Groundwater model calibration for the Amsterdam Water Supply dune area

THEO N. OLISTHOORN & PIERRE T. W. J. KAMPS
Amsterdam Water Supply, Vogelenzangseweg 21, 2114 BA Vogelenzang, The Netherlands

Abstract The groundwater model of the 36-km² dune area of the Amsterdam Water Supply has been calibrated in steady-state mode using MODINV. The calibration of the four parameters took two months work and 67 optimization runs. The optimizing code facilitated the elimination of many model errors by allowing us to focus on the hydrological causes of spatial error correlation, which has to be removed before optimized parameters can be accepted for further use. Throughout the process, the model has been greatly improved, finally yielding an ever low root mean square error of 0.17 m in the two topmost aquifers, and 0.15 m in the third aquifer. This result would not have been possible without the optimization.

INTRODUCTION

The Amsterdam Water Supply uses MODFLOW to model the groundwater flow in its dune area along the Dutch North-Sea coast (Olsthoorn et al., 1993). A first four-layer version of the model was finished in 1988 and calibrated by trial-and-error (Emke & Kooiman, 1988). The model was calibrated by comparing isohyps maps and the flow in extraction canals. Apart from the fact that the calibration results have not been quantified in terms of parameter reliability, today nobody remembers what steps have been consecutively made in the process.

In 1992, a refined version of the model was built, to be used in eco-hydrological scenario studies for the southern part of the dune area. That model was calibrated by us using a Monte Carlo approach, optimizing four parameters with the steady-state version of the model (Olsthoorn, 1995). Reasonable results were obtained, but the method lacks accuracy and cannot be extended to greater number of parameters.

In 1995, an even more refined model became ready, now containing seven aquifers. The extension to seven aquifers is due to the geological complexity of the northern part of the dune area, taken into study only recently. For the calibration of this new model it was decided to use a commercially available computer code. A selection was made from MODFLOWP, PEST and MODINV. In reality these codes perform the same tasks, apply more or less the same technology and provide the same results. We chose MODINV (Anonymous, 1995) because of its seamless connection with MODFLOW and the clarity of its manual.
DATA

We have tried to take a systematic approach. To this end all our more than 700 available
drillings have been reinterpreted and coded by a geologist and were stored in our GIS.
Maps were generated of all geological formations between ground surface and the
presumed bottom of the groundwater system at about 150 m below mean sea level.
Furthermore, all hydrologically relevant data (wells, recharge ponds, canals,
observation wells, interface between fresh and saline groundwater, our detailed
vegetation map and a 7-million-point remotely sensed digital elevation model) has been
stored in the same GIS. The GIS is connected to an ORACLE database containing
measured data such as flows, groundwater heads, canal levels and raingauge data. Using
ARC/INFO's AML language, the input for our MODFLOW model is generated
automatically, thus guaranteeing that the most recent data is always used in the
groundwater model.

"Soft" model parameters such as hydraulic conductivities are kept outside GIS. This
way, by multiplying the given layer thickness by an adaptable hydraulic conductivity,
the final model input is obtained, needed to run MODFLOW. Also, this way the model
can be optimized outside the GIS using a calibration code such as MODINV. The GIS
is indispensable afterwards to interpret calibration results in more detail and study and
correct model failures.

All calibrations described in this paper were performed with a steady-state model
using averaged data between 1974 and 1990, a period in which the hydrological situation
in the area has been almost constant. Transient calibration is difficult due to the fact that,
though we have about 1200 observation points, they are only measured four times per
year. Due to the very long response time of the dune system (about 7 years), short time
series are of little value to calibrate for anything else than the specific yield. Hence,
transient calibration was used to calibrate specific yield by fitting its value such that the
variance of the phreatic head measurements calculated by the model equalled that of the
quarterly measurements over 1974-1990.

PARAMETERS AND GROUNDWATER HEAD PROCESSING

Calibration parameters

The location of the model area and a schematic cross-section are depicted on Fig. 1.
Since the major scope of the model is its use for ecological studies, the emphasis of the
calibration had to be on the phreatic aquifer, located between 5 m above the mean sea
level and 20 m below mean sea level. However, in part of the model area, a low-
permeability peat layer of about 0.5 m thickness, located at the mean sea level elevation,
splits the aquifer in two layers. Hence focus had to be on the two top aquifers. Clearly
their groundwater heads are largely influenced by the resistance of the clay layer at the
bottom of second aquifer. Therefore, this clay layer has to be taken into account. From
this it follows that the hydraulic conductivities of the upper two aquifers and the vertical
hydraulic conductivities of the peat and the clay layers had to be taken into account
simultaneously in the first calibration stage.
The kriged groundwater heads of the third aquifer could be taken as a fixed boundary. Since the clay layer heavily smooths the head in the below aquifer, and as enough observation wells are available to fix the head-map, only negligibly small errors can originate from this assumption.

In a second stage, the hydraulic conductivity and the boundary conditions of the third aquifer are calibrated, and finally the model as a whole.

In order to regularize the parameters as much as possible, the four parameters, they are all hydraulic conductivities, were taken as spatially constant values, that yield spatially varying transmissivities or vertical conductances after multiplication with the spatially varying layer thickness or its inverse. Due to the large number of observed heads, this small number of parameters can be easily optimized.
To enable the use of MODINV, the spatially-variable parameters have to be divided into a maximum of 10 zones each (a peculiarity of MODINV, where MODFLOWP takes a more general approach). Hence, the spatially varying transmissivity and conductance parameters were divided into 10 classes. These parameter classes are optimized with fixed mutual proportion. This way sets of classes are made to behave as a single parameter, as we intended.

To further regularize the parameters, the log-values of hydraulic conductivities were optimized instead of the real values.

Declustering of data points

Clusters of groundwater head values in the same aquifer have been attenuated by weights, using as weights the inverse of the number of observation wells within a radius of 200 m, including the well itself. This way, the weights of the observations within a 200 m circle around each observation well add up to unity. Though obvious, no explicit study to relate the radius to the semi-variogram of the groundwater head data has yet been performed.

Presentation of the errors

As we are convinced that it is useless to make contour maps of calibration errors, which should be random, we went for an adapted presentation: The calibration results are plotted as coloured dots on the groundwater head map of each aquifer, the size of which correspond to the magnitude of the error, and the colour to its sign. GIS is used effectively for this purpose, facilitating enormously the search for explanation and reduction of model errors during the calibration process.

THE CALIBRATION PROCESS

Clearly, the optimized parameters are useless as long as the model is incorrect. We might at best use them in the negative sense, to conclude that the model is incorrect, if their values are far beyond reasonable boundaries.

In order to conclude that the model is (sufficiently) correct we have to prove that the errors comply with the assumptions of calibration theory. This demands that the errors obtained by the optimization (of our steady-state model) be spatially uncorrelated. Therefore most of the calibration work resides with the hydrologist, attempting to obtain uncorrelated errors. No mathematical model can solve this for him/her. One has to use one’s own judgement in searching for reasons (well-founded in hydrological reality) for the observed spatial correlation of the errors. We can now say that 66 of our 67 MODINV runs were needed for this matter and only the last run, run 67, was used to obtain the final parameters.
MODEL IMPROVEMENTS

Figure 2 shows the groundwater heads and the errors of the phreatic aquifer after the first optimization run. Clearly, the errors are far from random. The following improvements were then made during the 65 next optimizations in a two-month period:

Correction of obvious errors in the database, such as:
- incorrect canal bottom elevations (e.g., in reality, canal bottoms were penetrating the second aquifer, but some canal bottom elevations were above the top of this second aquifer);
- observation wells that had been attributed to the wrong aquifer;
- observation wells with wrong xy-coordinates.

Then the following improvements were implemented:
- A clay layer zonation due to Stuyfzand (1987) provided better results, probably because it included the geological origin of the clay.
- Evapotranspiration was wrong in the northeast part of the model, where a vegetation map was lacking. A new vegetation map for that area solved the problem.
- A more precise location of the rims of the peat-layer, using our georadar data (Van Overmeeren, 1993), with its very low vertical permeability, gave large improvements for some observation wells in the vicinity of that boundary where substantial groundwater head gradients are present.
- Some observation wells in high-gradient areas needed improvement of their exact location (had been estimated in the field by putting a dot on the local map during construction). Some errors are unavoidable due to the size of the cells of the MODFLOW model.
- Due to the bath-tub-shaped surface of the peat layer, grown in elongated dune depressions that have been overblown by an average of 5 m of sand, the hydraulic conductivity of the phreatic aquifer near the rim of the peat has been reduced in part of the area, thus yielding a fifth calibration parameter.
- The clay layer around the northern end of the southeasterly extraction canal has been made thinner, because it appeared from the drillings that it contains much sand, contrary to the general composition of that layer. This is an example where the optimization pointed out that more information should be extracted from existing drilling information than had actually been done in the geological coding process.

After no further hydrological evidence for improvement of the model could be found (without performing further drillings, pumping tests, etc.), the third aquifer, between 20 and 65 m below mean sea level, was calibrated separately, using as input the previously calibrated conductance of the overlying aquitard, and the groundwater head above this aquitard. This has been done by trial-and-error, by multiplying the conductance of the general groundwater head boundaries at the western coast and the eastern inland side of the aquifer. This lifts the average head in that aquifer and its west-eastern inclination, implying two degrees of freedom.

After this, the optimizer has been run for the 67th time for the three aquifers combined. The final result for the phreatic aquifer has a standard deviation of 17 cm and is shown in Fig. 3. Clearly, most errors have disappeared and, more importantly, most of the spatial structure in the error has been removed. The remaining errors have become more randomly distributed, though in some part of the area some error structure
persists. However, since we currently have no evidence as to what is their cause, and
so no ground for adding additional parameters that could remove it, we prefer to leave
it, and use the red and yellow dots in the figures as a remark to the model user,
indicating that results in that part of the area are less certain.

The results for the third aquifer after run 67 are shown in Fig. 4. Its root mean
square error is as low as 15 cm, and hence no further calibration was performed.

RESULTS

In a calibration by Monte Carlo simulation (Olsthoorn, 1995) for the southern half of the
dune area, we obtained a root mean square error of 24 cm. Even a direct optimization
using Marquardt-Levenberg (Olsthoorn, 1995) gave a root mean square error of the
southern part of the area of 23 cm. The success of the calibration using MODINV is due
to the time and effort dedicated to improve the model, in other words the work of the
hydrologist. We concluded from this experience that a good optimizer such as
MODINV, MODFLOWP or PEST is an invaluable tool for hydrologists.

We did not calibrate explicitly on groundwater flows. So the specified and fixed
precipitation surplus (groundwater recharge rate) is the major driving force of the
model. We did however validate the average groundwater flow from the aquifer system
into the two major extraction canals, for which independent measurements were
available (Kamps & Mosch, 1995). This resulted in an error of 3% for the northern and
7% for the southern canal, which is thus to be considered as an independent verification
with quite an acceptable result.

With acceptable model results, the optimized parameters can be considered to
represent physical reality instead of artifacts of model errors. Hence one can investigate
in more detail the remaining information produced by MODINV, such as the parameter
errors and eigenvalues. These errors should finally be used in scenario studies for which
most models have been built in the first place.

CONCLUSION

We expected MODINV to calibrate the model for us, but it turned out differently. The
calibration became one long search for model errors. The calibration of the four
parameters took us two man-months of intensive work and 67 MODINV optimization
runs. Finally, we ended up with a greatly improved model with an ever low root mean
square error of 0.17 m in the two topmost aquifers and 0.15 m in the third aquifer. This
achievement would have been impossible without a proper calibration code. In fact, the
MODINV-optimizations were so fast (it took only 10 minutes on our work station (DEC
ALPHA)), that we were able and obliged to focus all our attention on interpretation of
the calibration results, researching for hydrological and geological causes of the errors,
followed by hydrologically founded model improvements. The performance of the
optimization code thus resulted in a great deal of "free" time that could be used for the
"real" hydrological work. A fact not foreseen when we started.
Fig. 2 Phreatic aquifer, computed groundwater heads (colour of the model cells) and their deviations from measurements (yellow and red dots) after the first optimization run by MODINV. Purple lines show the boundary of the peat layer between the upper two aquifers.
Fig. 3 Phreatic aquifer, computed groundwater heads (colour of the model cells) and their deviations from measurements (yellow and red dots) after the final, 67th, optimization run by MODINV.
Fig. 4 Third aquifer: computed groundwater heads (colour of the model cells) and their deviations from measurements (yellow and red dots) after the final, 67th, optimization run by MODINV.
REFERENCES


