effectuated by means of the binary-directive. Its syntax is very similar to the output-directive, except that no output file name(s) can be specified. emerald uses default file names which consist of the parameter name of interest plus extension *rl4*, *e.g. Ea.rl4*, *Ep.rl4*, and *V.rl4*. A description of these binary files is given in a subsequent section.
example:	binary h qd

BYPASS *model parameters*

syntax: BYPASS <$f_{b,1}$> [$f_{b,2}$] ... [$f_{b,9}$]

purpose: Governs the amount of bypass flow through subsoil layers 1 to 9. $f_{b,i}$ should be expressed as a fraction of the incoming flux of subsoil layer *j*.

example: BYPASS 0.5 0.2 0.1

CALIBRATION *action*

syntax: CALIBRATION < [D] [II] [IF] [EF] > [tolerance]

purpose: Specifies the calibration method(s) to perform. The deterministic component is calibrated when D is encountered on the parameter line, the stochastic component is calibrated when II, IF, and/or EF are encountered. These abbreviations stand for Internal noise – Inverse model, Internal noise - Forward model, and External noise - Forward model respectively. Optionally, the tolerance of the termination criterion of the Downhill-Simplex method may be specified (Press *et al.*, 1989).
When II, IF, and/or EF are supplied as parameters, EMERALD expects the observations in the observation file to be separated by approximately equal time steps (section 3.2). Furthermore, if II is specified, EMERALD replaces missing values by predictions obtained by linear interpolation.

example: CALIBRATION D IF EF 1e-6

CROP *model parameters*

syntax: crop <f_c> <$h_{p,fc}$> <$h_{p,lp}$> <$h_{p,wp}$> <* | z_c>

purpose: Supplies emerald with crop and soil specific parameters. The parameter line should contain the following quantities:
f_c = crop factor [-];
$h_{p,fc}$ = pressure head at field capacity (cm);
$h_{p,lp}$ = pressure head at limiting point (cm);
$h_{p,wp}$ = pressure head at wilting point (cm);
z_c = critical depth (cm);
* = critical depth according to Wösten *et al.*, (1994) (cm).
EMERALD makes no distinction between positive and negative pressure heads.

example: CROP 1 -100 -500 -8000 *

DRAINAGE *model parameters*

syntax: DRAINAGE <z_d> <x> <L>

purpose: Specifies the drainage parameters:
z_d = level of drainage base with respect to soil surface (cm);
x = lateral space co-ordinate (see Figure 2.1) [L];
L = distance between drainage courses (see Figure 2.1) [L].

See to it that x and L have corresponding units.

example: DRAINAGE -100 50 100

FIX *action*

syntax: FIX <[gamma] [mu] [qv] [phi] [sigma2]>

purpose: Keeps the parameters on the parameter line fixed to their initial values during calibration. The initial values should be specified by the GROUNDWATER and NOISE directives.

example: FIX gamma phi

GROUNDWATER *model parameters*

syntax: GROUNDWATER <μ> <?> <q_v> <* | h_{avg}> <$q_{g,avg}$>

purpose: Provides the parameters of the saturated zone. The parameter line should contain the following quantities:

?	= specific yield [-];
G	= drainage resistance (days);
q_v	= infiltration/seepage flux (mm/d);
h_{avg}	= average groundwater level (cm);
*	= average groundwater level is based on observation file (cm);
$q_{g,avg}$	= average flux in percolation zone (mm/y).

Groundwater levels should be given with respect to the soil surface.

example: GROUNDWATER 0.2 200 0 * 250

INITIAL *model parameters*

syntax: INITIAL <S_0> <* | h_0> <q_{d0}> <q_{n0}> <N_0>

purpose: Gives the initial values of the following parameters:

S_0	= initial saturation grade root zone [-];
h_0	= initial groundwater level with respect to the soil surface (cm);
*	= h_0 equals the average groundwater level in the observation file (cm);
q_{d0}	= initial specific discharge (mm/d);
q_{n0}	= initial net input to the groundwater system (mm/d);
N_0	= initial value of noise process N.

The following default values are used if this directive is omitted: S_0: field capacity, h_0: drainage base, q_{d0}, q_{n0}, and N_0 are set to zero.

example: INITIAL 0.8 * 0 0 0

LAGS *action*

syntax: LAGS <n_{fit}> [n_{print}]

purpose: Specifies the number of lags involved in fitting the autocovariance function (section 3.2). The optional parameter n_{print} specifies the number of lags that should be written to the output file. n_{print} should always be greater than or equal to n_{fit}. If the LAGS - directive is omitted, n_{fit} and n_{print} are set to 10 by default, *i.e.* lags 0 to 9 are used for fitting.

example: LAGS 15 20

METEO	input
syntax:	METEO <file_name> [skip]
purpose:	Specifies the name of the file containing time series of precipitation amounts and evapotranspiration. Its file format is given in a subsequent section. The optional parameter [skip] denotes the number of lines to skip in the meteo file (default: skip=0).
example:	METEO Eelde.met 1

NOISE	*model parameters*
syntax:	NOISE < f > < s_e^2 > NOISE < f > < s_e^2 > <INTERNAL | EXTERNAL>
purpose:	Specifies the parameters of the noise process. The first is sufficient to supply the initial values in case of calibration, the latter is required in case of simulation. In case of prediction, the variance of the prediction errors of h and/or q_d are given if the noise type is set to INTERNAL.
example:	NOISE 0.25 0.01 INTERNAL

observations	*input*
syntax:	observations <*day1*> <*year1*> <*day2*> <*year2*> <*file_name*> [*skip*]
purpose:	Specifies which part of observation file <*file_name*> should be processed. The period of interest starts at <*day1*> of <*year1*> and ends at <*day2*> of <*year2*>. [*skip*] refers to the number of lines to skip in <*file_name*>, and is 0 by default.
example:	observations 1 1982 365 1991 12BL0015.dat 1

OUTPUT	*output*
syntax:	OUTPUT <*file_name*> OUTPUT <*file_name*> <* | [P] [Ea] [Er] [Ep] [V] [qp] [qg] [qc] [qn] [h] [qd]>
purpose:	Creates output file <*file_name*> which gives a summary of input parameters, and a detailed description of model output. In case of prediction or simulation, the requested output should be enumerated after <*file_name*>. The mnemonics on the parameter line correspond to those used in the report. If the symbol * is used, all parameters on the parameter line are written to <*file_name*>. In case of calibration, it is sufficient to supply <*file_name*>.
example:	OUTPUT Emerald.out h qd

PREDICTION	*action*
syntax:	PREDICTION <*day1*> <*year1*> <*day2*> <*year2*>
Purpose:	Performs prediction. The parameter line specifies the start and end of the prediction period, *i.e.* <*day1*> of <*year1*> and <*day2*> of <*year2*> respectively. If the script contains the OBSERVATION-directive, validation and/or verification measures are computed for the time span that is part of both the prediction period and the observation period.
Example:	PREDICTION 1 1982 365 1991

PRERUN	*action*
syntax:	PRERUN <* | *#years*>
Purpose:	Specifies the length of the prerun or warming up period. A prerun of <*#years*> years is required to eliminate the effect of the initial values. The prerun period

starts at the first day in the meteo file if <*> is encountered on the parameter line.

Example: PRERUN 2.5

SIMULATION	*action*

syntax: SIMULATION <*day1*> <*year1*> <*day2*> <*year2*> <*#runs*> <*#runs_out*> <*seed*>

Purpose: Performs simulation. The parameter line specifies the start and end of the simulation period, *i.e.* <*day1*> of <*year1*> and <*day2*> of <*year2*> respectively, the number of realisations to perform <*#runs*>, the number of realizations to write to the output file <*#runs_out*>, and the random seed <*seed*> to initiate the pseudo random number generator (*gasdev/ran1* of Press *et al.* (1989)).

Example: SIMULATION 1 1982 365 1991 1000 1 12534

SOIL	*model parameters*

syntax: SOIL <*topsoil=thickness*> <*subsoil1=thickness*> ... [*subsoil9=thickness*] [*cut_off*]

Purpose: Specifies the soil physical characteristics of the unsaturated zone. The parameter line contains building blocks of the *Staringreeks* (Wösten *et al.*, 1994), together with their associated thicknesses (cm). A total of nine subsoil layers may be specified. The root zone is represented by the topsoil layer, the percolation zone by the subsoil layers. The amount of percolation not reaching the groundwater system due to truncation of the pulse response function $U_p(? t,t)$ is governed by [*cut_off*]. This quantity should be expressed as a fraction of the total amount of percolation. It significantly affects the amount of CPU-time required. Default value: 1E-6.

Example: SOIL B3=30 B3=20 O3=100 1e-5

TMPDIR	*general*

syntax: TMPDIR <*path*>

Purpose: Designates the path to temporary files. Model performance may be significantly improved if <*path*> denotes a RAM-drive.

Example: TMPDIR e:\tmp

Examples of script files for EMERALD

In Figures A2 and A3 examples of script files are given.

```
+ + + + + + + + + + + + + + + + + + + + + + + + + + + + + +
+                                                         +
+  Calibration of the deterministic and stochastic        +
+  components (internal noise, forward model),            +
+  followed by prediction (verification)                  +
+                                                         +
+  period   : 1982-1986                                   +
+  well     : 12BL0015                                    +
+  location : Vries, Eelde                                +
+                                                         +
+ + + + + + + + + + + + + + + + + + + + + + + + + + + + + +

CALIBRATION D IF 1e-9
LAGS 20 20
PREDICTION 1 1982 365 1986
OBSERVATIONS 1 1982 365 1986 12BL0015.dat 5
METEO Eelde.met 1
PRERUN 2
OUTPUT prediction.out h
CROP 1 100 500 8000 *
SOIL B3=35 B3=20 O3=30 O4=95 1e-6
GROUNDWATER 0.2 165 0 * 261
DRAINAGE -80 1 2
INITIAL 1 * 0 0 0
BYPASS 1 1 1
NOISE 0 0.001 INTERNAL
```

Figure A2 Script to calibrate EMERALD and verify the deterministic model part with prediction

```
+ + + + + + + + + + + + + + + + + + + + + + + + + + + + + +
+                                                         +
+ Calibration of the deterministic and stochastic         +
+ components (internal noise, forward model),            +
+ followed by simulation.                                +
+                                                         +
+ period   : 1982-1991                                   +
+ well     : 12BL0015                                    +
+ location : Vries, Eelde                                +
+                                                         +
+ + + + + + + + + + + + + + + + + + + + + + + + + + + + + +

CALIBRATION D IF 1e-9
LAGS 20 20
SIMULATION   1 1982 365 1991 1000 2 12534
OBSERVATIONS 1 1982 365 1991 12BL0015.dat 5
METEO Eelde.met 1
PRERUN 2
OUTPUT Simulation.out h qd
BINARY h qd
CROP 1 100 500 8000 *
SOIL B3=35 B3=20 O3=30 O4=95 1e-6
GROUNDWATER 0.2 165 0 * 261
DRAINAGE -80 1 2
INITIAL 1 * 0 0 0
BYPASS 1 1 1
NOISE 0 0.001 INTERNAL
```

Figure A3 Script to calibrate EMERALD followed by stochastic simulation

Execution
The executable of emerald should be run in a ms-dos environment. If your operating system is windows 3.x/95/98/nt, emerald should be executed in a ms-dos-box. Execution starts after typing

EMERALD <script_file>

on the command-line, followed by pressing the enter/return-key. The name of the script file should satisfy the ms-dos conventions, even when emerald is installed on a windows-machine. File names may contain wildcarts in order to start several script files on a row. For instance

EMERALD *.scr

processes all script files with extension scr in the active directory.

Input files
The data flow diagram of emerald discerns three (ascii) input files. The first contains meteorological data, the second observed groundwater levels, and the third soil physical characteristics.

The meteo file consists of four columns, i.e. day number (1-365 or 366), year (4 digits), amount of precipitation (mm), and amount of evapotranspiration according to Makkink (mm). An example of a meteo file is given in Figure A4. Parameter skip equals 1, because one header line is present. Missing values are not allowed. Furthermore, the time step between successive lines should equal 1 day.

```
  day  year     P    ET
    1  1959   6.5   0.4
    2  1959   9.0   0.3
    3  1959   5.4   0.3
    4  1959   7.7   0.3
    5  1959   7.7   0.3
    6  1959   0.5   0.2
    7  1959   2.9   0.1
    8  1959  12.4   0.3
    9  1959   2.7   0.2
   10  1959   8.3   0.1
```

Figure A4 Example of a meteo file

The observation file consists of three columns, i.e. day number (1-365 or 366), year (4 digits), and observed groundwater levels (cm). The entries of the third column are defined with respect to the soil surface, and are decreasing in downward direction. All alphanumeric characters (including blanks) are regarded as missing values. An example of an observation file is given in Figure A5. It contains one header line, so parameter skip of the observation-directive should be set to 1.

```
  day    year       Y
   14    1982     -36
   28    1982     -34
   43    1982     -39
   57    1982     -33
   74    1982     -39
   88    1982
  104    1982     -64
  118    1982     -71
  134    1982     -74
  148    1982     -90
```

Figure A5 Example of an observation file

The file containing soil physical properties is called "Staring.dat" (Figure A6). This file contains the Van Genuchten parameters, the critical depth, and the thickness of the capillary fringe for all soil building blocks of the "Staringreeks" (Wösten et al., 1994). The user is allowed to add new building blocks, and edit existing ones. However, see to it that the codes referring to the soil building blocks consist of three characters at most.

```
STARINGREEKS, Wösten et al., 1994 (Technical Document 18)


topsoils
code    Tr     Ts     Ks       a       l      n     zc    dae
B1    0.01   0.43  17.46  0.0249  -0.140  1.507   94.0   19.2
B2    0.02   0.43   9.65  0.0227  -0.983  1.548  104.0   18.8
B3    0.01   0.45  17.81  0.0152  -0.213  1.412  151.0   19.3
B4    0.01   0.42  54.80  0.0163   0.177  1.559  201.3   19.8
B7    0.00   0.40  14.07  0.0194  -0.802  1.250   93.2   18.0
B8    0.00   0.43   2.25  0.0096  -2.733  1.284  102.3   15.1
B9    0.00   0.43   1.54  0.0065  -2.161  1.325  119.1   15.0
B10   0.01   0.42   1.17  0.0118  -4.795  1.224   58.2   10.8
B11   0.00   0.60   5.26  0.0243  -5.395  1.111   29.0   10.5
B12   0.00   0.55  15.46  0.0532  -8.823  1.081   24.5   10.2
B14   0.01   0.42   0.80  0.0051   0.000  1.305   65.0   12.3
B16   0.00   0.73  13.44  0.0134   0.534  1.320  120.4   18.8
B17   0.00   0.72   4.46  0.0180  -0.350  1.140   30.0   11.9
B18   0.00   0.77   6.67  0.0197  -1.845  1.154   43.9   14.0


subsoils
code    Tr     Ts     Ks       a       l      n     zc    dae
O1    0.01   0.36  13.21  0.0224   0.000  2.167   87.0   19.5
O2    0.02   0.38  15.56  0.0214   0.039  2.075   96.0   19.6
O3    0.01   0.34  18.30  0.0211  -0.522  1.564  127.0   19.4
O4    0.00   0.36  53.10  0.0216  -0.520  1.540  189.0   19.8
O5    0.01   0.32  43.55  0.0597   0.343  2.059   42.0   19.4
O6    0.00   0.41   5.48  0.0291  -6.864  1.152   61.6   12.4
O8    0.00   0.47   9.08  0.0136  -0.803  1.342  128.3   18.5
O9    0.00   0.46   2.23  0.0094  -1.382  1.400  107.0   16.3
O10   0.00   0.49   2.22  0.0107  -2.123  1.280   79.1   14.7
O11   0.00   0.42  13.79  0.0191  -1.384  1.152   66.5   16.3
O12   0.00   0.56   1.14  0.0095  -4.171  1.159   33.1    8.9
O13   0.00   0.57   3.32  0.0171  -4.645  1.110   25.1    9.6
O14   0.00   0.38   0.36  0.0025   0.057  1.686  140.0   12.2
O15   0.01   0.41   3.70  0.0071   0.912  1.298  109.2   17.1
O16   0.00   0.89   1.07  0.0103  -1.411  1.376   58.1   13.2
O17   0.00   0.86   2.75  0.0127  -1.832  1.274   71.0   15.0

user defined soils
code    Tr     Ts     Ks       a       l      n     zc    dae
D1    0.01   0.43  17.46  0.0249  -0.140  1.507   94.0   19.2
W1    0.01   0.36  13.21  0.0224   0.000  2.167   87.0   19.5
```

Figure A6 Example of "Staring.dat"

Output files

EMERALD always generates an ascii output file. It contains a summary of input data, and the results of executed calibration, prediction and/or simulation routines. In case of simulation, the results can also be stored in binary (direct-access) format by using the binary-directive. The contents of each record is given in table A1.

Table A1 Contents of binary (direct-access) files. Each record contains 4 bytes of information.

record	contents
0	first year of the simulation period
1	final year of the simulation period
2	number of runs
3-eof	realizations, where the year-loop is embedded in the run-loop

Disclaimer

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY. No author or distributor accepts responsibility to anyone for the consequences of using it or for whether it serves any particular purpose or work at all, unless he says so in writing. Everyone is granted permission to copy, modify and redistribute this program, but only under the condition that proper reference is given and that this notice and the above copyright notice remain intact.

Part 5: Input instructions for STATSIM

The program STATSIM can be used as a postprocessor after simulations with KALMAX, KALTFN, SSD and EMERALD. In case of EMERALD, the binary file made by the Pascal program EMERALD has to be converted to an unformatted Fortran file. To achieve this, two auxiliary programs are necessary: the program PAS2ASC to get from EMERALD binary output to an ascii file, and the program ASC2FOR to get from this ascii file to Fortran unformatted input for STATSIM. Both programs are also present in the STATSIM directory on the CD.

Header of the program STATSIM with input instructions

```
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
C Copyright (C) 2000, Alterra, Green World Research
C
C This program is distributed in the hope that it will be useful,
C but WITHOUT ANY WARRANTY.  No author or distributor accepts
C responsibility to anyone for the consequences of using it or for
C whether it serves any particular purpose or work at all, unless he
C says so in writing. Everyone is granted permission to copy, modify
C and redistribute this program, but only under the condition that proper
C reference is given and that this notice and the above
C copyright notice remain intact.
C
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c Program: STATSIM - calculation of STATistics from simulated data
c Goal: From NSIM realisations of simulated data with:
c          KALMAX
c          KALTFN
c          SSDSIM
c          EMERALD (after converting emerald output via PAS2ASC en
c                   ASC2FOR to a file readable by STATSIM)
c Note: Only whole years can be simulated with these programs if STATSIM
c       is to be used.
c
c The program estimates: 1) univariate statistics, including MHW/MLW
c       2) histogram; 3) frequency of exceedence, 4) regime graph,
c       5) auto-correlation function
c
c Author: M.F.P. Bierkens
c Date: April 1998; last modified August 2000
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c
c**********************************************************************
c input: - file (unformatted) with simulated data starting at januari 1
c          record 1: year1, nsim, nyears
c                    - year1:  first year of simulation
c                    - nsim:   number of realisations
c                    - nyears: number of years in simulation
c          do k = 1,nsim
c             do i = 1,nsteps
c                record 1 + (k-1)*nsim + i: gw1 – simulated water table
c                                               depth (or in case of
c                                               SSD soil saturation or
c                                               drainage discharge)
```

```
c             enddo
c          enddo
c note: nsteps = int(simtim/dstep)
c
c output:
c          - file with statistics and ghg/glg
c          - file with histogram data (can be used to estimate histogram)
c          - file with foe-curve
c          - file with regimecurve
c          - file with correlation function
c***********************************************************************
```