Université de Liège
Faculté des Sciences Appliquées
Département ArGEnCo
Architecture, Géologie, Environnement et Constructions
Secteur GEO³
Géotechnologies, Hydrogéologie, Prospection Géophysique

Assessing the impacts of technical and structure choices on groundwater model performance using a complex synthetic case

Thèse de doctorat
présentée par Samuel WILDEMEERSCH
en vue de l'obtention du grade de Docteur en Sciences de l'Ingénieur

2012

Université de Liège
Faculté des Sciences Appliquées
Département ArGEnCo
Architecture, Géologie, Environnement et Constructions
Secteur GEO³
Géotechnologies, Hydrogéologie, Prospection Géophysique

Assessing the impacts of technical and structure choices on groundwater model performance using a complex synthetic case

Thèse de doctorat
présentée par Samuel WILDEMEERSCH
en vue de l'obtention du grade de Docteur en Sciences de l'Ingénieur

2012

This research was carried out by:

Samuel WILDEMEERSCH
Hydrogeology and Environmental Geology
Building B52/3 – Sart-Tilman
B-4000 Liege

This research was funded by:

FRIA-F.R.S.-FNRS – Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture
Rue d Egmont, 5
B-1000 Brussels

and

Interuniversity Attraction Pole TIMOTHY (IAP Research Project P6/13) of the Belgian Federal Science Policy Office (BELSPO)

Board of examiners composed of:

Frédéric NGUYEN, Université de Liège (Belgium) – President
Alain DASSARGUES, Université de Liège (Belgium) – Supervisor
Serge BROUYERE, Université de Liège (Belgium) – Co-supervisor
Eric DELHEZ, Université de Liège (Belgium)
Benjamin DEWALS, Université de Liège (Belgium)
Okke BATELAAN, Vrije Universiteit Brussel (Belgium)
René THERRIEN, Université Laval (Canada)

Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.
George E. P. Box

Abstract

According to the European Water Framework Directive (2000/60/EC) and the specific Groundwater Directive (2006/188/EC), Member States have to manage groundwater at the groundwater body scale and in an integrated way. Given the objectives of “good quantitative and qualitative status” of groundwater for 2015 stated by the Directive, end-users want to know the quantitative and qualitative evolution of groundwater for several scenarios. Physically-based and spatially-distributed groundwater flow and transport models constitute useful management tools in this context since they take explicitely into account the heterogeneity and the physical processes occuring in the subsurface for predicting system responses to future stress factors. However, at such a scale, groundwater flow and transport modelling is challenging due to (1) the complexity of geological and hydrogeological contexts, (2) the uneven level of characterisation knowledge, and (3) the representativity of measured parameters. Furthermore, such models require long execution times. As a consequence, a series of choices and simplifications are made for dealing with these issues. Therefore, the outstanding question is to know whether end-users’ expectations can be met in spite of such choices and simplifications. This work focuses on choices and simplifications related to spatial discretisation and saturation–pressure relations in the unsaturated zone. The influence of stress factor time resolution is also tested.

Considering this general context, the objective of the present work is to evaluate the influence of some model technical (spatial discretisation) and structure (saturation–pressure relations) uncertainties on model results, parameter sensitivities, and optimisation performance in order to provide guidelines for model development. This is performed using a synthetic case inspired by typical groundwater bodies of Wallonia (Belgium). This synthetic case is used for obtaining reference observations in terms of flow rates and hydraulic heads. These reference observations are then compared with their simulated equivalent produced by simplified models differing by their spatial discretisation, their saturation–pressure relations in the unsaturated zone, or the time resolution of their stress factors. The simplified models are then ranked using several performance criteria measuring the discrepancies between reference observations and their simulated equivalent. This ranking leads to guidelines for large-scale groundwater flow model development with respect to typical end-users' expectations.

Whatever the time resolution of stress factors, the quantitative and qualitative analyses performed indicate that coarsening horizontal spatial discretisation deteriorates mainly the simulation of flow rates, coarsening vertical spatial discretisation deteriorates mainly the simulation of hydraulic heads, and (over)simplifying saturation–pressure relations in the unsaturated zone significantly impair the simulation of both flow rates and hydraulic heads. Although optimisation can compensate for errors induced by model technical and structure uncertainties, the improvement of model fit is limited, especially for the coarsest models. Furthermore, with respect to end-users' expectations, the weighted least-square objective function is not always the most relevant criteria for optimising models. Therefore, it is essential to use specific performance criteria for evaluating model performance depending on the

objectives of the study. The ideal would be to develop an end-users objective function for including such performance criteria in the optimisation process and stop the optimisation process once performance criteria would have reached the values specified by the end-users with respect to the objectives of the study.

Keywords: large-scale groundwater flow model, synthetic case, technical uncertainties, structure uncertainties, performance criteria.

Acknowledgements

Although a single name is visible on the cover of this thesis, I could not have completed this work without the help and the encouragements of numerous people.

I would like to express special thanks to Alain Dassargues and Serge Brouyère who supervised this thesis. Thank you for your guidance and support during these years. Thank you for your careful reading of my manuscript and for your proposals and suggestions for improving its quality. Thank you for having introduced me to research and for having given me the opportunity to participate in a series of international conferences and meetings. I am looking forward for our next trip in Tucson with Philippe, Pascal, René... and the crazy juke box girl!

Thank you to Frédéric Nguyen, Eric Delhez, Benjamin Dewals, Okke Batelaan, and René Therrien who have accepted to join the jury of this thesis and to take time to read this work.

I would also like to express special thanks to my colleagues and friends from the Hydrogeology and Environmental Geology group and from other teams of the Department: Philippe, Pascal, Julie G., Ingrid, Pierre J., Fabien, Frédérique, Pierre B., Mylène, Juan, Natalia, Joël, David, Jean, Tanguy, Thomas, and Arnaud. Working with you all is a great pleasure. Thank you in particular to Philippe and Pascal for their help on almost everything in groundwater

modelling, to Ingrid for her help on ArcGIS, and to Julie G. and Pierre J. for their help on the characterisation study they performed on groundwater body RWM021. Thank you also to Julie G. for her anticipative reading of my thesis. I would also thank former colleagues: Julie C., Mathieu, Jordi, Laurent, Piotr, Nicolas, François, Florence, Simon, Benjamin, Gaëlle, and Max. Thank you in particular to Julie C. for her help on SUFT3D.

Thank you to Martine, Nadia, and Christiane, the secretary staff, for their availability. Thank you to Annick Anceau and the other members of the staff of the Earth Sciences library for their help in the research of publications.

I would also like to thank David Colignon for helping me to use NIC3 for my simulations.

Thank you to René Therrien and Jean-Michel Lemieux for having given me the opportunity to spend three-month stay in their team at the University Laval (Quebec city, QC, Canada). I would particularly thank them and their family for their warm welcome. Thank you also to John Molson and Pierre Therrien. I would also like to thank Tanguy, Catherine, and their cute baby Esme for the very pleasant time we spent as roommates in Quebec city. Thank you also to Jalil, Tobias, Guillaume, Fabien, and Nelly for having replaced my family during this three months stay. I guess they know what I mean.

I would also like to thank Eileen Poeter, Mary Hill, and John Doherty for their quick replies to my questions about MMA, UCODE_2005, and PEST.

Thanks to the FRIA-F.R.S.-FNRS for having funded my thesis together with the Interuniversity Attraction Pole TIMOTHY (IAP Research Project P6/13) of the Belgian Federal Science Policy Office (BELSPO).

Last but certainly not least, I would like to express warm thanks to my family and my friends for their continuous support. Thank you in particular to my parents who are always there for me. Thank you also in particular to

Audrey for her comprehension and her patience during these last months. Thank you for having accompanied me in Quebec city... and thank you for having replied “yes” to my proposal three months ago!

Lastly, I would like to dedicate this thesis to my grand-parents, who never had the chance to go to university.

Samuel Wildemeersch

January 2012

Knowledge dissemination

Journal articles

- *Wildemeersch, S.*, S. Brouyère, Ph. Orban, J. Couturier, C. Dingelstadt, M. Veschkens, and A. Dassargues (2010). Application of the Hybrid Finite Element Mixing Cell method to an abandoned coalfield in Belgium. *Journal of Hydrology 392*(3-4), 188-200.

- *Wildemeersch, S.*, Ph. Orban, I. Ruthy, O. Grière, Ph. Olive, A. El Youbi, and A. Dassargues (2010). Towards a better understanding of the Oulmes hydrogeological system (Mid-Atlas, Morocco). *Environmental Earth Sciences 60*(8), 1753-1769.

- Brouyère, S., Ph. Orban, *S. Wildemeersch*, J. Couturier, N. Gardin, and A. Dassargues (2009). The Hybrid Finite Element Mixing Cell method: a new flexible method for modelling mine ground water problems. *Mine Water and the Environment 28*(2), 102-114.

- *Wildemeersch, S.*, P. Goderniaux, Ph. Orban, S. Brouyère, and A. Dassargues (2012). Assessing the impacts of technical choices on large-scale groundwater flow model results using a complex synthetic case. Manuscript in preparation.

- Goderniaux, P., *S. Wildemeersch*, S. Brouyère, and A. Dassargues

(2012). Are calculated climate change impacts on groundwater certain? Manuscript in preparation.

- Sauret, E., F. Nguyen, J. Beaujean, N. Gardin, *S. Wildemeersch*, S. Brouyère, A. Dassargues, and A. Biaou (2012). Contribution of Electrical Resistivity Tomography (ERT) and Horizontal and Vertical Spectral Ratio (H/V SR) to characterize the alluvial deposits in the Kou basin (Burkina Faso). Manuscript in preparation.

Conference proceedings

- Brouyère, S., *S. Wildemeersch*, Ph. Orban, M. Leroy, J. Couturier, and A. Dassargues (2011). The Hybrid Finite-Element Mixing-Cell method: a candidate for modelling groundwater flow and transport in karst systems. 9th Conference on Limestone Hydrogeology. Besançon, France. 1-4 September 2011. In C. Bertrand, N. Carry, J. Mudry, M. Pronk, and F. Zwahlen (Eds.), *Proceedings of H2Karst*, 79-82.

- Goderniaux, P., S. Brouyère, Ph. Orban, *S. Wildemeersch*, and A. Dassargues (2011). Uncertainty of climate change impact on groundwater resources considering various uncertainty sources. IUGG 2011 - Symposium H01: Conceptual and Modelling Studies of Integrated Groundwater, Surface Water, and Ecological Systems. Melbourne, Australia. 28 June-7 July 2011. In C. Abesser, G. Nutzmann, M. Hill, G. Bloschl, and E. Lakshmanan (Eds.), IAHS Press, *Conceptual and Modelling Studies of Integrated Groundwater, Surface Water, and Ecological Systems*, 139-144.

- *Wildemeersch, S.*, S. Brouyère, Ph. Orban, J. Couturier, N. Gardin, C. Dingelstadt, M. Veschkens, and A. Dassargues (2009). Application of the

HFEMC method to an abandoned coalfield in Belgium: from conceptualisation to scenario simulations. IMWA 2009: Mine Water & Innovative Thinking. Pretoria, South Africa. 19-23 October 2009. In Water Institute of Southern Africa & International Mine Water Association, *Proceedings, International Mine Water Conference*, 70-78.

- Brouyère, S., Ph. Orban, *S. Wildemeersch*, J. Couturier, N. Gardin, and A. Dassargues (2008). The Hybrid Finite Element Mixing Cell method: a new flexible method for modelling mine water problems. 10th International Mine Water Association Congress. Carlsbad, Czech Republic. 2-5 June 2008. In N. Rapantova and Z. Hrkal (Eds.), *Proceedings of IMWA 2008*, 429-431.

Conference abstracts

- *Wildemeersch, S.*, P. Goderniaux, Ph. Orban, S. Brouyère, and A. Dassargues (2011). Assessing the impacts of technical uncertainty on coupled surface/subsurface flow model predictions using a complex synthetic case. MODELCARE 2011: Models - Repositories of knowledge. Leipzig, Germany. 18-22 September 2011.

- *Wildemeersch, S.*, P. Goderniaux, M. Leroy, Ph. Orban, A. Dassargues, and S. Brouyère (2011). Regional scale groundwater flow and transport modelling: from conceptual challenges to pragmatic numerical solutions. IUGG 2011 - Earth on the Edge: Science for a Sustainable Planet. Melbourne, Australia. 28 June-7 July 2011.

- Leroy, M., Ph. Orban, J. Gesels, P. Jamin, *S. Wildemeersch*, P. Goderniaux, and A. Dassargues (2011). Regional scale flow and transport modelling for the management of groundwater and surface water bodies in

the framework of the EU Water Directive. HydroEco'2011: Ecosystems, Groundwater and Surface Water - Pressures and Options. Vienna, Austria. 2-5 May 2011.

- Orban, Ph., S. Brouyère, J. Couturier, *S. Wildemeersch*, P. Goderniaux, J. Batlle-Aguilar, and A. Dassargues (2009). Assessment of nitrate trends in groundwater using the regional scale HFEMC approach. Ground Water Summit 2009 and Ground Water Protection Council Spring Meeting 2009: Adapting to Increasing Demands in a Changing Climate. Tucson, AZ, USA. 19-23 April 2009.

- *Wildemeersch, S.*, Ph. Orban, S. Brouyère, J. Couturier, N. Gardin, and A. Dassargues (2009). Ground water flow simulation in mining works: application of the Hybrid Finite Element Mixing Cell method. Ground Water Summit 2009 and Ground Water Protection Council Spring Meeting 2009: Adapting to Increasing Demands in a Changing Climate. Tucson, AZ, USA. 19-23 April 2009.

- Orban, Ph., S. Brouyère, *S. Wildemeersch*, J. Couturier, and A. Dassargues (2008). The Hybrid Finite-Element Mixing-Cell method: a new flexible method for large scale groundwater modelling. CMWR XVII. San Francisco, CA, USA. 6-10 July 2008.

Scientific reports

- Beaujean, J., F. Nguyen, E. Sauret, N. Gardin, *S. Wildemeersch*, and S. Brouyère (2009). Prospection géophysique de la nappe alluviale dans la région de Nasso, Bobo-Dioulasso. 49 p.

- *Wildemeersch, S.*, J. Couturier, Ph. Orban, N. Gardin, S. Brouyère, and A.

Dassargues (2008). Modélisation d'une partie du bassin charbonnier de Liège selon une approche mixte au moyen du logiciel SUFT3D. Rapport RW08/01. Région Wallonne. 108 p.

Scientific reports - IAP-TIMOTHY

- ULB, VUB, UCL, ULg, MRAC, and UPMC (2008). **T**racing and **I**ntegrated **Mo**deling of Natural and Anthropogenic Effec**t**s on **Hy**drosystems: The Scheldt River Basin and Adjacent Coastal North Sea (**TIMOTHY**) - First annual scientific report. March 2008.

- ULB, VUB, UCL, ULg, MRAC, and UPMC (2009). **T**racing and **I**ntegrated **Mo**deling of Natural and Anthropogenic Effec**t**s on **Hy**drosystems: The Scheldt River Basin and Adjacent Coastal North Sea (**TIMOTHY**) - Second annual scientific report. March 2009.

- ULB, VUB, UCL, ULg, MRAC, and UPMC (2010). **T**racing and **I**ntegrated **Mo**deling of Natural and Anthropogenic Effec**t**s on **Hy**drosystems: The Scheldt River Basin and Adjacent Coastal North Sea (**TIMOTHY**) - Expost report.

- ULB, VUB, UCL, ULg, MRAC, and UPMC (2011). **T**racing and **I**ntegrated **Mo**deling of Natural and Anthropogenic Effec**t**s on **Hy**drosystems: The Scheldt River Basin and Adjacent Coastal North Sea (**TIMOTHY**) - Fourth annual scientific report. March 2011.

Contents

Chapter 1

General context and objectives of the research

According to the European Water Framework Directive (2000/60/EC) and the specific Groundwater Directive (2006/188/EC), a “good quantitative and qualitative status” of groundwater is expected for 2015. Furthermore, Member States have to manage groundwater at the groundwater body scale and in an integrated way. While a “good qualitative status” implies essentially to identify and reverse upward trends of groundwater contamination, a “good quantitative status” requires primarily that groundwater levels are high enough to feed surface water, wetlands, and related ecosystems.

Large-scale groundwater flow and transport models are particularly useful in such a context for their capacity of gathering every piece of information obtained on the system in order to simulate its quantitative and qualitative evolution for several predictive scenarios. Given the objectives of the Directive from a quantitative point of view, these models should provide information on both baseflow rates and hydraulic heads in order to evaluate the impacts of future changes in stress factors on groundwater resources. Typically, end-

users are interested in one or some of these model outputs:

- average magnitude, maximum value, or general evolution of baseflow rates,
- average magnitude, seasonal variations, or general evolution of hydraulic heads,
- hydraulic head maps.

Physically-based and spatially-distributed models prevail in such cases for their unique capacity of taking explicitely into account the heterogeneity and the physical processes occuring in the subsurface for predicting the quantitative evolution of groundwater resources (Dassargues et al., 1988; Beven, 1989; Beven and Freer, 1992; Grayson et al., 1992; Beven, 1993; Smith et al., 1994; Beven, 1996a,b; Refsgaard et al., 1996; Heng and Nikolaidis, 1998; Beven, 2000, 2001a,b, 2002a,b; Brouyère et al., 2004; Loague and VanderKwaak, 2004; Panday and Huyakorn, 2004; Beven, 2006; Ebel and Loague, 2006; Loague et al., 2006; Li et al., 2008; Goderniaux et al., 2009; Spanoudaki et al., 2009; Orban et al., 2010). However, large-scale groundwater flow and transport modelling is challenging due to (1) the complexity of geological and hydrogeological contexts, (2) the uneven level of characterisation knowledge, and (3) the representativity of measured parameters. These specific issues, especially the scarcity of data, can even prevent from using physically-based and spatially-distributed models at large-scale (Barthel et al., 2008). Additionally, such models require long execution times. Therefore, a series of choices and simplifications constituting sources of uncertainty for model simulations are inevitable for dealing with these issues and reducing the execution times of large-scale physically-based and spatially-distributed models. The present work focuses on choices and simplifications related to spatial discretisation

and saturation–pressure relations in the unsaturated zone. According to Refsgaard et al. (2007), the uncertainty coming from such choices and simplifications are part of model technical uncertainty and model structure uncertainty, respectively. Furthermore, the influence of stress factor time resolution is also tested. The outstanding question is to evaluate whether end-users' expectations can be met in spite of such choices and simplifications (Figure 1.1).

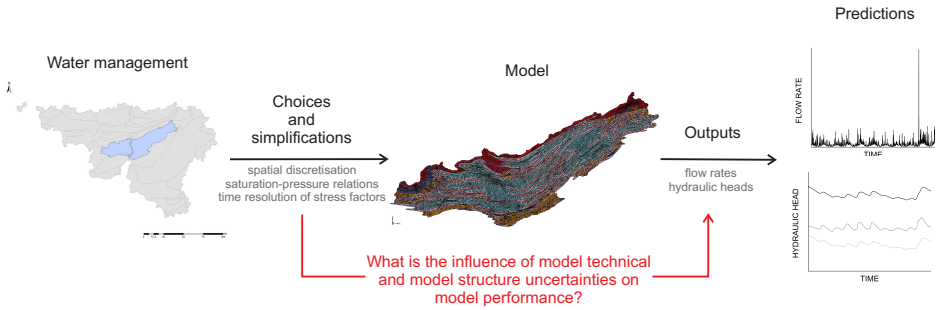

Figure 1.1: A series of choices and simplifications is inevitable for dealing with the specific issues of large-scale physically-based and spatially-distributed groundwater flow and transport models. The outstanding question is to know whether end-users' expectations can be met in spite of such choices and simplifications.

Given this general context, the objectives of this work are:

- to evaluate the influence of some model technical and structure uncertainties on modelling results, parameter sensitivities, and optimisation performance,
- to provide guidelines for large-scale groundwater flow model developement, including on the selection of specific performance criteria, with respect to typical end-users' expectations.

These tasks are undertaken using a synthetic case inspired by typical groundwater bodies of Wallonia (Belgium). While this synthetic case is developed with the fully-integrated and physically-based surface and subsurface flow model HydroGeoSphere (Therrien et al., 2005), this work focuses primarily

on groundwater.

Chapter 2 exposes the literature review performed in the field of model calibration, model evaluation, and model ranking. Chapter 3 presents the methodology developed for evaluating the influence of model technical and model structure uncertainties on model results, parameter sensitivities, and optimisation performance. Chapter 4 presents the governing equations and the main outputs of HydroGeoSphere. Chapter 5 exposes the generation of the synthetic case and its key characteristics. Chapter 6 presents the simplified models and the qualitative evaluation performed for each of them. The influence of model technical and model structure uncertainties on parameter sensitivities is also presented in this chapter. Chapter 7 exposes the quantitative evaluation performed for the simplified models using several performance criteria. Chapter 8 presents the model ranking obtained based on information theory and the model ranking obtained based on performance criteria. Chapter 9 provides general conclusions, including guidelines for large-scale groundwater flow model development, and perspectives for future works.

Chapter 2

Literature review

2.1 Complexity versus parsimony

The subsurface environment is inherently complex. Therefore, a model is always a simplification of the reality whose major objective is to provide predictions with the smallest uncertainty possible. The question of which level of model complexity is needed for achieving this objective is still an open debate within the scientific community, some defending the principle of complexity, others defending the principle of parsimony (Occam's razor). The strengths and weaknesses of each paradigm are presented in Gómez-Hernández (2006) and Hill (2006), respectively.

Advocates of complexity work in a stochastic framework. Their objective, consistent with the equifinality concept stating that a lot of model structures and parameter sets can reproduce the observed behaviour of a system (Beven and Freer, 2001), is to substitute the unknown subsurface environment by a series of plausible models given the available data (Gómez-Hernández, 2006). These plausible models can play on each element constituting a model, from the boundary conditions to the parameter values, and

probability functions are associated with each of them. The result is not a single model but a large set of uncalibrated model realisations whose likelihoods are measured with methods such as Generalised Likelihood Uncertainty Estimation (GLUE) (Beven and Freer, 1992), Bayesian Model Averaging (BMA) (Draper, 1995; Kass and Raftery, 1995; Hoeting et al., 1999), and Maximum Likelihood Bayesian Model averaging (MLBMA) (Neuman, 2003). A GLUE-BMA combination is also possible (Rojas et al., 2008, 2009). These methods measure the likelihood of each model realisation depending on their capacity of reproducing observed parameter values and state variables as well as their preservation of some predefined structure about the heterogeneity. This last condition most often leads to models striving for capturing the full heterogeneity of the system. Therefore, model realisations are often made up of hundreds of parameters which is a common indicator of model complexity.

Partisans of simplicity work in a deterministic framework with a single or a few calibrated models. Their objective is to keep the model as simple as possible while considering the characteristics and processes of the system that are evident from the observations and important for predictions, and while respecting all system information (Hill, 2006; Hill and Tiedeman, 2007). This is primordial for preserving refutability and transparency (Oreskes, 2000). The concept of refutability implies that each hypothesis made during model construction is testable. The concept of transparency implies that model dynamics are understandable. Generally, both concepts suffer when the number of parameters increases and the execution times get longer (Hill, 2006; Hill and Tiedeman, 2007). The importance of keeping simplicity in modelling is supported by the tradeoff observed between model fit and prediction error with respect to the number of parameters. This tradeoff is illustrated with a simple example in Hill (2006) and Hill and Tiedeman (2007). They compare the model

fit and the predictive capability of a linear model (simple model) and a polynomial model (complex model) in case of a true linear model (Figure 2.1). The polynomial model produces clearly a better fit but poorer predictions since it matches the errors in the data rather than the system processes. This problem was investigated for the first time in the groundwater literature by Yeh and Yoon (1981).

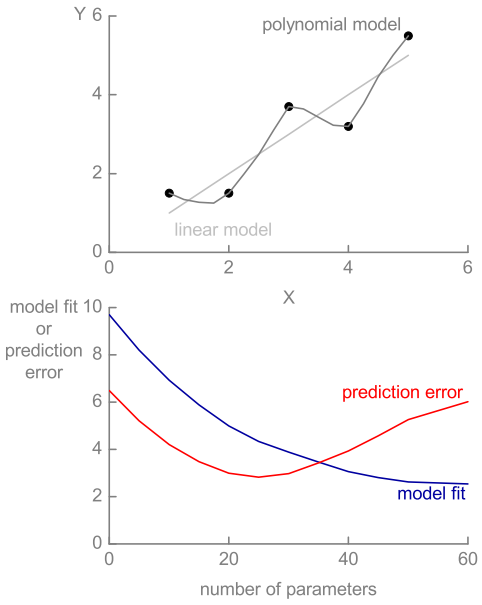

Figure 2.1: The tradeoff between model fit and prediction error is illustrated comparing linear and polynomial models. Considering a true linear model, it is clear that the polynomial model matches the errors in the data rather than the system processes. Therefore, in spite of a better model fit, its prediction error is greater (modified from Hill, 2006).

According to Hill (2006), a simple model with short execution times is also a prerequisite for performing effective sensitivity analysis, calibration, and uncertainty analysis such as proposed in her guidelines (Hill and Tiedeman, 2007). The sensitivity analysis is particularly useful for investigating the observations–parameters–predictions interactions in order to know which of the processes and parameters are essential for the predictions of interest.

Additionally, a parsimonious selection of spatial and time discretisations is also important to ensure short execution times. The selection of processes, parameters, and spatial and time discretisations should depend on the availability of data, the hydrogeological context, and the objectives of the study (Hill, 2006). This point of view is also supported by Refsgaard (1997) and Barthel et al. (2008).

The presentation of the opposite paradigms, complexity versus parsimony, shows that they differ essentially by their way of dealing with model uncertainty. The supporters of complexity consider a large set of complex models and estimate the likelihood of each of them in order to evaluate model uncertainty (stochastic framework). From their point of view, models taking explicitly into account the heterogeneity are the most plausible. The supporters of parsimony develop a single or a few simple models for which they perform sensitivity analysis, calibration, and uncertainty analysis (deterministic framework). From their point of view, a simple model with short execution times is preferable for investigating thoroughly the system functioning and selecting only the processes and parameters essential for the predictions of interest. The philosophy of the principle of complexity and the principle of parsimony are thus opposed, yet, they both have the same objectives of providing model predictions with the smallest uncertainty possible and quantifying this uncertainty. However, with current computer capabilities, the prohibitive number of model runs required for performing stochastic methods prevent from using them with large-scale physically-based and spatially-distributed models due to their long execution times.

2.2 Effective model calibration

2.2.1 Sensitivity analysis

The sensitivity analysis consists in studying how the variations in the model outputs (e.g. flow rates and hydraulic heads) are related to the variations in the model inputs (e.g. hydraulic conductivities) (Saltelli et al., 2000). Global sensitivity analysis performs this study over the full range of parameter values. This implies a lot of model runs and prevents from using global techniques for models with long execution times (Mishra et al., 2009). Local sensitivity analysis performs this study in the neighbourhood of a single set of parameter values. This requires far fewer model runs. Theoretically, local techniques are only valid for linear models. Practically, Hill and Tiedeman (2007) consider that models have to be extremely nonlinear for these techniques to fail completely. This is supported by the work of Foglia et al. (2009) who have successfully used local techniques on a large-scale physically-based and spatially-distributed hydrological model with long execution times. Therefore, local techniques are used in this work.

A series of statistics based on local sensitivities have been developed for facilitating the evaluation of interactions between observations and parameters (Hill and Tiedeman, 2007). These statistics are said to be fit-independent because they do not depend on the discrepancies between observed values and their simulated equivalent. They rely only on the sensitivities and the weighting.

The sensitivities J_{ij} are calculated as the derivatives of simulated equivalent to the observations with respect to the model parameters:

$$J_{ij} = \left(\frac{\partial y_i^{sim}}{\partial b_j} \right) \Bigg|_{\mathbf{b}} \tag{2.1}$$

where y_i^{sim} is the simulated equivalent to the i^{th} observation, b_j is the j^{th} parameter, and the subscript indicates that the sensitivities are calculated for the parameter values composing the vector **b**. This formulation indicates that the sensitivity reflects the slope of the input–output relationship at the reference point **b**. The sensitivity matrix **J**, also known as Jacobian matrix, is composed of the sensitivities:

$$\mathbf{J} = \begin{pmatrix} \left(\frac{\partial y_1^{sim}}{\partial b_1}\right)\Big|_{\mathbf{b}} & \cdots & \left(\frac{\partial y_1^{sim}}{\partial b_{npar}}\right)\Big|_{\mathbf{b}} \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial y_{nobs}^{sim}}{\partial b_1}\right)\Big|_{\mathbf{b}} & \cdots & \left(\frac{\partial y_{nobs}^{sim}}{\partial b_{npar}}\right)\Big|_{\mathbf{b}} \end{pmatrix} \tag{2.2}$$

where $nobs$ is the number of observations and $npar$ is the number of parameters.

Adjoint states and sensitivity-equation sensitivities produce the most accurate sensitivities (Hill and Tiedeman, 2007). However, these techniques are complex to program and sensitivities are most often estimated using backward, central or forward differences:

$$\left(\frac{\partial y_i^{sim}}{\partial b_j}\right)\Big|_{\mathbf{b}} \approx \frac{y_i^{sim}(\mathbf{b}+\Delta\mathbf{b}) - y_i^{sim}(\mathbf{b})}{\Delta b_j} \tag{2.3}$$

where $\Delta\mathbf{b}$ is a vector whose j^{th} element equals Δb_j (other elements equal zero).

The purposes of the weighting are to emphasise the most accurate observations and to produce dimensionless statistics. This is achieved by specifying weights that are proportional or, preferably, equal to the inverse of the observation error variances (Hill and Tiedeman, 2007). If the observation errors are independent, the weight matrix is diagonal, otherwise, it is a full matrix taking into account the correlations between observation errors. How-

ever, experience indicates that omitting such correlations has limited effect on paramer estimates (Hill and Tiedeman, 2007). Therefore, observation errors are assumed independent in this work and all the equations are written for a diagonal weight matrix whose diagonal elements w_{ii} are calculated as (Hill and Tiedeman, 2007):

$$w_{ii} = \frac{1}{\sigma_i^2} \tag{2.4}$$

where w_{ii} is the weight relative to the i^{th} observation and σ_i^2 is the true error variance of the i^{th} observation. This ensures that the most accurate observations, that is those with the smallest error variances, have the heaviest weights. The observation error variances are easily estimated following the procedure presented in Hill and Tiedeman (2007). This simple procedure is sufficient since only a rough evaluation of the weights is required (Foglia et al., 2009). When observation errors are estimated using coefficients of variation cv_i instead of standard deviations σ_i, the weights are calculated as (Hill and Tiedeman, 2007):

$$w_{ii} = \frac{1}{\left(cv_i \times y_i^{obs}\right)^2} \tag{2.5}$$

where y_i^{obs} is the i^{th} observed value. When working with transient models, a constant coefficient of variation is often used for calculating weights pertaining to flow rate observations. This is performed for providing weights inversely proportional to observed values and preventing low flow rates from being overwhelmed by high flow rates in sensitivity, model fit, and uncertainty analyses. This is illustrated in Table 2.1.

Table 2.1: Example illustrating the use of a constant coefficient of variation for providing weights proportional to observed values.

y_i^{obs}	cv_i	w_{ii}
1.0	0.10	100
10.0	0.10	1

The fit-independent statistics developed for investigating the relations between parameters and observations include the dimensionless scaled sensitivities (*dss*), the composite scaled sensitivities (*css*), and the parameter correlation coefficient (*pcc*). The dimensionless scaled sensitivities dss_{ij} are calculated as (Hill, 1992; Hill et al., 1998; Hill and Tiedeman, 2007):

$$dss_{ij} = \left(\frac{\partial y_i^{sim}}{\partial b_j} \right) \Bigg|_{\mathbf{b}} \times \left| b_j \right| \times w_{ii}^{1/2} \qquad i = 1, nobs \qquad j = 1, npar \tag{2.6}$$

This statistic evaluates the importance of a single observation y_i to the estimation of the single parameter b_j.

The composite scaled sensitivities css_j are calculated as (Hill, 1992; Anderman et al., 1996; Hill et al., 1998; Hill and Tiedeman, 2007):

$$css_j = \left[\frac{\sum_{i=1}^{nobs} \left(dss_{ij} \right)^2 \Big|_{\mathbf{b}}}{nobs} \right]^{1/2} \qquad j = 1, npar \tag{2.7}$$

This statistic measures the information provided by the entire set of observations for the estimation of the single parameter b_j. Large values correspond to parameters for which the observations provide a lot of information. The easiest way of using *css* is to compare the values obtained for each parameter. This is particularly useful for selecting the parameters to be included in the optimisation. According to Hill et al. (1998) and Hill and Tiedeman (2007),

parameters with composite scaled sensitivities less than 1 or less than 0.01 of the largest composite scaled sensitivity will produce problems in the regression or poorly estimated parameters which means parameters with large confidence intervals. The solution consists in fixing the values of these parameters prior to the calibration or merging some of them for increasing their composite scaled sensitivities.

The parameter correlation coefficients pcc_{kl} are calculated as (Hill and Tiedeman, 2007):

$$pcc_{kl} = \frac{cov(b_k, b_l)}{var(b_k)^{1/2} \times var(b_l)^{1/2}} \qquad k = 1, npar \qquad l = 1, npar \quad (2.8)$$

where $cov(b_k, b_l)$ are the covariances between parameter b_k and b_l and $var(b_k)$ are the variances of each of the parameters. This statistic measures whether coordinated changes in two parameters would result in the same simulated values and, so, the same model fit to the observations and the same objective function value. Therefore, parameter correlation coefficient indicates which parameter values are likely to be estimated uniquely during the optimisation. According to Hill et al. (1998) and Hill and Tiedeman (2007), pairs of parameters with correlation coefficients greater than 0.95 may not be estimated uniquely. Typically, groundwater flow models show a high negative correlation between hydraulic conductivity and recharge rate since varying simultaneously these parameters in an opposite way has almost no effect on simulated hydraulic heads. An effective way for decreasing the correlation between pair of parameters consists in using several types of observation for combining local (e.g. hydraulic heads) and global information (e.g. flow rates) on the system dynamics (Anderman et al., 1996; Poeter and Hill, 1997;

Hill and Tiedeman, 2007). Concentration and temperature data can also be used as observations (for example, see Woodbury and Smith, 1988; Manning and Ingebritsen, 1999; Bravo et al., 2002; Saiers et al., 2004; Sanford et al., 2004a,b; Anderson, 2005; Burow et al., 2005; Niswonger et al., 2005).

A combined use of *css* and *pcc* constitutes a powerful tool for investigating the interactions between parameters and observations. According to Hill and Tiedeman (2007), *css* and *pcc* are the key statistics for selecting the parameters to be included in the optimisation.

2.2.2 Optimisation

The objective of the optimisation is to obtain the set of parameter values that produces the smallest value of an objective function measuring the discrepancies between observed values and their simulated equivalent. This task is performed manually (trial-and-error procedure) or automatically using typically nonlinear regression methods (inverse modelling). The latter technique is preferable since it eliminates the subjectivity of trial-and-error procedure and it provides useful statistics about parameters and state variables (Poeter and Hill, 1997). However, trial-and-error procedure and inverse modelling share the same issues of non-identifiability, non-uniqueness, and instability (Carrera et al., 2005). Non-identifiability occurs when several set of parameter values lead to the same solution of the forward problem. Non-uniqueness occurs when several set of parameters minimise the objective function equally well. Instability occurs when small changes in parameter or observation values lead to large changes in estimated parameter values. Another issue, termed insensitivity by Hill and Tiedeman (2007), occurs when the information contained in the observations is not sufficient to support the parameter estimation. A review of these problems is proposed by Yeh (1986)

and Carrera and Neuman (1986b). The latter suggest to use prior information on parameters (e.g. field measurement of hydraulic conductivity) to reduce the ill-posedness of the inverse problem.

The optimisation is only possible for a finite number of parameters. Therefore, prior to optimisation, parameterisation is required for expressing system properties in terms of a small number of model parameters. The most widely used parameterisation techniques are zonation, point estimation, and pilot points (Carrera et al., 2005). The zonation consists in subdividing the system in a set of zones characterised by constant parameter values. This quite rigid technique is often used in conjunction with geological maps whose geological units correspond to parameter zones. The point estimation is an extreme case of zonation where the number of zones equals the number of cells or elements. This technique allows taking into account a maximum of heterogeneity at the cost of a huge number of parameters. The pilot points consists in defining parameters at a set of points distributed throughout the model domain. The set of parameter values estimated at each pilot point is used for interpolating parameter values over the whole domain at each optimisation iteration. Typically, interpolation is performed using kriging though any other geostatistical method is possible. This flexible technique is capable of defining the heterogeneity by itself. However, the interpolation leads to smoothed variations of parameter values over the model domain. Furthermore, questions remain concerning the number of pilot points to use and their location. The suggestions of Doherty (2003) simply indicate that pilot points should be placed literally throughout the model domain though preferably in zones where heterogeneity is suspected and where observations are numerous. Pilot points, initially developed by de Marsily et al. (1984), have gained popularity over the last decades (for example, see Certes and de Marsily, 1991;

RamaRao et al., 1995; LaVenue et al., 1995; Vesselinov et al., 2001; Doherty, 2003; Hernández et al., 2003; Kowalsky et al., 2004; Wood et al., 2005; Wylie and Doherty, 2005; Moore and Doherty, 2006; Batlle-Aguilar, 2008).

The optimisation consists in modifying the value of the parameters defined through parameterisation in order to minimise the value of a predefined objective function. This objective function measures the match of observed to simulated values. The most widely used objective function is the weighted least-squares objective function (L_2 norm) $\mathbf{\Phi(b)}$ (Hill and Tiedeman, 2007):

$$\Phi(\mathbf{b}) = \sum_{i=1}^{nh} w_{h_i} \times [y_{h_i}^{obs} - y_{h_i}^{sim}(\mathbf{b})]^2 + \sum_{j=1}^{nq} w_{q_j} \times [y_{q_j}^{obs} - y_{q_j}^{sim}(\mathbf{b})]^2 + \sum_{k=1}^{npr} w_{p_k} \times [y_{p_k}^{obs} - y_{p_k}^{sim}(\mathbf{b})]^2 \tag{2.9}$$

where, for a groundwater flow problem, nh is the number of hydraulic head observations, nq is the number of flow rate observations, npr is the number of prior information values, $y_{h_i}^{obs}$ is the i^{th} observed hydraulic head, $y_{h_i}^{sim}$ is the simulated value equivalent to the i^{th} observed hydraulic head, $y_{q_j}^{obs}$ is the j^{th} observed flow rate, $y_{q_j}^{sim}$ is the simulated value equivalent to the j^{th} observed flow rate, $y_{p_k}^{obs}$ is the k^{th} prior estimate included in the optimisation, $y_{p_k}^{sim}$ is the k^{th} simulated value, w_{h_i} is the weigth for the i^{th} hydraulic head observation, w_{q_j} is the weigth for the j^{th} flow rate observation, w_{p_k} is the weigth for the k^{th} prior estimate. This objective function is more commonly expressed as:

$$\Phi(\mathbf{b}) = \sum_{i=1}^{nobs+npr} w_i \times \left[y_i^{obs} - y_i^{sim}(\mathbf{b}) \right]^2 \tag{2.10}$$

$$= \sum_{i=1}^{nobs+npr} w_i \times r_i^2 \tag{2.11}$$

where $nobs$ is the number of observations of any kind, y_i^{obs} is the i^{th} observation or prior information value, $y_i^{sim}(\mathbf{b})$ is the simulated equivalent to the i^{th} observed value, w_i is the weight for the i^{th} contribution to the objective function, r_i is the i^{th} residual. The matrix form of this objective function is:

$$\Phi(\mathbf{b}) = [\mathbf{y} - \mathbf{y}(\mathbf{b})]^T \mathbf{w} [\mathbf{y} - \mathbf{y}(\mathbf{b})] \tag{2.12}$$

$$= \mathbf{r}^T \mathbf{w} \mathbf{r} \tag{2.13}$$

where $\mathbf{y}$ is a vector of order $(nobs+npr)$ containing the observation values, $\mathbf{y}(\mathbf{b})$ is a vector of order $(nobs+npr)$ containing the simulated values equivalent to the observations, $\mathbf{r}$ is the residual vector of order $(nobs+npr)$, and $\mathbf{w}$ is the weight matrix with dimensions $(nobs+npr) \times (nobs+npr)$.

The single objective function expressed by equation 2.9 is sometimes replaced by multiple objective functions (for example, see Vrugt et al., 2003). This allows for instance considering hydraulic head observations and flow rate observations separately using specific performance criteria. When using multiple objective functions, the improvement of one objective function causes the deterioration of one or more others. Therefore, the principle of the multiobjective optimisation is to find the set of points in the parameter space along which it is not possible to lower the objective functions simultaneously. This set of points is called the Pareto front. The simplest way for estimating

the Pareto front consists in weighing each performance criteria and then running a large number of independent optimisation using different values for the weights. The study of Foglia et al. (2009) comes to the conclusion that single objective optimisation is preferable for models with long execution times since it requires less model runs, and that multiobjective optimisation is of interest only when it is important to explore alternative observation weightings.

The objective of the optimisation is to find the optimal parameter values, that is the parameter values that minimise the value of the objective function. This implies solving the normal equations obtained by taking the derivative of the objective function with respect to the parameters and setting the derivative equal to zero:

$$\frac{\partial}{\partial \mathbf{b}} \left[[\mathbf{y} - \mathbf{y}(\mathbf{b})]^T \mathbf{w} [\mathbf{y} - \mathbf{y}(\mathbf{b})] \right] = \mathbf{0} \tag{2.14}$$

where $\mathbf{0}$ is a vector of $npar$ elements equal to zero and $npar$ is the number of parameters.

If the system is linear of the form:

$$\mathbf{y}(\mathbf{b}) = \mathbf{X}\mathbf{b} \tag{2.15}$$

where $\mathbf{X}$ is a matrix with dimensions $nobs \times npar$ containing the set of excitations of the system, $\mathbf{b}$ is a vector of order $npar$ containing the system parameters, and $\mathbf{y}(\mathbf{b})$ is a vector of order $nobs$ containing the system responses to the set of excitations, the corresponding least-squares objective function is:

$$\Phi(\mathbf{b}) = [\mathbf{y} - \mathbf{y}(\mathbf{b})]^T [\mathbf{y} - \mathbf{y}(\mathbf{b})] \tag{2.16}$$

$$= (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b}) \tag{2.17}$$

where $\mathbf{y}$ is a vector of order $nobs$ containing the observations measured in a laboratory or in the field.

If the number of observations $nobs$ is greater or equal to the number of parameters $npar$, the solution of the problem is unique and it can be shown that the parameter values embodied in $\mathbf{b}$ that minimises the objective function $\Phi(\mathbf{b})$ is given by:

$$\mathbf{b} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{2.18}$$

Furthermore, it can be shown that the parameter variance–covariance matrix $\mathbf{V}(\mathbf{b})$ is given by:

$$\mathbf{V}(\mathbf{b}) = s^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tag{2.19}$$

where s^2 is the calculated error variance that represents the estimate of the variance of the observation values embodied in vector $\mathbf{y}$:

$$s^2 = \frac{\Phi(\mathbf{b})}{nobs - npar} \tag{2.20}$$

When the observations are weighted such as each weight is inversely proportional to the standard deviation of the corresponding observation, the objective function equation 2.17 is modified into:

$$\Phi(\mathbf{b}) = [\mathbf{y} - \mathbf{y}(\mathbf{b})]^T \mathbf{w} [\mathbf{y} - \mathbf{y}(\mathbf{b})] \tag{2.21}$$

$$= (\mathbf{y} - \mathbf{X}\mathbf{b})^T \mathbf{w} (\mathbf{y} - \mathbf{X}\mathbf{b}) \tag{2.22}$$

where $\mathbf{w}$ is the weight matrix. This equation is equivalent to the equation 2.13.

The equation 2.18 for calculating the optimal set of parameter values is modified into:

$$\mathbf{b} = \left(\mathbf{X}^T \mathbf{w} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{w} \mathbf{y} \tag{2.23}$$

while the equation 2.19 for calculating the parameter variance–covariance matrix is modified into:

$$\mathbf{V}(\mathbf{b}) = s^2 \left(\mathbf{X}^T \mathbf{w} \mathbf{X}\right)^{-1} \tag{2.24}$$

However, most of the natural systems are nonlinear so linearisation is required in order to use these equations for finding the optimal set of parameter values. This linearisation is performed using Taylor's theorem:

$$\mathbf{y}(\mathbf{b}) \approx \mathbf{y}(\mathbf{b}_0) + \mathbf{J}(\mathbf{b} - \mathbf{b}_0) \tag{2.25}$$

where $\mathbf{b}_0$ differs only slightly of $\mathbf{b}$ and $\mathbf{J}$ is the Jacobian matrix defined in equation 2.2. This approximation improves with proximity of $\mathbf{b}_0$ to $\mathbf{b}$.

The corresponding weighted least-squares objective function is given by:

$$\Phi(\mathbf{b}) = [\mathbf{y} - \mathbf{y}(\mathbf{b})]^T \mathbf{w} [\mathbf{y} - \mathbf{y}(\mathbf{b})] \tag{2.26}$$

$$= [\mathbf{y} - \mathbf{y}(\mathbf{b}_0) - \mathbf{J}[\mathbf{b} - \mathbf{b}_0]]^T \mathbf{w} [\mathbf{y} - \mathbf{y}(\mathbf{b}_0) - \mathbf{J}[\mathbf{b} - \mathbf{b}_0]] \tag{2.27}$$

where $\mathbf{y}$ is the observation vector. The equations 2.22 and 2.27 are similar if $\mathbf{y}$ and $\mathbf{b}$ of equation 2.22 are replaced by $[\mathbf{y} - \mathbf{y}(\mathbf{b}_0)]$ and $[\mathbf{b} - \mathbf{b}_0]$ of equation 2.27, respectively. Therefore, the equations defined previously can be used for calculating the parameter upgrade vector $[\mathbf{b} - \mathbf{b}_0]$ from the vector $[\mathbf{y} - \mathbf{y}(\mathbf{b}_0)]$ which defines the discrepancy between the observed values $\mathbf{y}$ and their simulated equivalent $\mathbf{y}(\mathbf{b}_0)$. The parameter upgrade vector $\mathbf{u}$ is thus calculated as:

$$\mathbf{u} = [\mathbf{J}^T \mathbf{w} \mathbf{J}]^{-1} \mathbf{J}^T \mathbf{w} [\mathbf{y} - \mathbf{y}(\mathbf{b}_0)] \tag{2.28}$$

$$= \left(\mathbf{J}^T \mathbf{w} \mathbf{J}\right)^{-1} \mathbf{J}^T \mathbf{w} \mathbf{r} \tag{2.29}$$

and the parameter variance–covariance matrix is modified into:

$$\mathbf{V}(\mathbf{b}) = s^2 \left(\mathbf{J}^T \mathbf{w} \mathbf{J}\right)^{-1} \tag{2.30}$$

Equation 2.29 is used iteratively for calculating the set of optimal parameter values from a set of initial parameter values estimated by the user (Figure 2.2). This iterative method is known as the Gauss–Newton gradient method. The equation 2.30 indicates that the parameter variances depend on the sensitivities (equation 2.1) as measures of the information provided by the observations on the parameters, the weights (2.4) as measures of the error in the observations, and the calculated error variance (2.20) as a measure of model

fit to the observations.

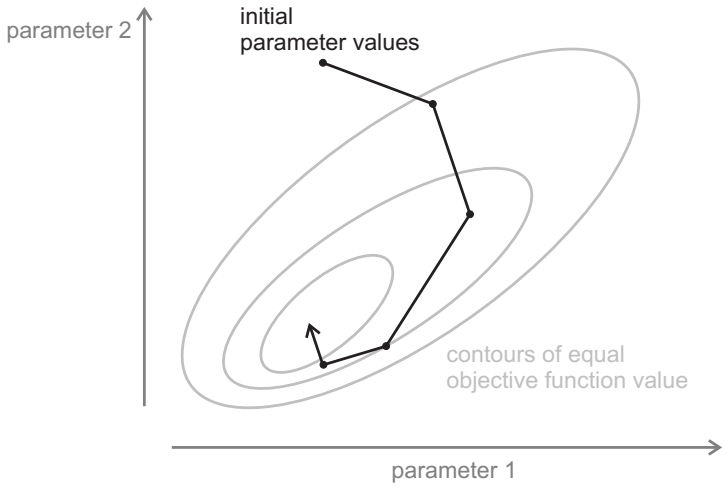

Figure 2.2: The Gauss-Newton method calculates iteratively a parameter upgrade vector for estimating the optimal parameter values, that is the parameter values that minimise the value of the objective function (modified from Doherty, 2005).

The Gauss-Newton method is prone to difficulties such as oscillations due to overshooting the optimal parameter values. Therefore, modifications have been developed for fixing these issues. These modifications consists in scaling the upgrade parameter vector **u** and modifying its direction and its magnitude. The scaling is performed for improving the accuracy of **u**. This is essential since observations, parameters, and thus sensitivities embodied in **J** have often values that differ by many orders of magnitude which can lead to roundoff error when calculating **u**. The scaling is implemented as:

$$\mathbf{S}^{-1}\mathbf{u} = ((\mathbf{JS})^T\mathbf{wJS})^{-1}(\mathbf{JS})^T\mathbf{wr} \tag{2.31}$$

where **S** is a diagonal scaling matrix with dimensions $npar \times npar$ whose element equals:

$$S_{ii} = \left(\mathbf{J}^T\mathbf{wJ}\right)_{ii}^{-1/2} \tag{2.32}$$

The modification of the direction of **u** is performed in incorporating the Marquardt parameter, developed by Levenberg (1944) and Marquardt (1963), in its calculation:

$$\mathbf{S}^{-1}\mathbf{u} = \left((\mathbf{JS})^T\mathbf{wJS} + \alpha_M\mathbf{S}^T\mathbf{S}\right)^{-1}(\mathbf{JS})^T\mathbf{wr} \tag{2.33}$$

where α_M is the Marquardt parameter. When α is equal to zero, equation 2.33 is equivalent to equation 2.29. When α_M is very high, the direction of the parameter upgrade vector gets close to the steepest descent direction given by the negative of the gradient of the objective function in the parameter space. The value of α_M is modified during the estimation process, high values are preferable in the initial stages of the optimisation and low values are preferable when the minimum value of the objective function is approached (Doherty, 2005) (Figure 2.3). The magnitude of **u** is also modified in order to estimate its optimum length and prevent overshooting problems. This Gauss–Newton gradient method improved with scaling and incorporation of the Marquardt parameter is termed Gauss–Netwon–Levenberg–Marquardt method or modified Gauss–Newton gradient method. Further information about this technique and its implementation in the inverse modelling codes PEST and UCODE_2005 can be found in Doherty (2005) and in Poeter et al. (2005) and Hill and Tiedeman (2007), respectively.

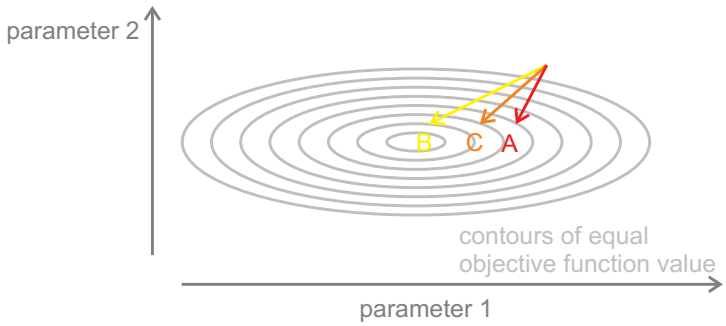

Figure 2.3: The direction of the upgrade parameter vector is optimised using the Marquardt parameter. This is illustrated using a linearised objective function with two parameters. The direction A points in the steepest descent direction. The direction B points in the direction estimated using the Gauss–Newton method (2.29 or its scaled version 2.31). The direction C points in the direction estimated using the Gauss–Newton–Levenberg–Marquardt with a nonzero Marquardt parameter (2.33). When the value of the Marquardt parameter is increased, the direction of the upgrade parameter vector gets closer to the steepest descent direction (gradient of the objective function with respect to the parameters) which is useful in the initial stages of the estimation process (modified from Hill and Tiedeman, 2007).

The Gauss–Newton–Levenberg–Marquardt method is powerful. However, in presence of local minima in the objective function, this method based on local sensitivities does not always provide the set of parameter values corresponding to the global minimum. The use of a temporary parameter immobilisation strategy, termed automatic user intervention (AUI) in PEST, greatly reduces this eventuality (Skahill and Doherty, 2006). This strategy consists in selectively withdrawing the most insensitive parameters from the estimation process when the objective function improvement during a particular iteration is poor. When a parameter is withdrawn, the upgrade parameter vector is recalculated and the model is rerun with the upgraded parameters to see whether the objective function is lowered. If it is the case, this parameter set constitutes the initial parameter set for the next iteration for which the temporarily frozen parameter is reinjected. If it is not the case, a further parameter is withdrawn and the procedure is repeated. According to Skahill and Doherty (2006), the heightened capacity of finding the global minimum of the objective function with the temporary parameter immobilisation strategy is

related to the large number of parameter upgrade vectors tested. Additionally, the directions of these parameter upgrade vectors tends to be maximally different. This is due to the fact that the upgrade parameter vector calculated with all the parameters is dominated by the most insensitive parameters which have to be strongly modified to have an influence on the objective function. When these insensitive parameters are frozen, the direction of the upgrade parameter vector calculated with the remaining parameters varies greatly (in fact, it tends to be orthogonal to the direction of the upgrade parameter vector calculated with all the parameters). Therefore, the temporary parameter immobilisation strategy allows investigating a large spectrum of the parameter space which is the key for finding the global optimum of the objective function. According to Doherty (2005), optimisation using truncated singular value decomposition (truncated SVD) gives similar results since this method also has the capacity of withdrawing insensitive parameters from the estimation process. The principle of this technique is to estimate only the sensitive combinations of parameters, these combinations corresponding to eigenvectors of the parameter variance–covariance matrix whose eigenvalues are significantly nonzero. Furthermore, this technique is stronger than the temporary parameter immobilisation in case of high parameter correlations (Doherty, 2005). An alternative for finding the global minimum of the objective function consists in using global optimisation methods rather than local optimisation methods such as the Gauss–Newton–Levenberg–Marquardt. These methods, instead of using sensitivities, find the next set of parameters using a long history of the model fit produced by previous set of parameters (for example, see Zheng and Wang, 1996; Solomatine et al., 1999; Tsai et al., 2003; Vrugt et al., 2003; Fazal et al., 2005; Moore et al., 2010; Keating et al., 2010). They have the capacity of always finding the global minimum of the

objective function. However, Hill and Tiedeman (2007) warn that global optimisation methods require a huge number of model runs inducing execution times tens or hundreds of times longer than the execution times required by local optimisation methods. This precludes using these methods with large-scale physically-based and spatially-distributed models due to their long execution times. Therefore, Gauss–Newton–Levenberg–Marquardt method enhanced with temporary parameter immobilisation is used in this work.

The Gauss–Newton–Levenberg–Marquardt method faces also issues when the number of parameters to be estimated exceeds the number of observations. The solution, in this case, is non-unique. This is one of the reason why Hill and Tiedeman (2007) plead to follow the principle of parsimony and to limit the number of parameters for complex systems. However, Hunt et al. (2007) say that this issue can be effectively circumvented using regularisation. This technique consists in introducing constraints on parameter values or on relationships between these values (e.g. smoothness) for obtaining parameter uniqueness. Therefore, zonation is an implicit form of regularisation since it imposes a constraint of parameter homogeneity in each zone. The principle of regularisation is to penalise the objective function when the predefined constraints are violated. This type of regularisation is termed Tikhonov regularisation (Tikhonov and Arsenin, 1977). When using regularisation, the objective function of equation 2.22 is modified into (Doherty, 2003):

$$\Phi_{tot} = (\mathbf{y} - \mathbf{Xb})^T \mathbf{w}_{obs} (\mathbf{y} - \mathbf{Xb}) + (\mathbf{y}' - \mathbf{Rb})^T \mathbf{w}_{reg} (\mathbf{y}' - \mathbf{Rb}) \tag{2.34}$$

$$= \gamma_L \times \Phi_{obs} + \Phi_{reg} \tag{2.35}$$

where $\mathbf{y}'$ is a vector containing the regularisation observations, $\mathbf{R}$ is the reg-

ularisation matrix depending on the parameter vector **b**, $\mathbf{w}_{reg}$ is a weight matrix containing the weights relative to the regularisation observations, and γ_L is a Lagrange multiplier used for formulating the constrained minimisation problem as an unconstrained minimisation problem. The subscripts obs and reg state for observation and regularisation, respectively.

The objective of the constrained optimisation is to minimise Φ_{reg} for preserving a preferred system state while keeping a suitable goodness of fit Φ_{obs}. This suitable goodness of fit is verified when Φ_{obs} is less or equal to a certain value Φ_{obs}^{max} specified by the user. Practically, Φ_{obs} will be equal to Φ_{obs}^{max} since a decrease of Φ_{reg} will nearly always lead to an increase of Φ_{obs} when parameter values are close to optimum. Therefore, with regularisation constraints, the optimisation consists in finding the parameter values **b** that minimise the total objective function Φ_{tot} while simultaneously finding the value of γ_L which satisfies $\Phi_{obs} = \Phi_{obs}^{max}$. This technique allows estimating a large number of parameters and stabilising the optimisation process by introducing a preferred system state which increases the parameter sensitivities (Doherty, 2003). Tikhonov regularisation has been used successfully in groundwater flow modelling (for example, see Skaggs and Kabala, 1994; Liu and Ball, 1999; Doherty, 2003; van den Doel and Ascher, 2006). Furthermore, a combination of Tikhonov regularisation and truncated singular value decomposition is possible for accelerating the optimisation process by solving the inverse problem in a parameter space reduced to the sensitive parameters only (Tonkin and Doherty, 2005; Hunt et al., 2007).

The use of regularisation allows including many parameters like in a stochastic framework (principle of complexity) while producing a single model like in a deterministic framework (principle of parsimony). Therefore, Hunt et al. (2007) say that regularisation is a middle ground between complexity

and parsimony. However, they recognize that neither Tikhonov regularisation nor hybrid Tikhonov–truncated singular value decomposition reduces the computational cost induced by a highly parameterised model because sensitivities have to be calculated with respect to all parameters during each iteration of the optimisation process. Therefore, these techniques are only possible for models with short execution times.

2.2.3 Uncertainty analysis

The uncertainty analysis consists in quantifying prediction uncertainty. The techniques developed for this purpose evaluate the magnitude of prediction uncertainty depending on the uncertainty in model parameters and the sensitivity of predictions to model parameters.

The inferential techniques are the simplest techniques for quantifying prediction uncertainty. They produce linear or nonlinear confidence intervals on predictions. The larger the interval, the greater the uncertainty. The calculation of linear confidence intervals involves calculating the prediction standard deviations $s_{z_k^{sim}}$ (Hill and Tiedeman, 2007):

$$s_{z_k^{sim}} = \left[\sum_{i=1}^{npar} \sum_{j=1}^{npar} \frac{\partial z_k^{sim}}{\partial b_j} \times \mathbf{V(b)} \times \frac{\partial z_k^{sim}}{\partial b_i} \right]^{1/2} \tag{2.36}$$

$$= \left[s^2 \left(\mathbf{j}_{z_k} \left(\mathbf{J}^T \mathbf{w} \mathbf{J} \right)^{-1} \mathbf{j}_{z_k}^T \right) \right]^{1/2} \tag{2.37}$$

where z_k^{sim} is the k^{th} prediction, $\mathbf{j}_{z_k}$ is a vector containing the sensitivities of prediction z_k^{sim}, and $\mathbf{V(b)}$ is the parameter variance–covariance matrix of equation 2.30. Then, linear confidence intervals are calculated as:

$$\left[z_k^{min}; z_k^{max} \right] = z_k^{sim} \pm [\text{critical value}] \times s_{z_k^{sim}} \tag{2.38}$$

For individual linear confidence interval, the critical value is calculated as:

$$[\text{critical value}] = t_S(n, 1.0 + \alpha/2) \tag{2.39}$$

where t_S is the Student-t distribution, α is the significance level (typically 5 or 10 percent which results in 95- or 90-percent intervals, respectively), n is the degrees of freedom equals to $nobs + npr - npar$. The linear confidence intervals are easy to calculate. However, for nonlinear models, nonlinear confidence intervals are often preferable. The method proposed by Vecchia and Cooley (1987) for calculating nonlinear confidence intervals involves estimating a parameter confidence region. This parameter confidence region is composed of the set of parameter values for which the objective function values are only slightly greater than the minimum objective function value obtained via optimisation (Hill and Tiedeman, 2007):

$$\Phi(\mathbf{b}) \leq \Phi(\mathbf{b}') + s^2 \times [\text{critical value}] + a \tag{2.40}$$

where $\Phi(\mathbf{b}')$ is the objective function for the vector of parameter values $\mathbf{b}'$ and a is a term reflecting the accuracy of a measured observed equivalent of the predictions. For individual linear confidence interval, the critical value is calculated as:

$$[\text{critical value}] = c_c \times [t_S(n, 1.0 + \alpha/2)]^2 \tag{2.41}$$

where c_c is a correction factor defined by Christensen and Cooley (2005). Then, the method finds the minimum and maximum values of the prediction on the boundaries of this parameter confidence region. These extreme values are the lower and the upper limits of the nonlinear confidence interval on the prediction. However, estimating the parameter confidence region requires a

lot of model runs which prevents calculating nonlinear confidence intervals for models with long execution times. The same reason prevents from using Monte-Carlo techniques for quantifying prediction uncertainty since they require to run the model under calibration and then prediction conditions for a large number of parameter sets randomly generated.

2.3 Performace criteria and model ranking

The quality of a model is quantified using performance criteria. These performance criteria evaluate the level of agreement between model and reality (Refsgaard and Henriksen, 2004). Typically, they are function of the discrepancies between observed values and their simulated equivalent for a particular type of observations (e.g. flow rates or hydraulic heads).

The most widely used performance criterion for flow rates is the Nash–Sutcliffe efficiency criterion NSE_q (Nash and Sutcliffe, 1970):

$$NSE_q = 1 - \frac{\sum_{t=1}^{nt} \left(q_t^{sim} - q_t^{obs}\right)^2}{\sum_{t=1}^{nt} \left(q_t^{obs} - \mu^{obs}\right)^2} \quad \in \quad]-\infty; 1] \tag{2.42}$$

where nt is the total number of timesteps, q_t^{sim} is the simulated flow rate at timestep t, q_t^{obs} is the observed flow rate at timestep t, and μ^{obs} is the mean of the observed values. If the simulated values match perfectly the observed values, $NSE_q = 1$. The lower the value of NSE_q, the poorer the model, negative values indicating that the mean observed value μ^{obs} gives a better description of the data than the simulated values q_t^{sim}. According to Gupta et al. (2009), this criteria is a convenient and popular indicator of model skill. However, since it uses the observed mean as a baseline, they warn that it can

lead to the overestimation of model skill for highly seasonal variables such as runoff in snowmelt dominated basins.

The Nash-Sutcliffe efficiency criterion can be decomposed into different components for facilitating its interpretation. A decomposition is proposed by Weglarczyk (1998):

$$NSE_q = r_{lin}^2 - \left(r_{lin} - \frac{\sigma^{sim}}{\sigma^{obs}} \right)^2 - \left(\frac{\mu^{sim} - \mu^{obs}}{\sigma^{obs}} \right)^2 \tag{2.43}$$

where r_{lin} is the linear correlation coefficient between q^{sim} and q^{obs}, σ^{sim} is the standard deviation of q^{sim}, μ^{sim} is the mean of q^{sim}, σ^{obs} is the standard deviation of q^{obs}, and μ^{obs} is the mean of q^{obs}. The first component measures the strength of the linear relationships between the observed flow rates and their simulated equivalent, the second component measures the conditional bias, and the third component measures the unconditional bias (Murphy, 1988).

An alternative decomposition is proposed by Gupta et al. (2009):

$$NSE_q = 2 \times \frac{\sigma^{sim}}{\sigma^{obs}} \times r_{lin} - \left(\frac{\sigma^{sim}}{\sigma^{obs}} \right)^2 - \left(\frac{\mu^{sim} - \mu^{obs}}{\sigma^{obs}} \right)^2 \tag{2.44}$$

where the first component uses the linear correlation coefficient for measuring the capacity of the model to reproduce timing and shape of the signal, the second component measures the capacity of the model to reproduce the standard deviation of the observations, and the third component measures the capacity of the model to reproduce the mean of the observations. These decompositions show that the Nash-Sutcliffe efficiency criterion takes implicitly into account the capacity of the model to reproduce the mean flow, the spread of flows, and the timing and shape of the hydrograph. Therefore, it is

a quite complete criterion.

The literature provides a lot of other perfomance criteria for evaluating the capacity of the model to simulate flow rates (for example, see Loague and Green, 1991; Chiew and McMahon, 1994; Yapo et al., 1996; Gupta et al., 1998, 1999; Legates and McCabe, 1999; Beldring, 2002; Vázquez et al., 2002; Henriksen et al., 2003; Smith et al., 2008; Das et al., 2008; Aricò et al., 2009). Among others, the mass balance error, the peak error, and the time peak error.

The mass balance error MBE_q, also known as bias, percent bias or relative bias, is calculated as (Gupta et al., 1999):

$$MBE_q = \frac{\sum_{t=1}^{nt} \left(q_t^{sim} - q_t^{obs} \right)}{\sum_{t=1}^{nt} q_t^{obs}} \times 100 \qquad \in \qquad]-100; +\infty[\tag{2.45}$$

An alternative formulation of this performance criterion is:

$$MBE_q = \frac{\mu^{sim} - \mu^{obs}}{\mu^{obs}} \times 100 \tag{2.46}$$

where μ^{sim} is the mean of the simulated values and μ^{obs} is the mean of the observed values. This performance criterion measures the tendency of the simulated values to be larger or smaller than their observed counterparts. If the fit is perfect, $MBE_q = 0$. If $MBE_q > 0$, simulated values are, on average, greater than observed values, and vice versa.

The peak error PE_q is calculated as (Aricò et al., 2009):

$$PE_q = \left(\frac{h_{peak}^{sim}}{h_{peak}^{obs}} - 1 \right) \times 100 \qquad \in \qquad]-100; +\infty; [\tag{2.47}$$

where h_{peak}^{sim} is the simulated peak value, and h_{peak}^{obs} is the observed peak value. This performance criterion measures the capacity of the model to reproduce the peak in the hydrograph. If the observed peak is equal to the simulated peak, $PE_q = 0$. If $PE_q > 0$, the simulated peak is greater than the observed peak, and vice versa.

The time peak error TPE_q is calculated as (Aricò et al., 2009):

$$TPE_q = t_{peak}^{sim} - t_{peak}^{obs} \qquad \in \qquad]-\infty;+\infty[\tag{2.48}$$

where t_{peak}^{sim} is the simulated peak time value, and t_{peak}^{obs} is the observed peak time value. This performance criterion measures the time lag between the simulated peak and the observed peak. If there is no time lag, $TPE_q = 0$. If $TPE_q > 0$, the simulated peak occurs later than the observed peak.

The most widely used performance criterion for hydraulic heads is the root mean square error criterion RMS_h:

$$RMS_h = \sqrt{\frac{1}{nt} \times \sum_{t=1}^{nt} \left(h_t^{sim} - h_t^{obs}\right)^2} \qquad \in \qquad [0;+\infty[\tag{2.49}$$

where h_t^{sim} is the i^{th} simulated hydraulic head value, and h_t^{obs} is the i^{th} observed hydraulic head value. This performance criterion measures the discrepancies between observed hydraulic heads and their simulated equivalent for a particular piezometer. If the simulated values match perfectly the observed values, $RMS_h = 0$. The greater the value of RMS_h, the poorer the model.

The other criteria typically used for hydraulic heads are the mean absolute error MAE_h:

$$MAE_h = \frac{1}{nt} \times \sum_{t=1}^{nt} \left| h_t^{sim} - h_t^{obs} \right| \quad \in \quad [0; +\infty[\tag{2.50}$$

and the maximum absolute error $MxAE_h$:

$$MxAE_h = \max \left| h_t^{sim} - h_t^{obs} \right| \quad \in \quad [0; +\infty[\tag{2.51}$$

A combined use of these performance criteria allows performing a thorough evaluation of model perfomance with respect to both flow rates and hydraulic heads. A convenient way for performing such a multiple criteria analysis is proposed by Henriksen et al. (2003). They develop an aggregated score system based on performance intervals defined for each criterion used (e.g. if $NSE_q > 0.85$, then the model gets 5 points, if $0.65 < NSE_q < 0.85$, then the model gets 4 points, etc.). However, special care should be taken with multiple performance criteria since some of them are closely related to each other and could lead to contradictory conclusions (Weglarczyk, 1998).

Performance criteria are sometimes used for comparing and ranking models. However, the growing tendency in the scientific community to develop several alternative models for a single study has lead to the development of other criteria specially dedicated to model ranking and model averaging (for example, see Neuman, 2003; Neuman and Wierenga, 2003; Ye et al., 2004; Poeter and Anderson, 2005; Ye et al., 2008; Singh et al., 2010; Ye et al., 2010). These criteria are based on the Kullback–Leibler (K-L) information $I(f, g)$ which is the information I lost when full truth f is approximated by a candidate model g (Kullback and Leibler, 1951). Given a set of candidate models g_i, it is possible to calculate the K-L information for each of the N models and then to select the one that minimises $I(f, g)$ across models. However, for flow and transport models of natural systems, K-L information cannot be calculated since

the truth (e.g. boundary conditions and parameters) is not known (Anderson, 2003). Therefore, the solution consists in considering the change in K-L information between pairs of models. The most widely used criteria for estimating the expected K-L information loss are the Akaike Information Criterion AIC (Akaike, 1973, 1974), the modified Akaike Information Criterion $AICc$ (Sugira, 1978; Hurvich and Tsai, 1989, 1994), the Bayesian Information Criterion BIC (Schwarz, 1978), and the Kashyap Information Criterion KIC (Kashyap, 1982).

The AIC criterion is an estimator of twice the expected K-L information loss (Poeter and Hill, 2007):

$$AIC = nobs \times \ln\left(\sigma^2\right) + 2 \times k \tag{2.52}$$

where $nobs$ is the number of observations, $k = npar + 1$ with $npar$ equals the number of parameters, and s^2 is the residual variance estimated in these equations based on maximum likelihood theory as:

$$\sigma^2 \approx s^2_{ML} = \frac{\Phi(\mathbf{b})}{nobs} \tag{2.53}$$

where $\Phi(\mathbf{b})$ is the objective function of equation 2.11.

The $AICc$ criterion is a correction of AIC required when $nobs/k$ is lower than 40 (Burnham and Anderson, 2002; Poeter and Hill, 2007):

$$AICc = nobs \times \ln\left(\sigma^2\right) + 2 \times k + \left(\frac{2 \times k \times (k+1)}{nobs - k - 1}\right) \tag{2.54}$$

The BIC criterion is calculated as (Poeter and Hill, 2007):

$$BIC = nobs \times \ln\left(\sigma^2\right) + k \times \ln(nobs) \tag{2.55}$$

The KIC criterion is calculated as (Poeter and Hill, 2007):

$$KIC = [nobs - (k-1)] \times \ln\left(\sigma^2\right) - (k-1) \times \ln(2\pi) + \ln\left|\mathbf{J}^T\mathbf{wJ}\right| \quad (2.56)$$

where **J** is the Jacobian matrix of equation 2.2 and **w** is the weight matrix of equation 2.4.

There is a lack of consensus on which criterion to use for model selection. According to Burnham and Anderson (2004), BIC and KIC criteria assume that the true (or quasi-true) model exists in the set of candidate models which is not the case of AIC and $AICc$ criteria. Additionally, Poeter and Anderson (2005) have found that BIC and KIC criteria tend to select models with a lot of parameters regardless of the number of observations. Therefore, they suggest to use in priority AIC and $AICc$ criteria since the theoretical underpinnings of BIC and KIC are philosophically weak. However, Ye et al. (2008) have shown that the justifications of Burnham and Anderson (2004) and Poeter and Anderson (2005) are wrong and that BIC and especially KIC are preferable for model selection. This is supported by others within the scientific community (Carrera and Neuman, 1986a; Neuman, 2003; Neuman and Wierenga, 2003; Ye et al., 2004). The question is still open and most of the publications using K-L information for model selection include each of these criteria (for example, see Ye et al., 2010). Whatever the criterion used, the best model is the one with the smallest criterion value.

The criterion values are often used for calculating the delta values Δ_i of each candidate model i. Using the AIC criterion, these delta values are calculated as:

$$\Delta_i = AIC_i - AIC_{min} \qquad i = 1, N \tag{2.57}$$

where N is the number of candidate models. These delta values represent the K-L information loss of model i relative to the best model in the set. A simple transformation yields posterior model probabilities:

$$p_i = \frac{\exp^{-\Delta_i/2}}{\sum_{j=1}^{N} \exp^{-\Delta_j/2}} \tag{2.58}$$

where p_i is the posterior model probability corresponding to the model candidate i and reflects the evidence in favor of model i being the best model in the sense of minimum K-L information loss. This way of calculating posterior model probabilities is based on information theory. However, it is possible to calculate alternative posterior model probabilities using this general form:

$$p_i = \frac{F_i}{\sum_{j=1}^{N} F_j} \tag{2.59}$$

where F_i is a function of a model criterion. As an example, (Poeter and Hill, 2007) shows a way of calculating posterior model probabilities that increase linearly from the model with the smallest value of the model criterion to the model with the largest value of the model criterion:

$$F_i = 1 + \left(\frac{crit_{min} - valcrit_i}{crit_{max} - crit_{min}} \right) \tag{2.60}$$

where $crit_{min}$ is the smallest value of the chosen criterion, $crit_{max}$ is the largest value of the chosen criterion, and $valcrit_i$ is the criterion calculated for the candidate model i. However, this way of calculating posterior model probabilities does not have any theoretical basis.

Additional statistics have been developed by Poeter and Anderson (2005)

and Poeter and Hill (2007) for facilitating the interpretation of posterior model probablities. These statistics are the evidence ratios and the inverted evidence ratios as a percent. The evidence ratios ER are calculated as:

$$ER = \frac{p_i}{p_j} \tag{2.61}$$

When i is the best model, evidence ratios are used to make statement such as "there is ER times more evidence supporting the best model relative to the second best model".

The inverted evidence ratios as a percent IER are calculated as:

$$IER = \frac{p_j}{p_i} \times 100 \tag{2.62}$$

These statistics are used to make statement such as "the evidence supporting model j is only IER percent of the evidence supporting the best model".

The code MMA (Poeter and Hill, 2007) performs model ranking using classic information criteria (AIC, $AICc$, BIC, and KIC) or others criteria specified by the user. Additionaly, it calculates useful statistics such as the evidence ratios ER and the inverted evidence ratios IER.

Chapter 3

Methodology

The assessment of the influence of some model technical and structure uncertainties on model performance is undertaken using the concept of synthetic case. When working with a synthetic case, the conceptual model, the parameters, and the stress factors are exactly known since the system is entirely generated with a computer program. Furthermore, there is no measurement error on the observations produced. Therefore, with this "perfect knowledge" of the system, it is possible to test specific model features such as the impacts of spatial discretisation on model results. The concept of synthetic case is commonly used in hydrogeology. Among others, this concept has already been used for investigating the importance of combining hard and soft data for reducing the uncertainty associated with groundwater flow and transport predictions (Poeter and McKenna, 1995), for proving the efficiency of nonlinear regression methods in groundwater model calibration (Hill et al., 1998), and for testing the reliability of several monitoring techniques used for the evaluation of natural attenuation, first-order degradation rate, and plume length in contaminated aquifers (Schäfer et al., 2004; Bauer et al., 2006; Beyer et al., 2006, respectively).

The methodology developed in this work involves three main steps (Figure 3.1):

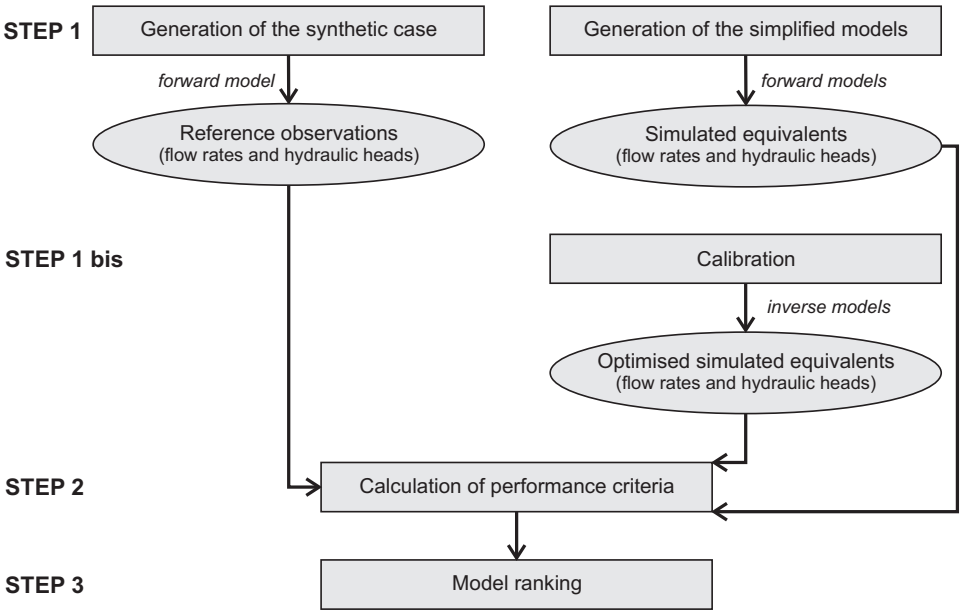

Figure 3.1: The methodology consists in (step 1) generating the synthetic case for producing reference observations and the simplified models for obtaining their simulated equivalent, (step 1 bis) performing optimisation for obtaining optimised simulated equivalent to the reference observations (for models with short execution times only), (step 2) calculating model performance criteria for each simplified model, (step 3) ranking the simplified models depending on their performance.

STEP 1 - *Generation of the synthetic case/Generation of the simplified models*: When choosing the main characteristics of the synthetic case, the objective is to reproduce the key features of typical groundwater bodies of Wallonia (Belgium). This is achieved by assigning to the synthetic case similar elevation, surface materials, and subsurface materials than those known from characterisation studies and hydrogeological maps covering typical groundwater bodies of Wallonia. Additionally, parameter values and stress factors are also specified depending on the values found in such studies. A 5-year reference simulation is then run using refined time and space discretisations

for producing reference observations (flow rates and hydraulic heads). The simplified models are generated in parallel using the same elevation, materials, and parameter values than in the reference simulation. These simplified models differ by their horizontal spatial discretisation (element side length of 250 m, 500 m, 750 m, or 1000 m), their vertical spatial discretisation (8 layers, 6 layers, or 3 layers), and the saturation–pressure relations used in the unsaturated zone (van Genuchten or linear). A forward run with daily stress factors and then a forward run with average monthly stress factors are performed with each simplified model for obtaining simulated values equivalent to the reference observations (flow rates and hydraulic heads). A graphical model fit analysis is performed for qualitatively evaluating model performance.

STEP 1 bis - *Calibration*: The calibration includes a sensitivity analysis, an optimisation, and an uncertainty analysis. However, using a synthetic case, there is no prediction in the sense that the future state of the system is exactly knowable. Consequently, the uncertainty analysis turns into model validation which consists in testing model performance against independent field data (i.e. calculating model perfomance criteria with observations not used during the optimisation) (Refsgaard and Henriksen, 2004; Refsgaard et al., 2006, 2007). This is the subject of the next step. The current step involves only sensitivity analysis and optimisation. The sensitivity analysis is performed to evaluate the influence of spatial discretisation and saturation–pressure relations used in the unsaturated zone on parameter sensitivities. The optimisation is performed for evaluating how far parameter values can compensate for model technical and structure uncertainties. However, this step is only possible for models with short execution times. Therefore, optimisation is only performed for the simplified models with average monthly stress factors. As

for the forward models, a graphical model fit analysis is performed for qualitatively evaluating model performance.

STEP 2 - *Calculation of performance criteria*: The most widely used performance criteria for flow rates and hydraulic heads are calculated for each simplified model. This allows comparing the simplified models and evaluating the impacts of some model technical and structure uncertainties on model performance. The observed values and the simulated values in the performance criterion equations correspond to the reference observations and their simulated equivalent, respectively.

STEP 3 - *Model ranking*: The model ranking is performed based on information theory. Additionally, a ranking based on performance criteria is proposed. The results obtained with both methods are discussed.

All models in this work are developed with the fully-integrated and physically-based surface and subsurface flow model HydroGeoSphere (Therrien et al., 2005). The sensitivity analyses and the optimisations are performed using UCODE_2005 (Poeter et al., 2005) and PEST with AUI (Doherty, 2005), respectively. The model ranking using information criteria is performed using MMA (Poeter and Hill, 2007).

The simulations performed in this work are run on NIC3, the super computer for intensive calculation of the University of Liege, whose calculation blades Dell PEM600 are equipped with 2 quadcore processors Intel Xeon L5420 2.50Ghz with 16 or 32 GB RAM.

Chapter 4 presents the governing equations and the main outputs of HydroGeoSphere. Chapter 5 presents the synthetic case and the 5-year refer-

ence simulation. The simplified models are presented in Chapter 6. Forward models, sensitivity analyses, and inverse models performed for the simplified models are also presented in this chapter. Chapter 7 discusses the values of performance criteria calculated for each simplified model. Chapter 8 discusses the model rankings obtained using the information theory and the performance criteria.

Chapter 4

Governing equations and main outputs of HydroGeoSphere

4.1 Governing equations of HydroGeoSphere

4.1.1 Groundwater flow

Three-dimensional groundwater flow in both the saturated and the unsaturated zone is represented using the Richard's equation:

$$-\nabla \cdot \mathbf{q} + \sum \Gamma \pm Q = S_w \times S_S \times \frac{\partial \psi}{\partial t} + \frac{\partial (\phi \times S_w)}{\partial t} \qquad (4.1)$$

where Γ_{ex} is the volumetric fluid exchange rate between the subsurface domain and the other domains within the model [$L^3L^{-3}T^{-1}$], Q is a source/sink term [$L^3L^{-3}T^{-1}$], S_S is the specific storage [L^{-1}], ψ is the pressure head [L], ϕ is the porosity [-], and S_w is the water saturation [-]. The fluid flux $\mathbf{q}$ [LT^{-1}] is calculated as:

$$\mathbf{q} = -\mathbf{K} \cdot k_r \times \nabla h \tag{4.2}$$

where $\mathbf{K}$ is the hydraulic conductivity tensor [LT^{-1}], k_r is the relative permeability of the medium [-], and $h = \psi + z$ is the hydraulic head [L], and z is the elevation head [L]. The relations linking water saturation, pressure head, and relative permeability are most often expressed using the model of van Genuchten (1980). Given that the water saturation S_w [-] is calculated as:

$$S_w = \frac{\theta}{\theta_s} \tag{4.3}$$

where θ is the water content [-] and given that the effective saturation S_e [-] is calculated as:

$$S_e = \frac{S_w - S_{wr}}{1 - S_{wr}} \tag{4.4}$$

where S_{wr} is the residual water saturation [-], the saturation–pressure relation proposed by van Genuchten (1980) is written:

$$\begin{cases} S_w = S_{wr} + (1 - S_{wr}) \left[1 + |\alpha_{vG} \times \psi|^{\beta_{vG}}\right]^{-\gamma_{vG}} & \text{for} \quad \psi < 0 \\ S_w = 1 & \text{for} \quad \psi \geq 0 \end{cases} \tag{4.5}$$

with the relative permeability k_r given by:

$$k_r = S_e^{1/2} \left[1 - \left(1 - S_e^{1/\gamma_{vG}}\right)^{\gamma_{vG}}\right]^2 \tag{4.6}$$

where α_{vG} [L^{-1}], β_{vG} [-], and γ_{vG} [-] are the van Genuchten parameters.

4.1.2 Surface water flow

Two-dimensional surface water flow is represented using the two-dimensional depth-averaged diffusion-wave approximation to the Saint-Venant equation:

$$-\nabla \cdot (d_s \times \mathbf{q}_s) - d_s \times \Gamma_s \pm Q_s = \frac{\partial (\phi_s \times h_s)}{\partial t} \tag{4.7}$$

where d_s is the depth of the surface water flow [L], $h_s = d_s + z_s$ is the water surface elevation [L] and z_s is the river bed elevation [L], Γ_s is the volumetric fluid exchange rate between the surface domain and the other domains within the model [$L^3L^{-3}T^{-1}$], ϕ_s is the surface water domain "porosity" [-] which varies from unity in a river to zero over rills, Q_s is a source/sink term [L^3T^{-1}]. The fluid flux $\mathbf{q}_s$ [LT^{-1}] is calculated as:

$$\mathbf{q}_s = -\mathbf{K}_s \cdot k_{rs} \times \nabla (d_s + z_s) \tag{4.8}$$

where k_{rs} is the relative permeability of the surface water domain. The conductivity $\mathbf{K}_s$ is derived using Manning's formula:

$$\begin{cases} K_{s_{xx}} = \frac{d_s^{2/3}}{n_{xx}} \times \frac{1}{(\partial h_s / \partial s)^{1/2}} \\ K_{s_{yy}} = \frac{d_s^{2/3}}{n_{yy}} \times \frac{1}{(\partial h_s / \partial s)^{1/2}} \\ K_{s_{xy}} = K_{s_{yx}} = 0 \end{cases} \tag{4.9}$$

where s is the length along the direction of maximum slope [L], and n_{xx} and n_{yy} are the Manning roughness coefficients [$L^{-1/3}T$].

When using the dual node approach, the interactions between groundwater and surface water are simulated using a physically-based exchange flux relation:

$$d_s \times \Gamma_s = \frac{k_r \times \mathbf{K}_{zz}}{L_c} \times (h - h_s) \tag{4.10}$$

where $\mathbf{K}_{zz}$ is the vertical saturated hydraulic conductivity of the underlying porous medium and L_c is a coupling length [L]. When water flows from subsurface to surface, k_r is the relative permeability of the porous medium. When water flows from surface to subsurface, it is given by the ratio of the water depth in the surface to the total obstruction height H_s [L] such that:

$$k_r = \begin{cases} S_{exch}^{2\times(1-S_{exch})} & \text{when } d_s < H_s \\ 1 & \text{when } d_s > H_s \end{cases} \tag{4.11}$$

with S_{exch} given by:

$$S_{exch} = \frac{d_s}{H_s} \tag{4.12}$$

4.1.3 Evapotranspiration and interception

The processes of interception and evapotranspiration are simulated using physically-based equations taking into account potential evapotranspiration measurements, plant characteristics, and soil moisture content in the unsaturated zone obtained using the subsurface component of the model.

Actual evapotranspiration is calculated using a combination of plant transpiration T_p and evaporation from the surface and subsurface domain E_s. Plant transpiration T_p [LT^{-1}] is estimated using the model of Kristensen and Jensen (1975):

$$T_p = f_1(LAI) \times f_2(\theta) \times RDF(L_r) \times \left(E_p - E_{can}\right) \tag{4.13}$$

where E_p is the potential evapotranspiration [LT^{-1}] and E_{can} is the canopy

evaporation of water stored via interception [LT^{-1}].

The vegetation term $f_1(LAI)$ [-] is a function of the leaf area index LAI calculated as:

$$f_1(LAI) = \max[0, \min(1, C_2 + C_1 \times LAI)] \tag{4.14}$$

The function f_1 relates the plant transpiration rate in a linear manner to the leaf area index LAI.

The water content term $f_2(\theta)$ [-] is calculated as:

$$f_2(\theta) = \begin{cases} 0 & \text{for} \quad 0 \le \theta \le \theta_{t2} \\ 1 - \left(\frac{\theta_{t1} - \theta}{\theta_{t1} - \theta_{t2}}\right)^{C_3} & \text{for} \quad \theta_{t2} \le \theta \le \theta_{t1} \\ 1 & \text{for} \quad \theta_{t1} \le \theta \end{cases} \tag{4.15}$$

where C_1 [-], C_2 [-], and C_3 [-] are fitting parameters, and θ_{t1} [-] and θ_{t2} are the transpiration limiting saturations. The function f_2 relates plant transpiration to the water content of the subsurface surrounding the root zone. Full transpiration can occur when the water content is greater than θ_{t1}. Conversely, no transpiration can occur when water content is lower than θ_{t2}. When water content is greater than θ_{t2} and lower than θ_{t1}, transpiration decreases following a law governed by C_3.

The role of the root distribution function $RDF(L_r)$ [-] is to distribute the water extracted from the root zone over the entire root depth L_r following a quadratic law. The quantity of water extracted is greater near the ground surface where the density of roots is greatest. The function f_2 and RDF couple the plant transpiration rate to the water content θ in a nonlinear manner.

When plant transpiration and canopy evaporation have not met the potential evapotranspiration, extra evaporation of water from the surface and subsurface domains occur. This extra evaporation occurs along with transpiration

due to extraterrestrial radiation energy penetrating the vegetation cover and evaporation of water from the soil surface and subsurface soil layers. This process is modelled as:

$$E_s = \alpha^* \times \left(E_p - E_{can}\right) \times \left[1 - f_1\left(LAI\right)\right] \times EDF\left(L_e\right) \tag{4.16}$$

with the wetness factor α^* given by:

$$\alpha^* = \begin{cases} 0 & \text{for} \quad \theta < \theta_{e2} \\ \frac{\theta - \theta_{e2}}{\theta_{e1} - \theta_{e2}} & \text{for} \quad \theta_{e2} \leq \theta \leq \theta_{e1} \\ 1 & \text{for} \quad \theta > \theta_{e1} \end{cases} \tag{4.17}$$

where L_e is the evaporation depth [L] and θ_{e1} [-] and θ_{e2} [-] are the evaporation limiting saturations. Full evaporation can occur when the water content is greater than θ_{e1}. Conversely, no evaporation can occur when water content is lower than θ_{e2}. When water content is greater than θ_{t2} and lower than θ_{t1}, transpiration decreases following a law governed by C_3.

The role of the evaporation distribution function $EDF\left(L_e\right)$ [-] is to distribute the water extracted from the evaporative zone over the entire evaporation depth L_e following a quadratic law.

The interception of water by the canopy is simulated with the bucket model where precipitations in excess of interception storage and evapotranspiration reach the ground surface. The interception storage capacity S_{int}^{max} is calculated as:

$$S_{int}^{max} = C_{int} \times LAI \tag{4.18}$$

where C_{int} is the canopy storage parameter [L]. The interception storage capacity is the maximum quantity of water that the canopy can intercept.

4.2 Main outputs of HydroGeoSphere

The control volume finite element method is used to solve simultaneously the equations describing surface water flow and groundwater flow. At each time step, HydroGeosphere calculates simultaneously water depth and fluid flux at each node of the surface domain, and hydraulic head, saturation, and Darcy flux at each node of the subsurface domain. A thorough description of HydroGeosphere including verification examples is proposed by Therrien et al. (2005).

Chapter 5

Generation of the synthetic case

5.1 Groundwater body template for the synthetic case

The synthetic case is inspired by a typical groundwater body of Wallonia (Belgium). This groundwater body, consisting in Famennian sandstone and Carboniferous limestone aquifers, is known as RWM021[1] (Figure 5.1). The objective is to reproduce its hydrodynamic functioning by taking into account its key features. This is performed using the characterisation studies of this groundwater body proposed by Brouyère et al. (2009).

[1] *Calcaires et grès du Condroz*

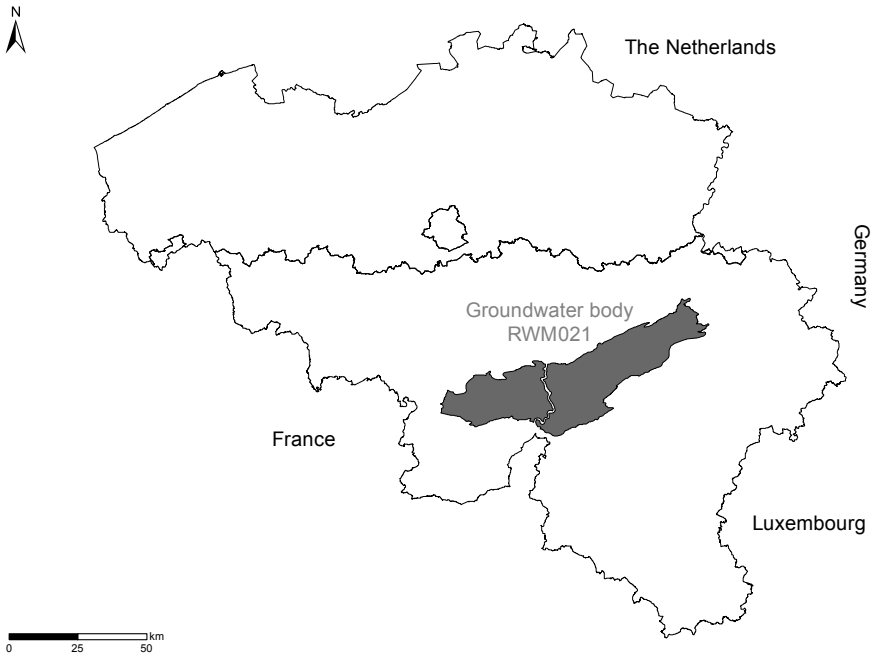

Figure 5.1: Groundwater body RWM021 is located in the Condroz region of Wallonia (Belgium).

5.2 Key features of groundwater body RWM021

5.2.1 Geology

Groundwater body RWM021 is located in the Condroz region. This region is characterised by a succession of limestone synclines and sandstone anticlines oriented West-East to South-West-North-East. This structure is due to the Variscan orogeny and to the differential erosion of limestones and sandstones. Limestones, easily eroded, are found in the centre of the valleys, while sandstones, hardly eroded, outcrop on the crests. The main rivers are located in the centre of the limestone synclines and are fed by tributaries flowing from the sandstone crests.

5.2.2 Hydrogeology

5.2.2.1 Aquifers and aquicludes

Geological formations are grouped together into hydrogeological units depending on their hydrodynamic properties (Figure 5.2). The main hydrogeological units of groundwater body RMW021 are (from the oldest to the youngest geological formations):

- Famennian sandstone aquifers,
- Hastarian aquifer with aquiclude formations,
- Carboniferous limestone aquifer,
- Upper Carboniferous aquiclude with aquifer levels.

Groundwater is essentially withdrawn from Hastarian and Carboniferous limestones whose hydrogeological potential is closely related to their fissuration. When altered and fissured, Famennian sandstones can also have a relatively high hydrogeological potential. This is not the case of Upper Carboniferous shales whose hydrogeological potential is poor due to low hydraulic conductivity values.

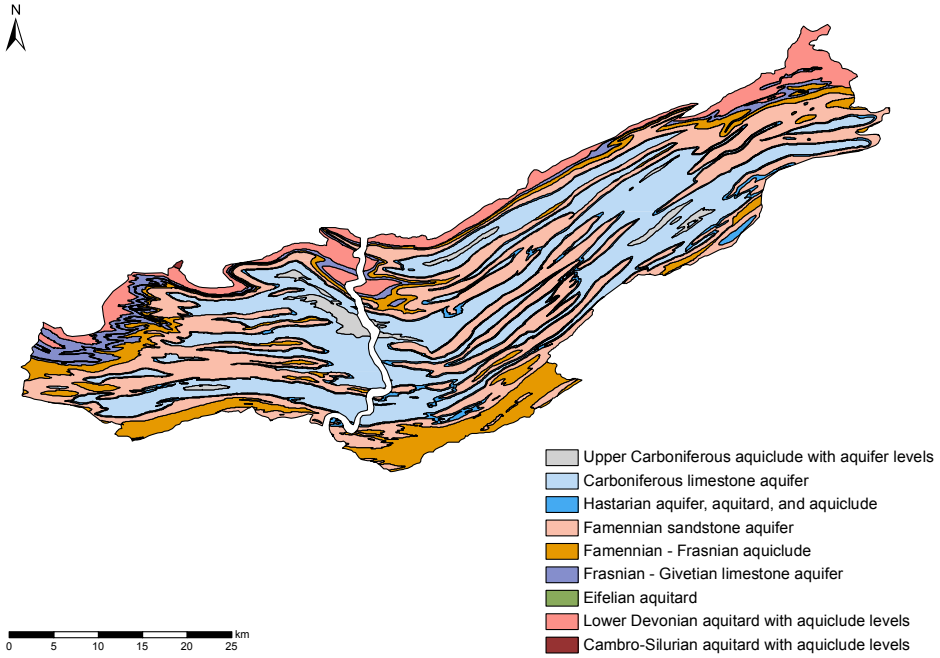

Figure 5.2: Grounwater body RWM021 is characterised by a succession of limestone synclines and sandstone anticlines typical of the Condroz region. When fissured, limestones are intensively exploited for groundwater.

5.2.2.2 Groundwater balance

Annual precipitations measured in meteorological stations located in groundwater body RWM021 from 1970 to 2008 range from 550 mm to 1,050 mm. According to the groundwater balances calculated for 15 hydrogeological catchments of this groundwater body, evapotranspiration, infiltration, and runoff represent 60 %, 10 % to 25 %, and 15 % to 30 % of precipitations, respectively. Aquifer recharge and surface runoff occur almost exclusively from October to April. Baseflow recession is usually observed from May to September.

5.2.2.3 Hydraulic head variations

According to hydraulic head measurements performed regularly in 141 piezometers located in groundwater body RMW021, seasonal variations of the piezometry are observed. The largest variations occur in piezometers located far from the gaining rivers that constrain the hydraulic heads in their neighbourhood. Variations are most often limited to a few meters but they can reach locally up to 10 m in the Famennian sandstones and up to 25 m in the Carboniferous limestones.

5.2.2.4 Hydrodynamic parameters

According to pumping tests performed in 24 piezometers located in groundwater body RWM021, hydraulic conductivity of the Famennian sandstones range from 7.40×10^{-7} m/s to 4.80×10^{-4} m/s with a geometric mean of 2.29×10^{-5} m/s. Concerning Carboniferous limestones, hydraulic conductivity estimated from pumping tests in 100 piezometers of RWM021 range from 4.00×10^{-10} m/s to 2.55×10^{-3} m/s with a geometric mean of 1.74×10^{-5} m/s. This very large range of hydraulic conductivity for the limestones is typical of fissured and karstified hard rocks.

5.2.2.5 Typical hydrodynamic functioning

Groundwater flow in the Famennian sandstone anticlines is mostly topography-driven. Groundwater flow in the Carboniferous limestone synclines is more complex and influenced by the topography, the orientation of the geological layers, and preferential flowpaths in fissure networks. Limestones are fed by infiltration through the loess formations covering the entire zone. Additionally, they can be fed laterally by groundwater flowing through the sandstones. Lines of springs occur in the hillslopes at the contact between

sandstones and low permeable horizons underlying the limestones. These springs feed a lot of streams, with possible losing portions, that constitute the tributaries of a gaining river located in the centre of the synclines (Figure 5.3).

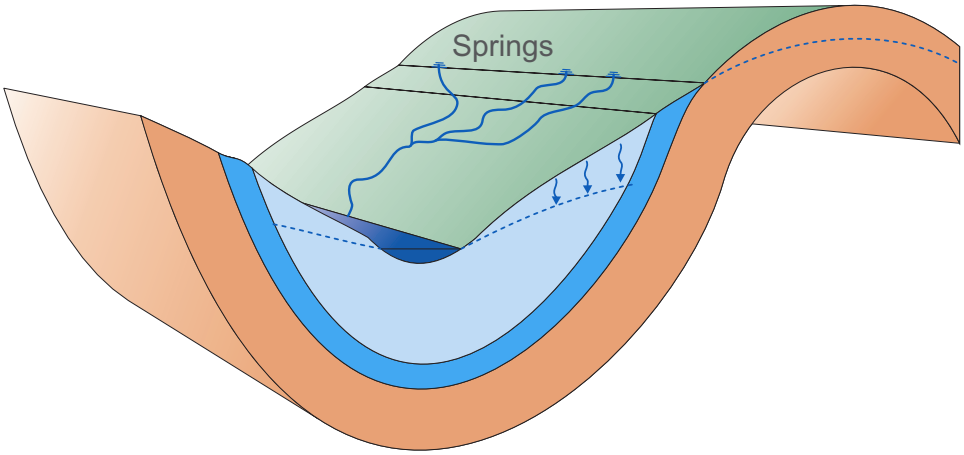

Figure 5.3: While groundwater flow in the Famennian sandstones is mostly topography-driven, groundwater flow in the Carboniferous limestones is influenced by the topography, the orientation of the geological layers, and preferential flowpaths in fissure networks.

5.3 Key features of the reference model and reference simulation

5.3.1 Elevation

The objective is to generate a hydrological catchment with a syncline structure typical of groundwater body RWM021. This is performed using ordinary kriging with SGeMs (Remy et al., 2009). The set of experimental points is obtained by randomly extracting 50,000 elevation points from the Digital Elevation Model (DEM) of RWM021. The omnidirectional experimental variogram is calculated using 25 lags with a nominal lag spacing of 500 m and a lag tolerance of 100 m. At the scale of groundwater body RWM021, the directional experimental variograms at 0, 45, 90, and 135 degrees do not indicate any significant anistropy, so the omnidirectional experimental variogram is used.

This omnidirectional experimental variogram is then fitted using an exponential model without nugget effect. This exponential model is in turn used for kriging elevation data using a set of constraining points on the crests and in the valley, and a circular search ellipsoid with a radius of 15,000 m, requiring a minimum of 5 data points to be found within 15,000 m of each grid point and using no more than 15 nearest neighbouring points in the estimation for each grid point. The elevation grid obtained (30,000 m by 15,000 m with 30 m by 30 m cells) is used for delineating the hydrological catchment (364 km²) constituting the reference model (Figure 5.4).

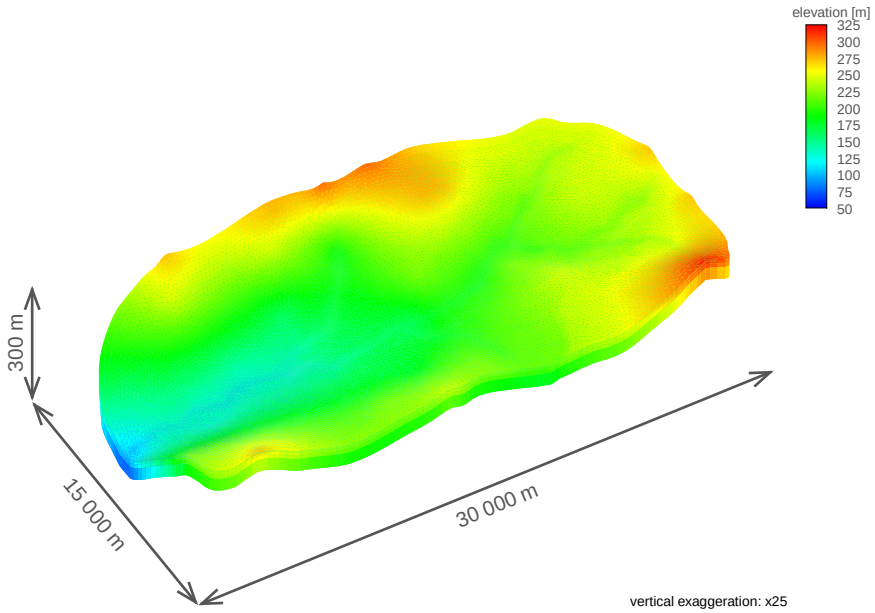

Figure 5.4: The elevation of the reference model is simulated using ordinary kriging with an omnidirectional experimental variogram derived from the DEM of groundwater body RWM021 and an exponential model. The hydrological catchment delineated from the kriged elevation data covers 364 km².

5.3.2 Materials

5.3.2.1 Surface materials

According to a study outlining the land use in Wallonia (Grandjean et al., 2006), groundwater body RWM021 are characterised by rural (crops and grasslands) and forest land use. Crops and grasslands are found preferably in low elevation zones (< 300 m) with gentle slopes (< 7 %) while forests are rather located in high elevation zones (> 300 m) with steep slopes (> 7 %). Typically, urbanised zones are found in very low elevation zones (< 200 m). A similar criterion depending on elevation and slope constraints is considered for defining the surface materials of the reference model (Figure 5.5).

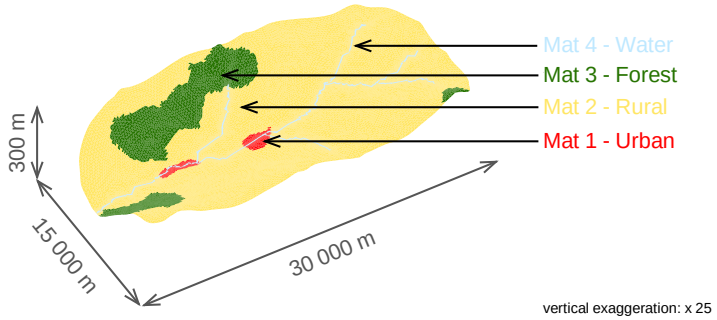

Figure 5.5: The surface materials of the reference model are assigned depending on elevation and slope constraints: Mat 1 - Urban in very low elevation zones, Mat 2 - Rural in low elevation zones with gentle slopes, Mat 3 - Forest in high elevation zones with steep slopes, and Mat 4 - Water in the valleys.

5.3.2.2 Subsurface materials

According to the geology of groundwater body RWM021, Famennian sandstones, Hastarian limestones, Carboniferous limestones, and Upper Carboniferous shales constitute, from the crests to the centre of the valley, the sub-

surface materials of the reference model (Figure 5.6). Additionally, these formations are covered by alluvial deposits and loess.

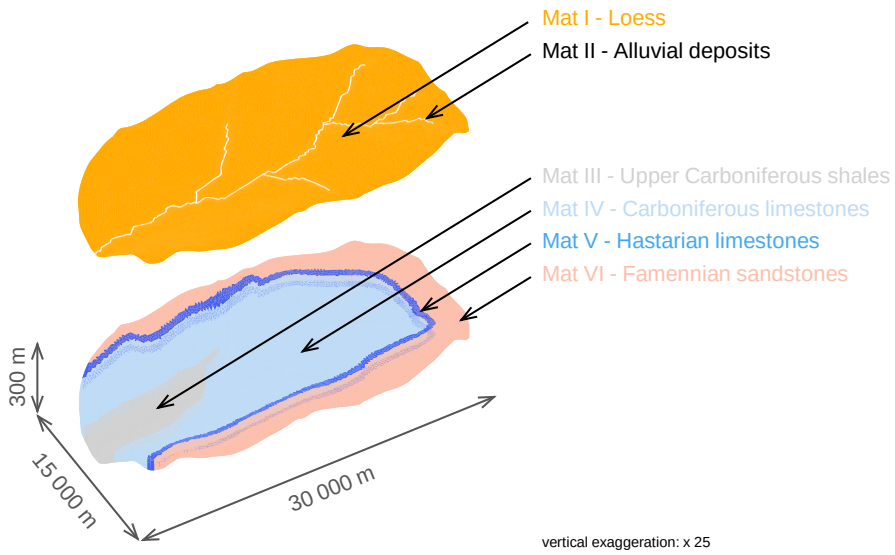

Figure 5.6: The subsurface materials of the reference model are assigned following the typical syncline structure of groundwater body RWM021: Mat IV - Carboniferous limestones is found in the centre of the valley and Mat VI - Famennian sandstones is found on the crests.

5.3.3 Parameters

5.3.3.1 Surface parameters

All surface materials involved in the interception and evapotranspiration processes are assigned a series of interception and evapotranspiration parameters listed in Table 5.1.

Table 5.1: List of interception and evapotranspiration parameters required by HydroGeoSphere.

Parameter	Symbol	Unit
Leaf area index	LAI	[-]
Root depth	L_r	[L]
Evaporation depth	L_e	[L]
Evaporation & transpiration limiting saturations	θ_{e2} & θ_{t2}	[-]
	θ_{e1} & θ_{t1}	[-]
Transpiration fitting parameters	C_1	[-]
	C_2	[-]
	C_3	[-]
Canopy storage parameter	C_{int}	[L]
Initial interception storage	S^0_{int}	[L]

Appropriate values for leaf area index and root depth are available in the global syntheses for the major species and terrestrial biomes proposed by Asner et al. (2003) and Canadell et al. (1996), respectively. Typical values for the other parameters can be found in publications about evapotranspiration models (Kristensen and Jensen, 1975; Dickinson et al., 1991) and hydrological models including evapotranspiration processes (Schroeder et al., 1994; Andersen et al., 2002; Graham and Kilde, 2002; Vázquez et al., 2002; Vázquez and Feyen, 2003; Islam, 2004; Panday and Huyakorn, 2004; Therrien et al., 2005; Li et al., 2008; Goderniaux, 2010). The set of evapotranspiration parameter values extracted from these sources for use in the reference simulation is given in Table 5.2.

Table 5.2: Values of interception and evapotranspiration parameters used in the reference simulation.

	Mat 1	Mat 2	Mat 3
LAI [-]	0.40	3.53	5.12
L_r [m]	0	2.30	2.90
L_e [m]		2.00	
θ_{e2} & θ_{t2} [-]		0.60	
θ_{e1} & θ_{t1} [-]		0.96	
C_1 [-]		0.31	
C_2 [-]		0.15	
C_3 [-]		10.00	
C_{int} [m]		5.00×10^{-5}	
S^0_{int} [m]		0	

The LAI values presented in Table 5.2 are reduced by 50 % for fall and winter months since the cover of leaves is lower during these seasons.

A series of surface flow parameters, listed in Table 5.3, is also prescribed for each surface material.

Table 5.3: List of surface flow parameters required by HydroGeoSphere.

Parameter	Symbol	Unit
Manning roughness coefficients	n_{xx} & n_{yy}	[$L^{-1/3}T$]
Rill storage height	H_{sto}	[L]
Coupling length	L_c	[L]

Typical values of Manning roughness coefficients for common land use categories are presented in McCuen (1989), Hornberger et al. (1998), Fetter (2001), and Brutsaert (2005). As the values proposed by McCuen (1989) were used successfully in complex hydrological models (Jones, 2005; Li et al., 2008), they are also used in the reference model. The set of surface flow parameter values used in the reference simulation is given in Table 5.4.

Table 5.4: Values of surface flow parameters used in the reference simulation.

	Mat 1	Mat 2	Mat 3	Mat 4
n_{xx} & n_{yy} [m$^{-1/3}$s]	0.012	0.200	0.600	0.025
H_{sto} [m]	0.002	0.002	0.002	0.002
L_c [m]	1.00×10^{-1}			

5.3.3.2 Subsurface parameters

All subsurface materials are assigned a series of subsurface flow parameters listed in Table 5.5.

Table 5.5: List of subsurface flow parameters required by HydroGeoSphere.

Parameter	Symbol	Unit
Saturated hydraulic conductivity	K	[LT^{-1}]
Specific storage	S_S	[L^{-1}]
Saturated water content	θ_s	[-]
Residual water saturation	S_{wr}	[-]
van Genuchten parameters	α_{vG}	[L^{-1}]
	β_{vG}	[-]
	γ_{vG}	[-]

The covering sediment materials, namely Mat I - Loess and Mat II - Alluvial deposits, are assigned typical parameter values of sediment materials found in Jones (2005) who had extracted them from Freeze and Cherry (1979) and Radcliffe (2000). Parameter values for the fissured hard rock materials, namely Mat III - Upper Carboniferous shales, Mat IV - Carboniferous limestones, Mat V - Hastarian limestones, and Mat VI - Famennian sandstones, come from several sources. Saturated hydraulic conductivity values are extracted from the characterisation studies of groundwater body RWM021 in the context of which pumping tests were performed (Brouyère et al., 2009) . Van Genuchten parameter values for shales and limestones are taken from studies

focusing on the unsaturated rock mechanics of an Upper Carboniferous shale extracted from a coal mine in Beringen (Belgium) (Ramos da Silva et al., 2008) and on the leaching of a contaminant through an unsaturated fractured limestone in Brévilles (France) (Roulier et al., 2006), respectively. Van Genuchten parameter values for the fissured limestones are also used for the fissured sandstones due to the lack of studies about the unsaturated properties of this type of rocks. The set of subsurface flow parameter values used in the reference model are given in Table 5.6 for the sediment materials, and in Table 5.7 for the fissured hard rock materials.

Table 5.6: Values of subsurface flow parameters for the covering sediment materials used in the reference simulation.

	Mat I	Mat II
K [m/s]	5.00×10^{-7}	1.00×10^{-6}
S_S [1/m]	1.00×10^{-4}	1.00×10^{-4}
θ_s [-]	4.10×10^{-1}	4.10×10^{-1}
S_{wr} [-]	9.76×10^{-2}	
α_{vG} [m^{-1}]	2.67×10^{0}	
β_{vG}	1.45×10^{0}	
γ_{vG} [-]	$1 - \frac{1}{\beta_{vG}}$	

Table 5.7: Values of subsurface flow parameters for the fissured hard rock materials used in the reference simulation.

	Mat III	Mat IV	Mat V	Mat VI
K [m/s]	1.00×10^{-5}	1.00×10^{-4}	2.50×10^{-4}	5.00×10^{-5}
S_S [1/m]	1.00×10^{-4}	1.00×10^{-4}	1.00×10^{-4}	1.00×10^{-4}
θ_s [-]	2.50×10^{-2}	1.00×10^{-1}	1.00×10^{-1}	7.50×10^{-2}
S_{wr} [-]		0		
α_{vG} [m^{-1}]	6.08×10^{-12}		3.65×10^{-2}	
β_{vG} [-]	0.62		1.83	
γ_{vG} [-]	38671.00		$1 - \frac{1}{\beta_{vG}}$	

5.3.4 Observation points and stress factors

A set of observation points, including one gauging station (G1) and twelve piezometers (Pz1 to Pz12), is used to follow the time evolution of flow rates and hydraulic heads in the synthetic case (Figure 5.7). The gauging station is located at the outlet of the catchment. The piezometers are located evenly over the synthetic case.

Precipitations, potential evapotranspiration, and groundwater withdrawals, constitute the stress factors of the reference model. Precipitations and potential evapotranspiration measured daily in a meteorological station close to groundwater body RWM021 are prescribed on faces using specified fluxes. These fluxes are distributed nodally using the contributing area normal to the face. Groundwater is withdrawn using two galleries (GAL1 and GAL2) and four wells (W1 to W4) (Figure 5.7). Galleries GAL1 and GAL2, of 320 m and 180 m long respectively, are located in the Carboniferous limestones. The induced drainage is simulated using drain nodes with a conductance of 1.00×10^{-3} m²/s. Well W1 is located in the Carboniferous limestones. Wells W2 and W3 are both located in the Hastarian limestones. Well W4 is located in the Famennian sandstones. They are simulated in prescribing volumetric flow rate at the corresponding node. These volumetric flow rates are 1000 m³/h for W1, W2, and W3, and 500 m³/h for W4.

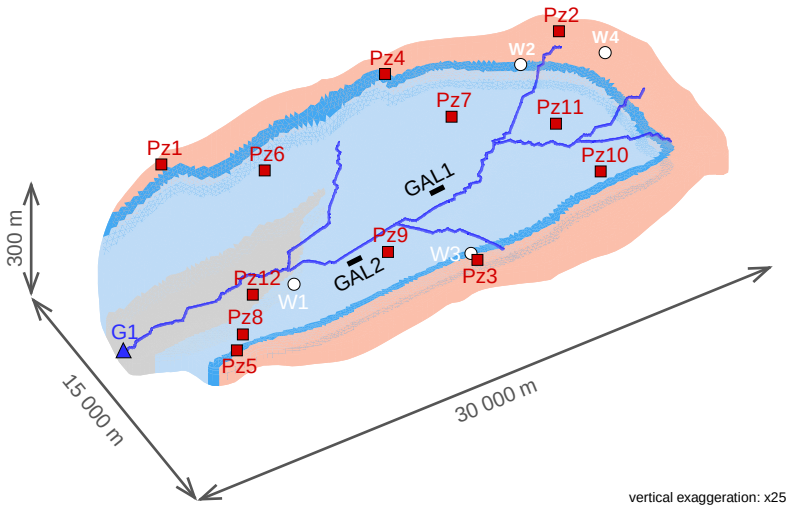

Figure 5.7: A gauging station (G1) and twelve piezometers (Pz1 to Pz12) are used to obtain reference observations in terms of flow rates and hydraulic heads, respectively. Additionally, two galleries (GAL1 and GAL2) and four wells (W1 to W4) are used to simulate groundwater withdrawals.

5.3.5 5-year reference simulation

A 5-year integrated surface and subsurface flow simulation with daily stress factors is performed with the reference model for producing reference flow rate and hydraulic head observations. This reference simulation is performed using a refined mesh with 8 layers, 153,027 nodes, and 269,872 elements (Figure 5.8). The element side length increases progressively from 25 m near the surface water network to 250 m far from the surface water network. The layer thickness increases progressively from 1 m for the top layers corresponding to the unsaturated zone to 30 m for the bottom layers corresponding to the saturated zone (i.e. 5 layers of 1 m, 1 layer of 5 m, 1 layer of 10 m, and 1 layer of 30 m). These horizontal and vertical mesh refinements are performed for simulating accurately surface water–groundwater interactions as

well as recharge processes at the interface between surface and subsurface domains.

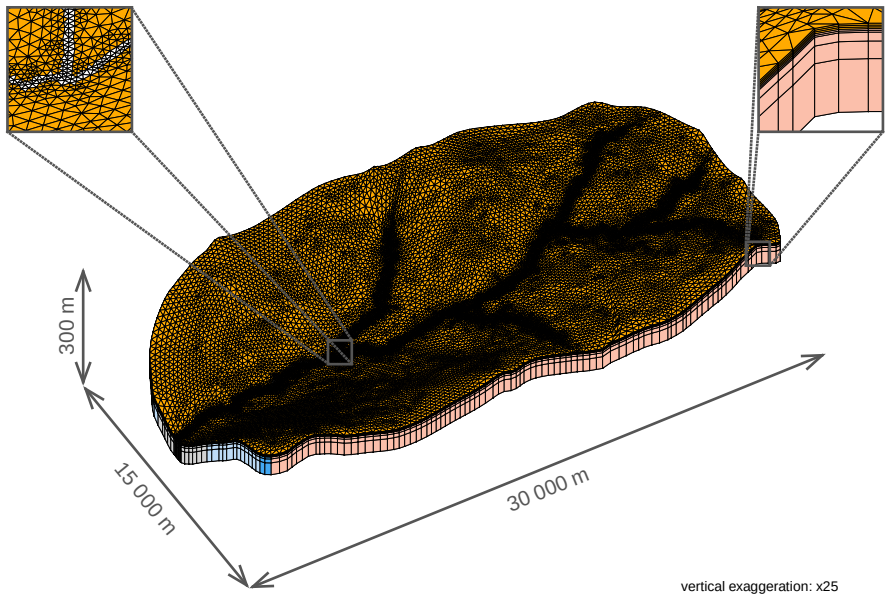

Figure 5.8: The mesh for the 5-year reference simulation is refined horizontally (element side length from 25 m to 250 m) and vertically (layer thickness from 1 m to 30 m). The total number of nodes is 153,027.

Critical-depth boundary conditions are assigned to boundary nodes of the surface domain. This is quite usual for this kind of model (for example, see Li et al., 2008; Goderniaux et al., 2009). This type of boundary condition forces the water elevation at the boundary to be equal to the water elevation for which the energy of the flowing water relatively to the stream bottom is minimum (Therrien et al., 2005). No-flow boundary conditions are applied to boundary nodes of the subsurface domain. Water depth and hydraulic head extracted from preliminary tests performed with the reference model are used as initial conditions for the surface domain and the subsurface domain, respectively.

Flow rates and hydraulic heads produced each day of the 5-year reference

simulation constitute the set of reference observations to calculate performance criteria for the simplified models with daily stress factors (Figure 5.9 and Figure A.1 to Figure A.2 in Appendix A). The set of reference observations for the simplified models with average monthly stress factors is obtained by monthly averaging the daily observations (arithmetic mean). This procedure is required for ensuring time-consistency, that is working with observations that are consistent with the simulated transient processes (Hill and Tiedeman, 2007).

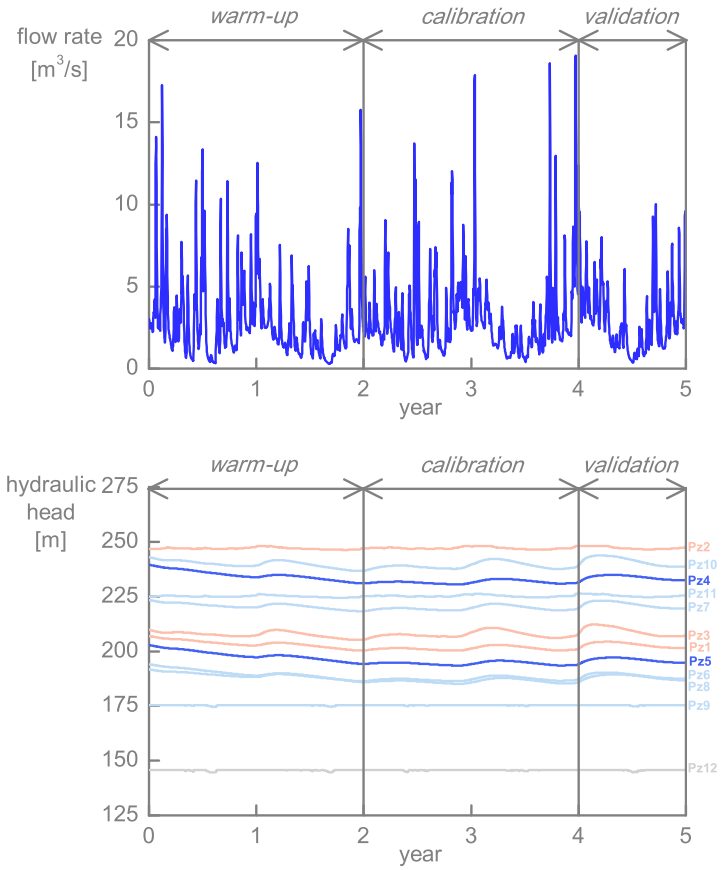

Figure 5.9: A set of transient reference flow rate and hydraulic head observations is produced with the 5-year reference simulation.

The reference simulation is subdivided into warm-up, calibration, and val-

idation periods. The warm-up period is necessary for obtaining simulated values independent of the initial conditions. Flow rates and hydraulic heads produced during this warm-up period are not included in the set of reference observations. Perfomance criteria are calculated only for flow rates and hydraulic heads produced during calibration and validation periods.

The average water balance calculated for the reference simulation (warm-up period excluded) is presented in Table 5.8. The values obtained are consistent with those obtained for several hydrogeological catchments of groundwater body RWM021 (see Section 5.2.2.2). The hydraulic head variations simulated with the reference simulation range between 0.75 m and 6.97 m for the piezometers located in the Carboniferous limestones and between 1.76 m and 6.92 m for piezometers located in the Famennian sandstones. The smallest variations occur in piezometers close to the gaining river located in the centre of the synthetic case. The highest variations occur in piezometers far from the gaining river. This is consistent with the hydraulic head variations observed in piezometers of groundwater body RWM021 (see Section 5.2.2.3). Therefore, with regards to water balance and hydraulic head variations, the synthetic case fairly reproduces the typical hydrodynamic functioning of groundwater body RWM021.

Table 5.8: Average water balance calculated for the reference simulation (warm-up period excluded).

	Flux [mm/yr]	Flux [% precipitation]
Precipitations	644.4	100.0
Actual evapotranspiration	-336.4	-52.2
Outlet	-270.7	-42.0
Groundwater withdrawals	-18.0	-2.8
Error	-19.3	-3.0

Chapter 6

Generation and qualitative evaluation of the simplified models

6.1 Key features of the simplified models

A series of simplified models is generated for evaluating the impacts of some model technical and structure uncertainties on model performance. The models developed for testing technical uncertainties focus on horizontal (element size) and vertical (number of layers) spatial discretisations. The models developed for testing structure uncertainties focus on the representation of the unsaturated zone (saturation–pressure relations). Additionally, each simplified model is executed with daily and average monthly stress factors (precipitations and potential evapotranspiration) for testing the influence of time resolution on model results. The average monthly stress factors are obtained by monthly averaging the daily stress factors. The simplified models and their key characteristics are listed in Table 6.1.

Table 6.1: List of the simplified models with their key characteristics.

	Element size	Number of layers	Saturation–pressure relation
250 m	250 m	8	van Genuchten
500 m	500 m	8	van Genuchten
750 m	750 m	8	van Genuchten
1000 m	1000 m	8	van Genuchten
6 layers	250 m	6	van Genuchten
3 layers	250 m	3	van Genuchten
linear	250 m	8	linear

A forward model is executed with each simplified model using the same elevation, materials, boundary conditions and parameter values than in the reference model (see Section 5.3). A sensitivity analysis is then performed on the calibration period for evaluating whether model technical and structure uncertainties modify parameter sensitivities. As reference observations are produced by a synthetic case, there is no measurement error. Therefore, heavy weights are specified to reference observations. These weights are calculated using coefficient of variation values of 5×10^{-3} and 1×10^{-3} for flow rates and hydraulic heads, respectively (Equation 2.5). Additionally, an inverse model is executed on the calibration period to see how parameters can possibly compensate for model technical and structure uncertainties. However, for execution time issues, optimisation is only performed for simplified models with average monthly stress factors.

6.2 Graphs selected for qualitative evaluation of model performance

A qualitative evaluation of both forward and inverse models is performed using a graph comparing reference observations with their simulated equiv-

alent together with a graph of unweighted simulated values versus weighted residuals. This graphical model fit analysis is somewhat subjective. However, it is good practice to perform such a visual inspection prior to use performance criteria for objective evaluation of model quality (Legates and McCabe, 1999; Hill and Tiedeman, 2007; Moriasi et al., 2007). Furthermore, the graph of unweighted simulated values versus weighted residuals is particularly useful for detecting model bias (Hill and Tiedeman, 2007). Given that the weighted residuals wr_i are calculated as:

$$wr_i = w_i^{1/2} \times \left[y_i^{obs} - y_i^{sim} \right] \tag{6.1}$$

$$= w_i^{1/2} \times r_i \tag{6.2}$$

where w_i is the weight of the i^{th} observation, y_i^{obs} is the i^{th} observed value, y_i^{sim} is the i^{th} simulated value, and r_i is the i^{th} residual, a cluster of negative values of wr_i indicate that simulated values are systematically overestimated, and vice versa. Furthermore, with weights calculated using a constant coefficient of variation, residuals are emphasised proportionally to their observed value. Therefore, similar wr_i values indicate similar relative errors. This is illustrated in Table 6.2.

Table 6.2: A simple numerical example illustrating that similar weighted residual values correspond to similar relative errors when using a constant coefficient of variation for calculating the weights.

y_i^{obs}	y_i^{sim}	r_i	relative error	cv_i	w_i	wr_i
1.0	0.9	0.1	10 %	0.10	100	1
10.0	9.0	1.0	10 %	0.10	1	1

6.3 Qualitative evaluation of models with daily stress factors

6.3.1 Influence of horizontal spatial discretisation

The simplified models focusing on horizontal spatial discretisation are the *250 m*, *500 m*, *750 m*, and *1000 m* models. They are characterised by an element side length of 250 m, 500 m, 750 m, and 1000 m, respectively. They each have 8 layers (5 layers of 1 m for the unsaturated zone, 1 layer of 5 m, 1 layer of 10 m, and 1 layer of 30 m for the saturated zone) and they each use van Genuchten saturation–pressure relations in the unsaturated zone. The number of nodes, number of elements, and execution time of each model for a 5-year simulation are given in Table 6.3.

Table 6.3: Comparison of the number of nodes, number of elements, and execution time of the *250 m*, *500 m*, *750 m*, and *1000 m* models for a 5-year simulation with daily stress factors.

	Number of nodes	Number of elements	Execution time
250 m	61,884	107,536	34.88 h
500 m	16,200	27,544	5.93 h
750 m	7,245	12,040	2.21 h
1000 m	4,302	7,016	1.27 h

6.3.1.1 Forward model

Whatever the horizontal spatial discretisation, the graphs comparing reference observations with their simulated equivalent indicate that flow rates are underestimated during low flow periods and overestimated during high flow periods (Figure 6.1). The underestimation is almost equivalent for each model. The overestimation is higher for models with a coarse horizontal spatial discretisation. This is clearly visible on peak flow rates. These graphs shows also that hydraulic heads are more likely to be systematically underes-

timated or overestimated (observed and simulated values shifted) by models with a coarse horizontal spatial discretisation (see also Figure B.1 to Figure B.5 in Appendix B). The graphs of unweighted simulated values versus weighted residuals plotted for the calibration period support these findings (Figure 6.2). Furthermore, these graphs highlight that the relative errors on high flow rates are minor relative to the relative errors on low flow rates for models with a fine horizontal spatial discretisation. These flow rate issues induce mass balance errors which increase progressively from -8 % for the *250 m* model to 7 % for the *1000 m* model (Figure 6.3). This transition from negative to positive values is related to the fact that the systematic underestimation of low flow rates is progressively compensated and then exceeded by the overestimation of high flow rates when the horizontal spatial discretisation gets coarser. The compensation of the underestimated low flow rates by the overestimated high flow rates is almost perfect for the *500 m* model which leads to a very small mass balance of -1 %. The overall graphical model fit analysis suggests that flow rates are more sensitive to horizontal spatial discretisation than hydraulic heads.

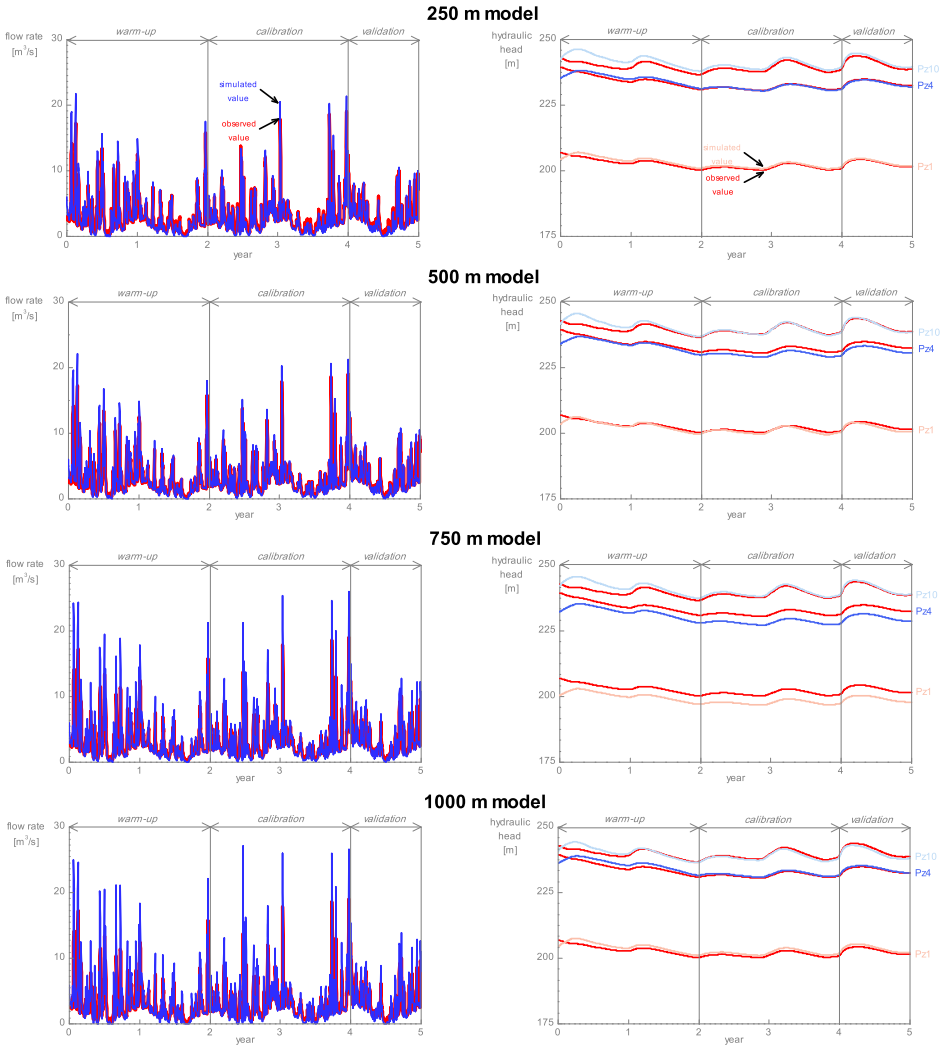

Figure 6.1: Flow rates and hydraulic heads simulated with the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors are compared with flow rate and hydraulic head observations produced with the 5-year reference simulation. These graphs indicate that flow rates are poorly simulated by models with a coarse horizontal spatial discretisation.

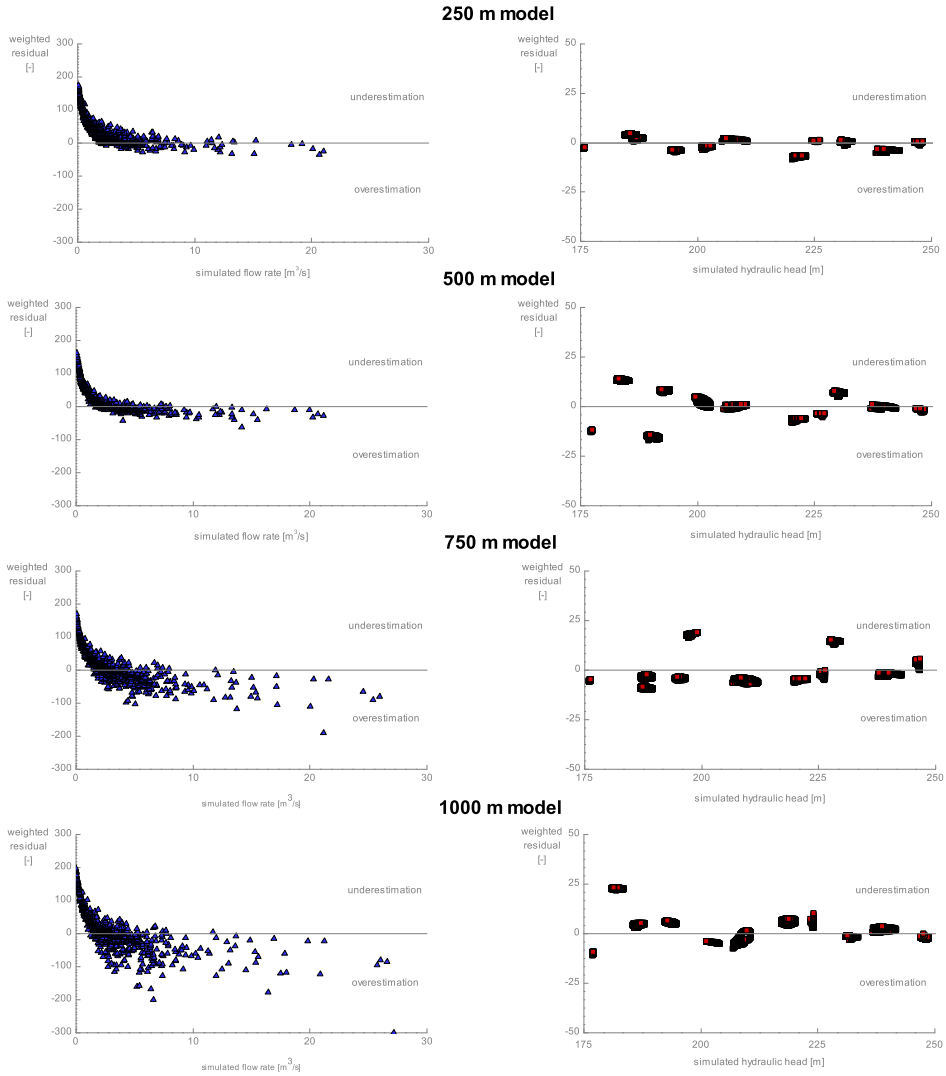

Figure 6.2: The graphs of unweighted simulated values versus weighted residuals shows that the major issues of models with a coarse horizontal spatial discretisation is the underestimation of flow rates during low flow periods and the overestimation of flow rates during high flow periods.

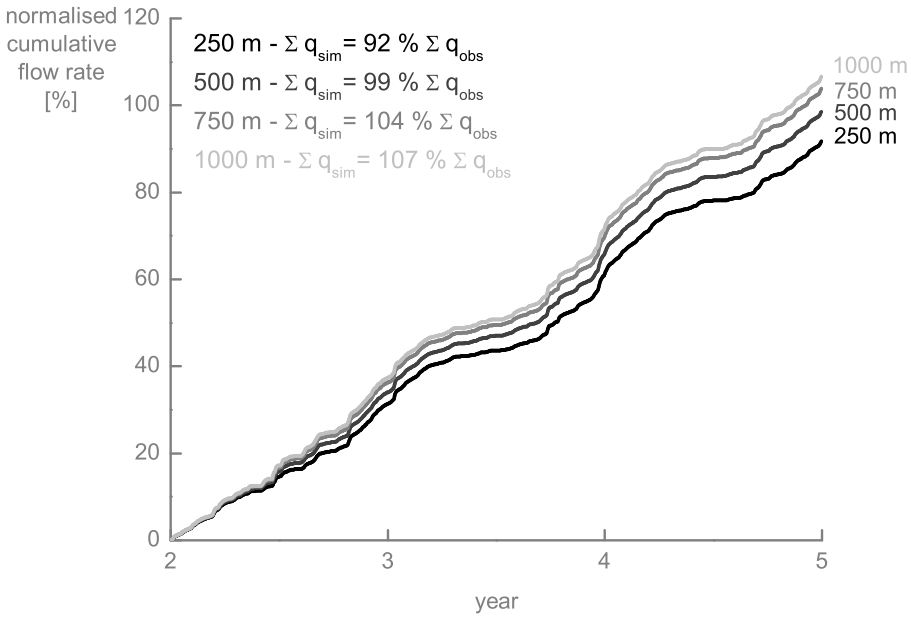

Figure 6.3: The mass balance errors calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors are -8 %, -1 %, 4 %, and 7 %, respectively.

6.3.1.2 Sensitivity analysis

The set of observations is composed of 731 flow rates (i.e. 1 per day for 2 years) and 8,772 hydraulic heads (i.e. 1 per day and per piezometer for 2 years) (see Section 5.3.4). The set of parameters includes 44 parameters (each parameter with a non-zero value) (see section 5.3.3).

Composite scaled sensitivities are quite similar for each model (Figure 6.4). The most sensitive parameters are the van Genuchten parameter β_{vG} of Mat I and Mat II and the evapotranspiration fitting parameters $\theta_{e2\,\&\,t2}$ of Mat 1, Mat 2, and Mat 3. This suggests that physically-based and spatially-distributed models are particularly sensitive to parameters governing flow in the unsaturated zone and evapotranspiration processes. The saturated hydraulic conductivity K of Mat IV is also slightly more sensitive than the other

parameters. This is probably related to the fact that most of the piezometers are located in the Carboniferous limestones (Figure 5.7). However, no parameter is significantly less sensitive than the others since the ratios between maximum and minimum *CSS* values for the *250 m*, *500 m*, *750 m*, and *1000 m* models are 13 %, 8 %, 11 %, and 17 %, respectively. This is far greater than the critical value of 1 %.

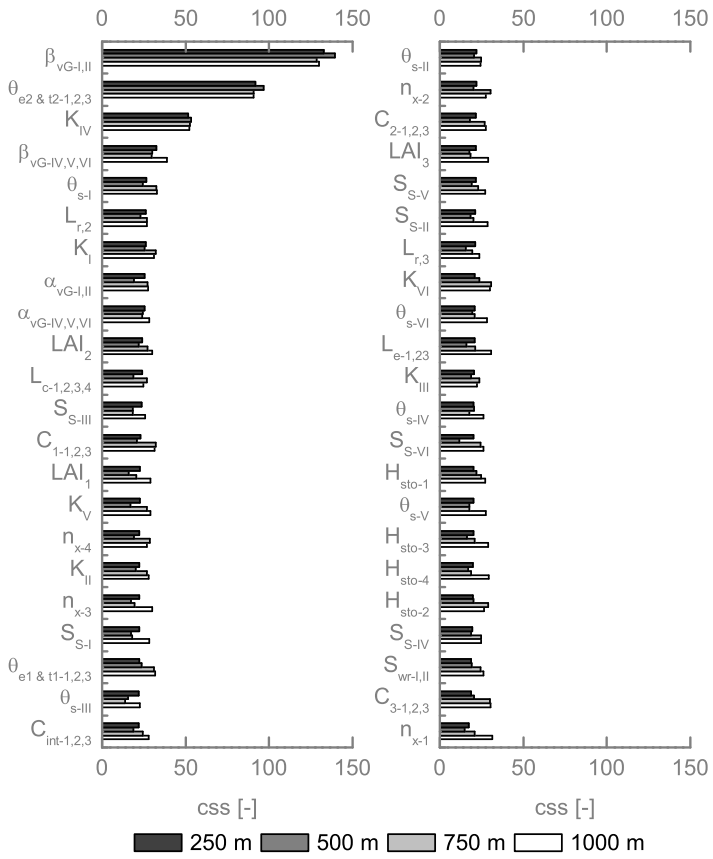

Figure 6.4: Composite scaled sensitivities calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* with daily stress factors indicate that the most sensitive parameters are related to flow in the unsaturated zone ($\beta_{vG-I,II}$) and evapotranspiration processes ($\theta_{e2\ \&\ t2-1,2,3}$). However, the other parameters are also relatively sensitive.

Parameter correlation coefficient values are slightly greater for models with a coarse horizontal spatial discretisation. However, the maximum abso-

lute value of pcc is lower than the critical value of 0.95 for each model.

6.3.2 Influence of vertical spatial discretisation

The simplified models focusing on vertical spatial discretisation are the *8 layers*, *6 layers*, and *3 layers* models. The *8 layers* model is identical to the *250 m* model of the previous section. They are characterised by the number and the thickness of their layers, so, the way of discretising the unsaturated zone (top layers) and the saturated zone (bottom layers). The *8 layers* model is composed of 5 layers of 1 m, 1 layer of 5 m, 1 layer of 10 m, and 1 layer of 30 m. The *6 layers* model is composed 5 layers of 1 m and 1 layer of 45 m. The *3 layers* model is composed of 1 layer of 3 m, 1 layer of 2 m, and 1 layer of 45 m. The other features of these models are an element side length of 250 m and van Genuchten saturation–pressure relations in the unsaturated zone. The number of nodes, number of elements, and execution time of each model for a 5-year simulation are given in Table 6.4.

Table 6.4: Comparison of the number of nodes, number of elements, and execution time of the *8 layers*, *6 layers*, and *3 layers* models for a 5-year simulation with daily stress factors.

	Number of nodes	Number of elements	Execution time
8 layers	61,884	107,536	34.88 h
6 layers	48,132	80,652	24.22 h
3 layers	27,504	40,326	15.03 h

6.3.2.1 Forward model

As for the horizontal spatial discretisation, the graphs comparing reference observations with their simulated equivalent indicate that flow rates are underestimated during low flow periods and overestimated during high flow periods irrespective of the vertical spatial discretisation (Figure 6.5). However, compared to models with a coarse horizontal spatial discretisation, the

overestimation of high flow rates is slightly lower. These graphs show also that hydraulic heads are poorly simulated (observed and simulated values shifted and wrong hydraulic head variations) by models with a coarse vertical spatial discretisation (see also Figure B.6 to Figure B.10 in Appendix B). The graphs of unweighted simulated values versus weighted residuals plotted for the calibration period support these findings (Figure 6.6). The mass balance errors for the *8 layers*, *6 layers*, and *3 layers* models are -8 %, -11 %, and 10 %, respectively (Figure 6.7). As the *8 layers* and *6 layers* models hardly overstimate high flow rates, the underestimation of low flow rates is not compensated and the mass balance error is negative. Conversely, since the *3 layers* model overestimates significantly high flow rates, the underestimation of low flow rates is exceeded and the mass balance error is positive. The overall graphical model fit analysis suggests that hydraulic heads are more sensitive to vertical spatial discretisation than flow rates.

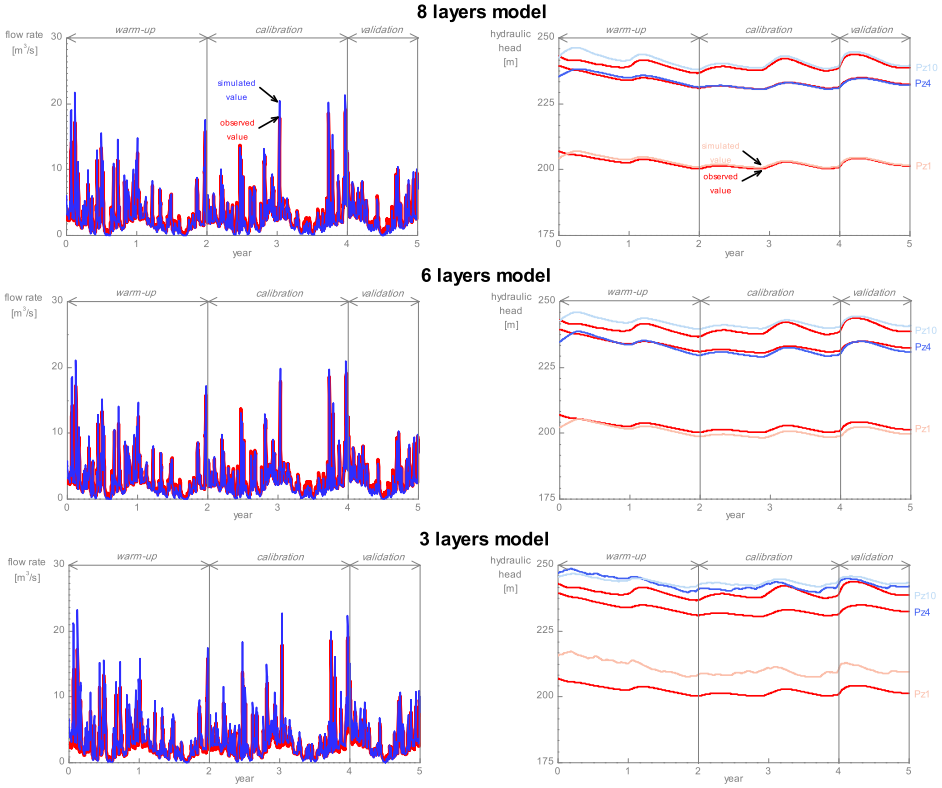

Figure 6.5: Flow rates and hydraulic heads simulated with the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors are compared with flow rate and hydraulic head observations produced with the 5-year reference simulation. These graphs indicate that hydraulic heads are poorly simulated by models with a coarse vertical spatial discretisation.

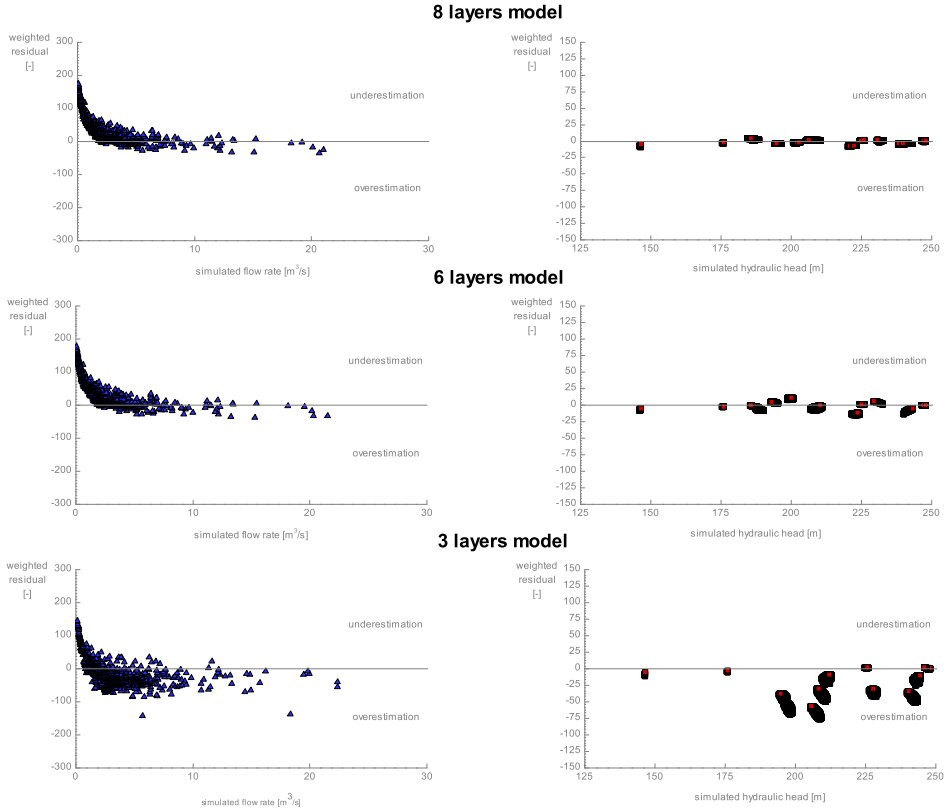

Figure 6.6: The graphs of unweighted simulated values versus weighted residuals for the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors show that the major issue of models with a coarse vertical spatial discretisation is the simulation of hydraulic heads.

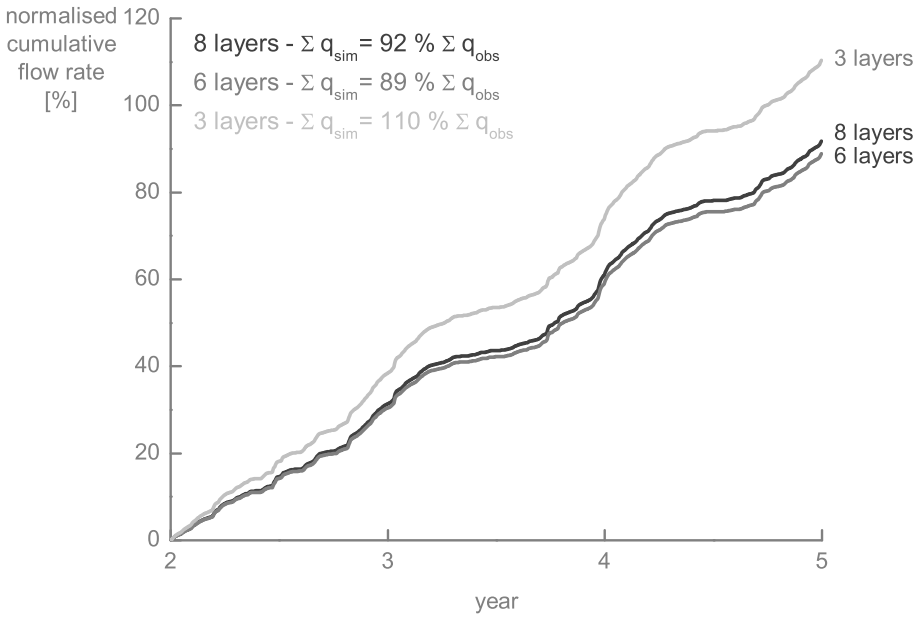

Figure 6.7: The mass balance errors calculated for the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors are -8 %, -11 %, and 10 %, respectively.

6.3.2.2 Sensitivity analysis

The sets of observations and parameters are identical to those used for the models with different horizontal spatial discretisations (731 flow rate observations, 8,772 hydraulic head observations, and 44 parameters).

As for the models differing by their horizontal spatial discretisation, the most sensitive parameters are the van Genuchten parameter β_{vG} of Mat I and Mat II, the evapotranspiration fitting parameters $\theta_{e2\,\&\,t2}$ of Mat 1, Mat 2, and Mat 3, and, to a lesser extent, the saturated hydraulic conductivity K of Mat IV (Figure 6.8). However, the values of css for $\beta_{vG-I-II}$ and $\theta_{e2\,\&\,t2-1,2,3}$ are far higher for the *3 layers* model. This indicates that the coarse representation of the unsaturated zone increases the relative sensitivity of parameters governing flow in the unsaturated zone and evapotranspiration processes. The

ratios between maximum and minimum *css* values for the *8 layers*, *6 layers*, and *3 layers* models are 13 %, 10 %, and lower than 1 %, respectively. This suggests that some parameters are insensitive relative to the most sensitive parameter for the *3 layers* model.

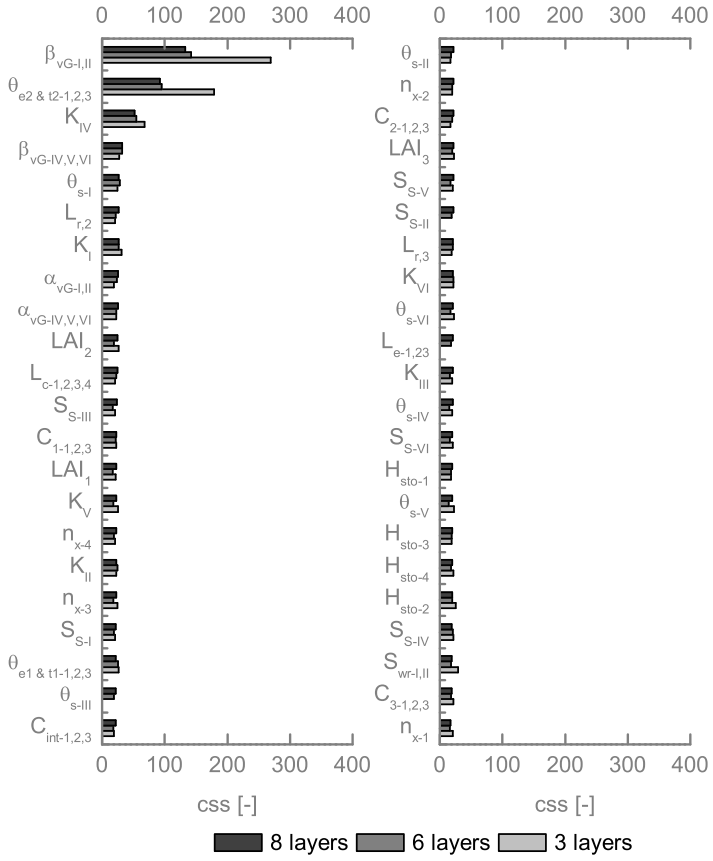

Figure 6.8: Composite scaled sensitivities calculated for the *8 layers*, *6 layers*, and *3 layers* with daily stress factors indicate that the most sensitive parameters are related to flow in the unsaturated zone ($\beta_{vG-I,II}$) and evapotranspiration processes $\theta_{e2\ \&\ t2-1,2,3}$.

As for the models with different horizontal spatial discretisations, parameter correlation coefficient values are slightly greater for models with a coarse vertical spatial discretisation. However, the maximum absolute value of *pcc* is lower than the critical value of 0.95 for each model.

6.3.3 Influence of the representation of the unsaturated zone

The simplified models focusing on the representation of the unsaturated zone are the *van Genuchten* and *linear* models. The *van Genuchten* model is similar to the *250 m* and *8 layers* models of the previous sections. As proposed by Lehmann et al. (2008), the linearisation of the van Genuchten saturation–pressure curve is performed by taking the tangent at the inflection point. The pressure heads h_a and h_b corresponding respectively to $S_e = 0$ and $S_e = 1$ constitute the parameters of the linear saturation–pressure relation (Figure 6.9). The values of h_a and h_b estimated for the covering sediment materials (Mat I and Mat II) and the fissured hard rock materials (Mat IV, Mat V, and Mat VI) of the synthetic case are presented in Table 6.5. The other features of the *van Genuchten* and *linear* models are an element side length of 250 m and 8 layers (5 layers of 1 m for the unsaturated zone, 1 layer of 5 m, 1 layer of 10 m, and 1 layer of 30 m for the saturated zone). Their number of nodes, number of elements, and execution time for a 5-year simulation are given in Table 6.6. The execution time of the *linear* model is longer than the execution time of the *van Genuchten* model which suggests that simplifying saturation–pressure relations in the unsaturated zone is not an effective way to save time.

Table 6.5: Values of h_a and h_b parameters for the materials of the synthetic case.

	h_a [m]	h_b [m]
Mat I and Mat II	-12.00	-0.21
Mat IV, Mat V, and Mat VI	-220.00	-12.95

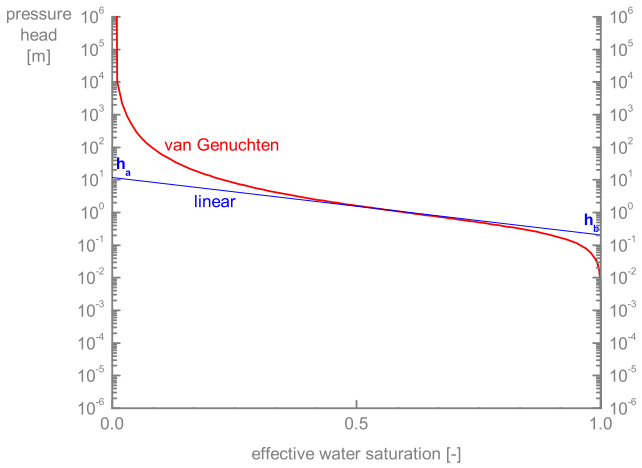

Figure 6.9: The linearisation of the van Genuchten saturation–pressure curve is performed by taking the tangent at the inflection point for obtaining the pressure heads h_a and h_b corresponding to $S_e = 0$ and $S_e = 1$, respectively.

Table 6.6: Comparison of the number of nodes, number of elements, and execution time of the *van Genuchten* and *linear* models for a 5-year simulation with daily stress factors.

	Number of nodes	Number of elements	Execution time
van Genuchten	61,884	107,536	34.88 h
linear	61,884	107,536	47.69 h

6.3.3.1 Forward model

The graphs comparing reference observations with their simulated equivalent indicate that the simulation of both flow rates and hydraulic heads is severely impaired with the *linear* model (Figure 6.10 and Figure B.11 to Figure B.15 in Appendix B). Flow rates are in general overestimated and hydraulic heads suffer from systematic errors and wrong variations. This is also visible on the graphs of unweighted simulated values versus weighted residuals plotted for the calibration period (Figure 6.11). Furthermore, with a mass balance error of 36 %, the *linear* model presents a severe problem of mass (Figure

6.12). The overall graphical model fit analysis suggests that simplifications of saturation–pressure relations in the unsaturated zone can severely impair model results in terms of both flow rates and hydraulic heads.

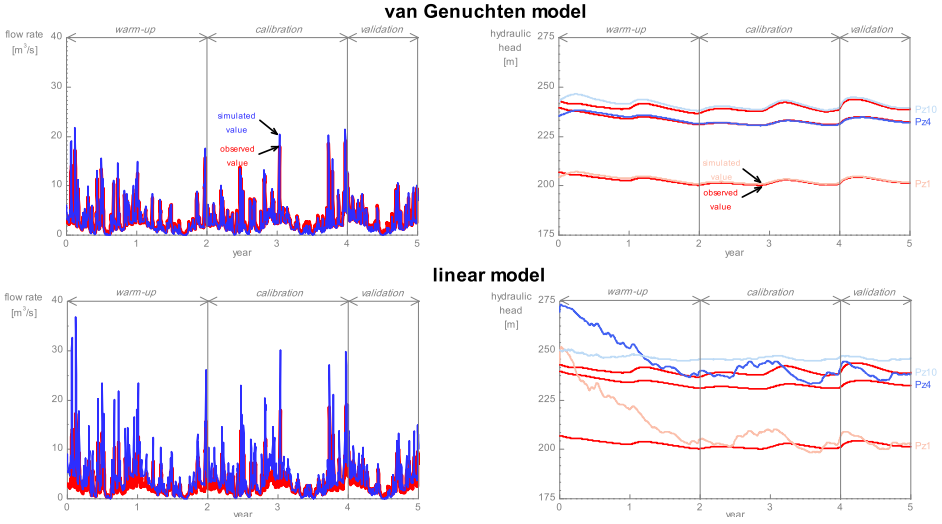

Figure 6.10: Flow rates and hydraulic heads simulated with the *van Genuchten* and *linear* models with daily stress factors are compared with flow rate and hydraulic head observations produced with the 5-year reference simulation. These graphs indicates that both flow rates and hydraulic heads are poorly simulated by the model with linear saturation–pressure relations in the unsaturated zone.

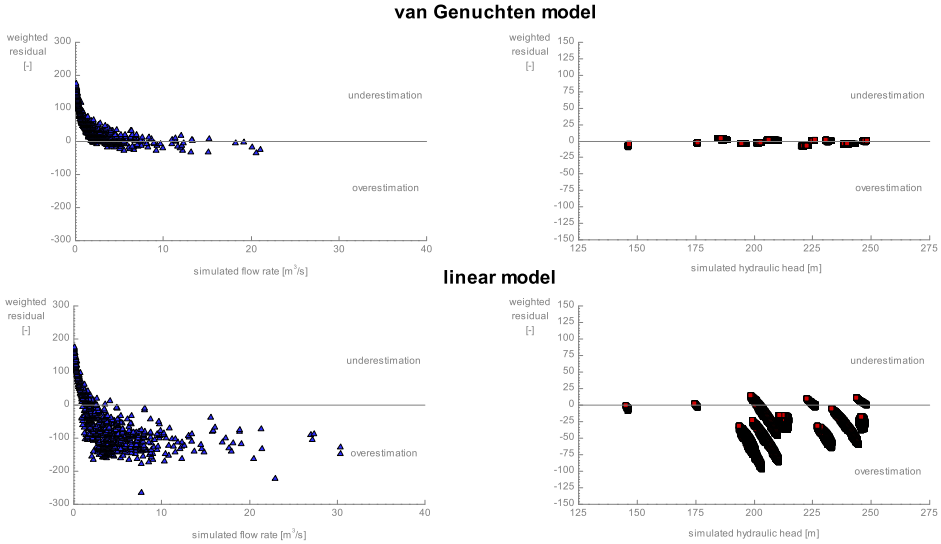

Figure 6.11: The graphs of unweighted simulated values versus weighted residuals of the *van Genuchten* and *linear* models with daily stress factors show that both flow rates and hydraulic heads are poorly simulated by the model with linear saturation–pressure relations in the unsaturated zone.

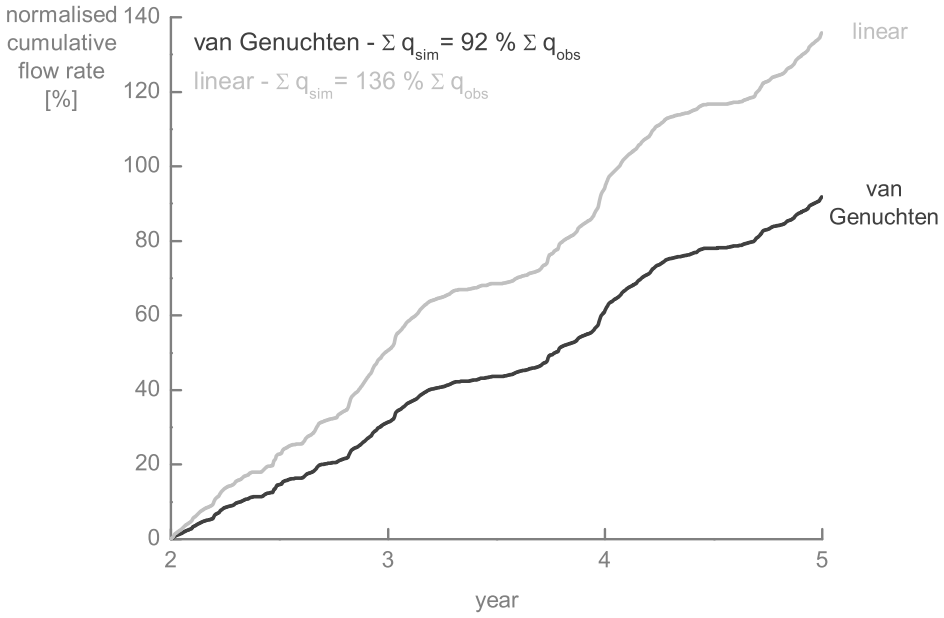

Figure 6.12: The mass balance errors calculated for the *van Genuchten* and *linear* models with daily stress factors are -8 % and 36 %, respectively.

6.3.3.2 Sensitivity analysis

The sets of parameters and observations are identical to those used for the models with different horizontal spatial discretisations, excepting van Genuchten parameters (α_{vG} and β_{vG}) that are replaced by their linear equivalent (h_a and h_b) for the *linear* model (731 flow rate observations, 8,772 hydraulic head observations, and 44 parameters).

Although the highest composite scaled sensitivities of the *van Genuchten* and *linear* models are in the same range of values, their distribution is quite different (Figure 6.13). The css values for the *van Genuchten* model indicate that the van Genuchten parameter β_{vG} of Mat I and Mat II, the evapotranspiration fitting parameters $\theta_{e2\,\&\,t2}$ of Mat 1, Mat 2, and Mat 3, and, to a lesser extent, the saturated hydraulic conductivity K of Mat IV are comparatively more sensitive. This is not the case of the *linear* model for which no parameter is significantly more sensitive than the others. This is supported by the ratios of 13 % and 32% between maximum and minimum css values calculated for the *van Genuchten* and *linear*, respectively. The higher value of this ratio for the *linear* model suggests that no parameter is predominant in terms of sensitivity.

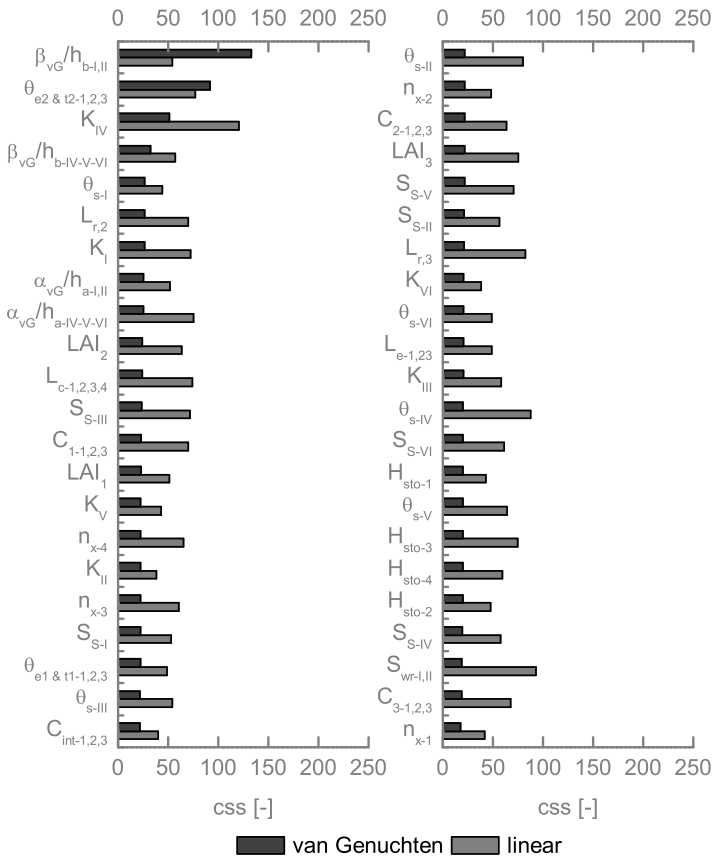

Figure 6.13: Composite scaled sensitivities calculated for the *van Genuchten* model with daily stress factors indicate that the most sensitive parameters are related to flow in the unsaturated zone ($\beta_{vG-I,II}$) and evapotranspiration processes $\theta_{e2\ \&\ t2-1,2,3}$. This is not the case of the *linear* model whose parameters have almost the same composite scaled sensitivity.

Whatever the representation of the unsaturated zone, van Genuchten or linear saturation–pressure relations, the maximum absolute value of *pcc* is lower than the critical value of 0.95.

6.4 Qualitative evaluation of models with average monthly stress factors

6.4.1 Influence of horizontal spatial discretisation

The execution times of the *250 m*, *500 m*, *750 m*, and *1000 m* models for a 5-year simulation with daily and average monthly stress factors are compared in Table 6.7. The mean relative gain in execution time for these models is 88 %. This is tremendous. Therefore, for problems which are satisfied with model results at such a time resolution, such a simplification constitutes an efficient way of saving time.

Table 6.7: Comparison of the execution times of the *250 m*, *500 m*, *750 m*, and *1000 m* models for a 5-year simulation with average monthly stress factors and daily stress factors.

	Execution time	
	Daily stress factors	Average monthly stress factors
250 m	34.88 h	6.08 h
500 m	5.93 h	0.71 h
750 m	2.21 h	0.20 h
1000 m	1.27 h	0.13 h

6.4.1.1 Forward model

The graphs comparing reference observations with their simulated equivalent indicate that flow rates are most often underestimated during low flow periods and overestimated during high flow periods (Figure 6.14). The underestimation is almost equivalent for each model. The overestimation is sligthly higher for models with a coarse horizontal spatial discretisation. This is visible on peak flow rates. These graphs show also that hydraulic heads are more likely to be systematically underestimated or overestimated (observed and simulated values shifted) by models with a coarse horizontal spatial dis-

cretisation (see also Figure B.16 to Figure B.20 in Appendix B). These findings, supported by the graphs of unweighted simulated values versus weighted residuals plotted for the calibration period (Figure 6.15), are identical to those obtained for the same models with daily stress factors. Furthermore, the same problems of mass conservation occur (Figure 6.16). However, with monthly stress factors, the lowest mass conservation error is obtained for the *1000 m* model and not the *500 m* model. This suggests that the underestimation of low flow rates is compensated by the overestimation of high flow rates only for the model with the coarsest horizontal spatial discretisation.

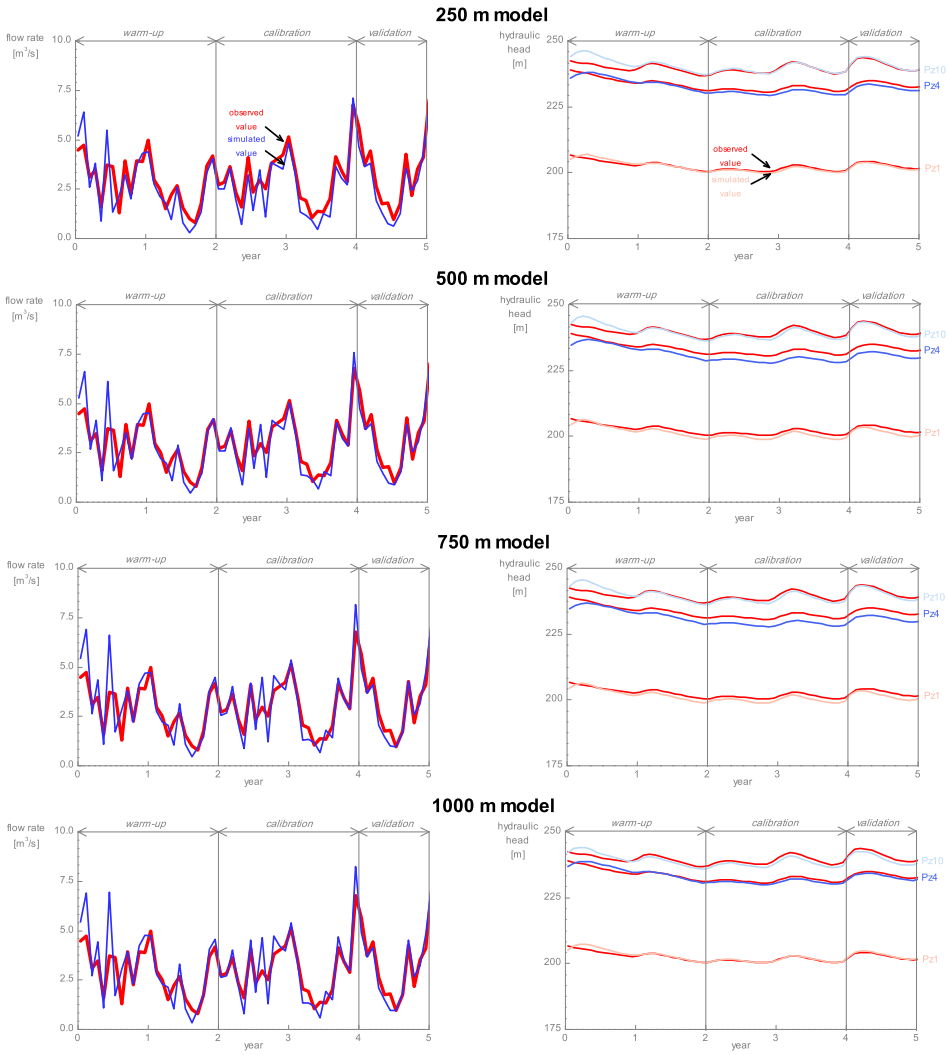

Figure 6.14: Flow rates and hydraulic heads simulated with the *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors are compared with flow rate and hydraulic head observations produced with the 5-year reference simulation. These graphs indicate that flow rate simulation is progressively impaired when horizontal spatial discretisation is coarsened.

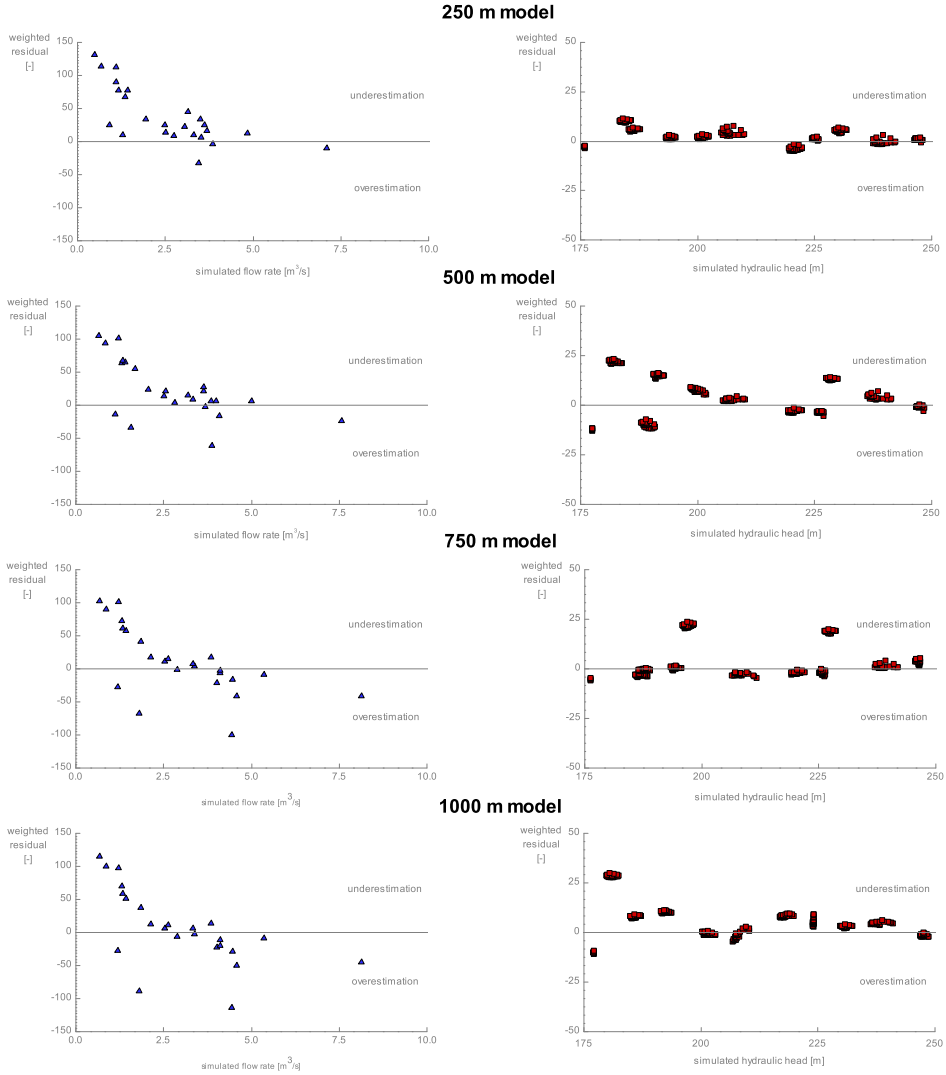

Figure 6.15: The graphs of unweighted simulated values versus weighted residuals for the *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors show that the major issue of models with a coarse horizontal spatial discretisation is the underestimation of flow rates during low flow periods.

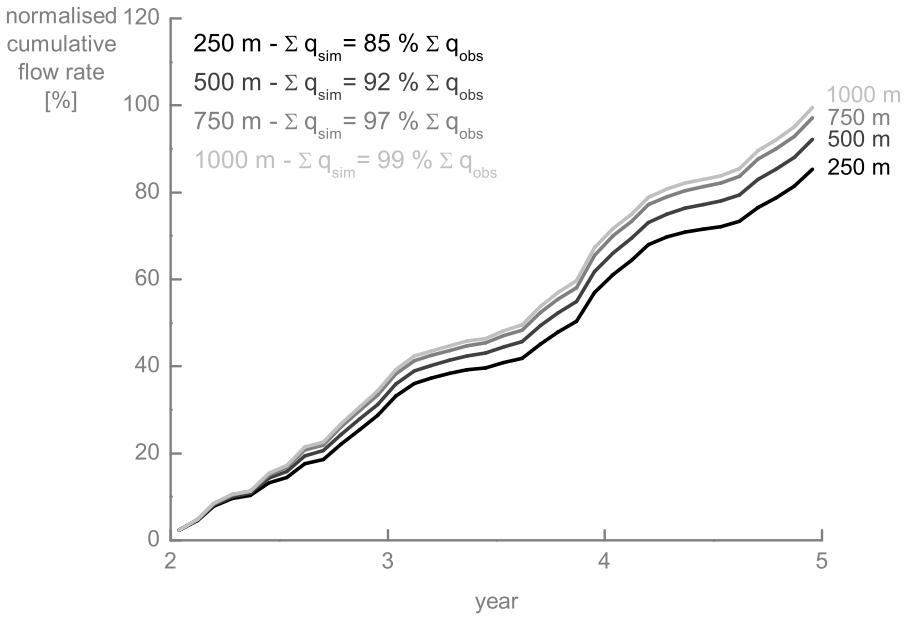

Figure 6.16: The mass conservation errors calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors are -15 %, -8 %, -3 %, and -1 %, respectively.

6.4.1.2 Sensitivity analysis

The sensitivity analysis is performed on the calibration period. The set of observations is composed of 24 flow rates (i.e. 1 per month for 2 years) and 288 hydraulic heads (i.e. 1 per month and per piezometer for 2 years). These monthly observations are obtained by monthly averaging the daily observations (arithmetic mean). The set of parameters includes 44 parameters (each parameter with a non-zero value).

The most sensitive parameters are the van Genuchten parameter β_{vG} of Mat I and Mat II, the evapotranspiration fitting parameters $\theta_{e2\ \&\ t2}$ of Mat 1, Mat 2, and Mat 3, and, to a lesser extent, the saturated hydraulic conductivity K of Mat IV (Figure 6.17). These results are similar to those obtained for the same models with daily stress factors. However, the ratios between maximum

and minimum *css* values are quite different since the values obtained for the *250 m*, *500 m*, *750 m*, and *1000 m* with average monthly stress factors are 3 %, 3 %, 2 %, and lower than 1 %, respectively. This is up to 16 % lower than the ratio values obtained with daily stress factors which indicates a significant loss of sensitivity for the less sensitive parameters of models with average monthly stress factors.

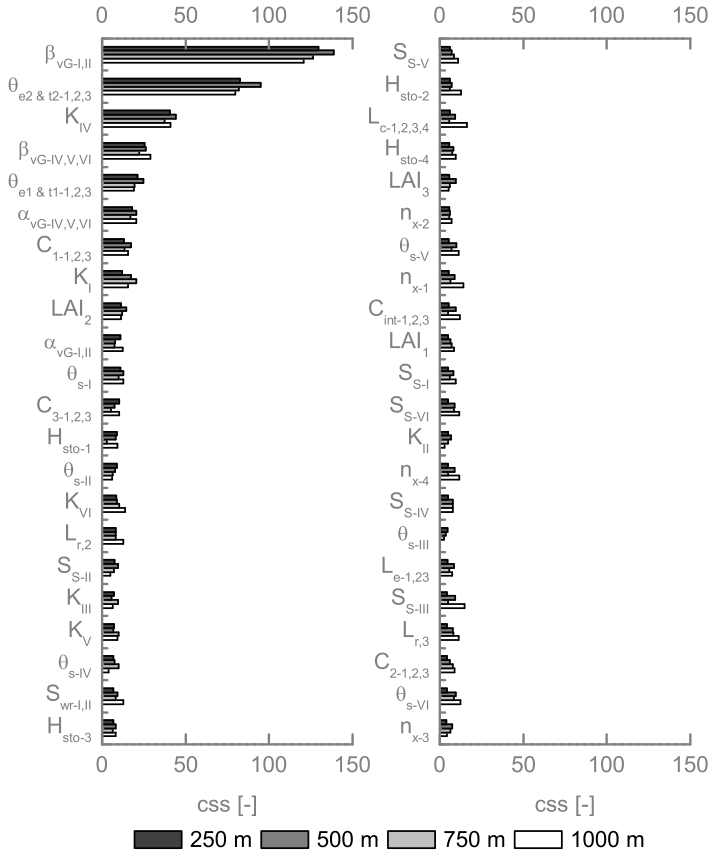

Figure 6.17: As for their equivalent with daily stress factors, composite scaled sensitivities calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors indicate that the most sensitive parameters are related to flow in the unsaturated zone ($\beta_{vG-I,II}$) and evapotranspiration processes ($\theta_{e2\ \&\ t2-1,2,3}$).

The maximum absolute value of *pcc* is lower than the critical value of

0.95 for each model.

6.4.1.3 Inverse model

The optimisation is performed on the calibration period with lower and upper constraints on parameters in order to prevent them from taking unrealistic values. Compared to sensitivity analysis, the set of observations is identical (24 flow rates and 288 hydraulic heads) but the set of parameters is slightly reduced (32 parameters instead of 44) since interception and evapotranspiration parameters are withdrawn. This choice is made for saving time and focusing only on surface and subsurface flow parameters familiar to hydrogeologists.

The graphs of model fit indicate that optimisation significantly improves the simulation of flow rates and, to a lesser extent, hydraulic heads for each model (Figure 6.18 and Figure B.31 to Figure B.35 in Appendix B). Additionally, the distributions of weighted residuals look closer from randomness (Figure 6.19), which suggests that models are less biased, and mass conservation errors are lower (Figure 6.20). This shows that optimisation is capable of modifying parameter values in order to compensate for coarse horizontal discretisations. However, model fit is not perfect. Furthermore, though most of them are still plausible, several optimised parameter values are far from their reference values (Figure 6.21). This is especially the case for the coarsest models whose improvement requires more significant changes in parameter values (Table B.1 in Appendix B). The highest relative changes are calculated for friction (n_x) and unsaturated zone parameters (θ_s and α_{vG}). The calculation of the relative changes in parameter is unfortunately possible only for synthetic cases for which reference parameter values are known. The only verification possible for real cases consists in making sure that optimised parameter val-

ues are plausible with regards to field data.

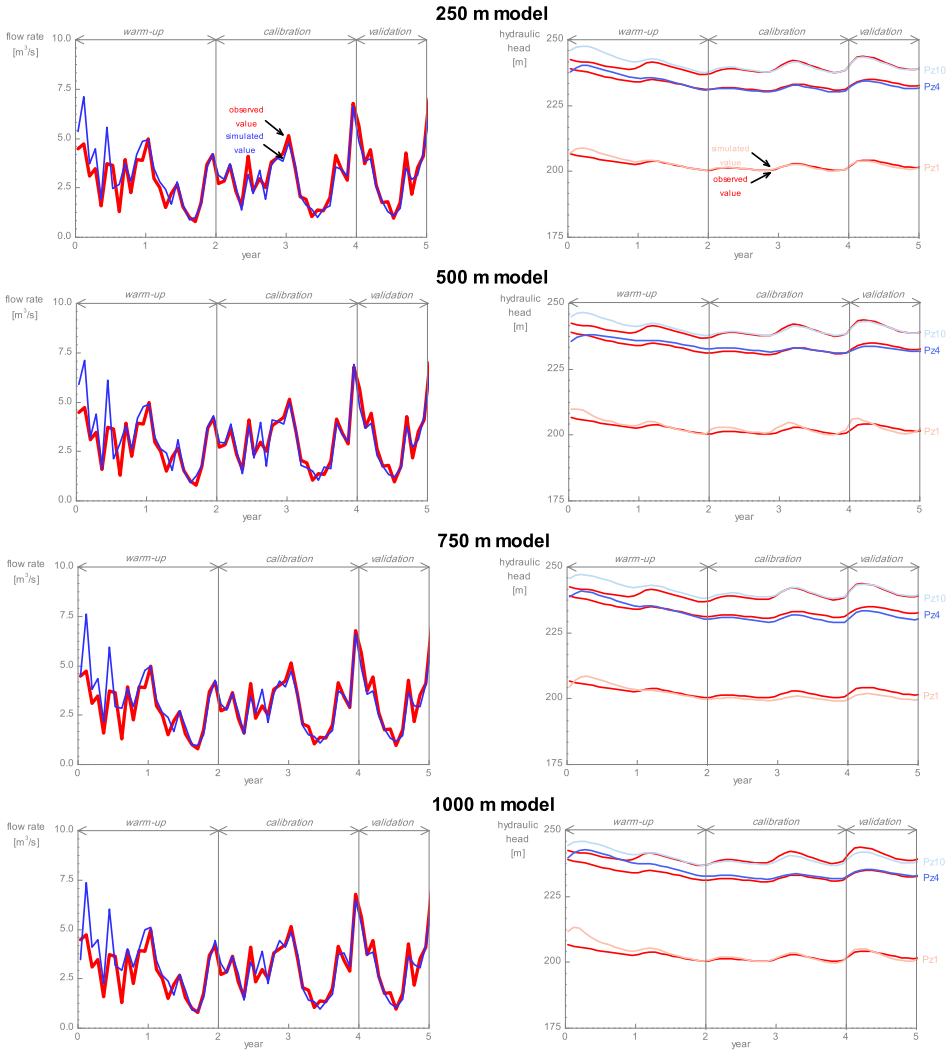

Figure 6.18: The optimisation significantly improves the simulation of flow rates for the *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors, even for models with a coarse horizontal spatial discretisation.

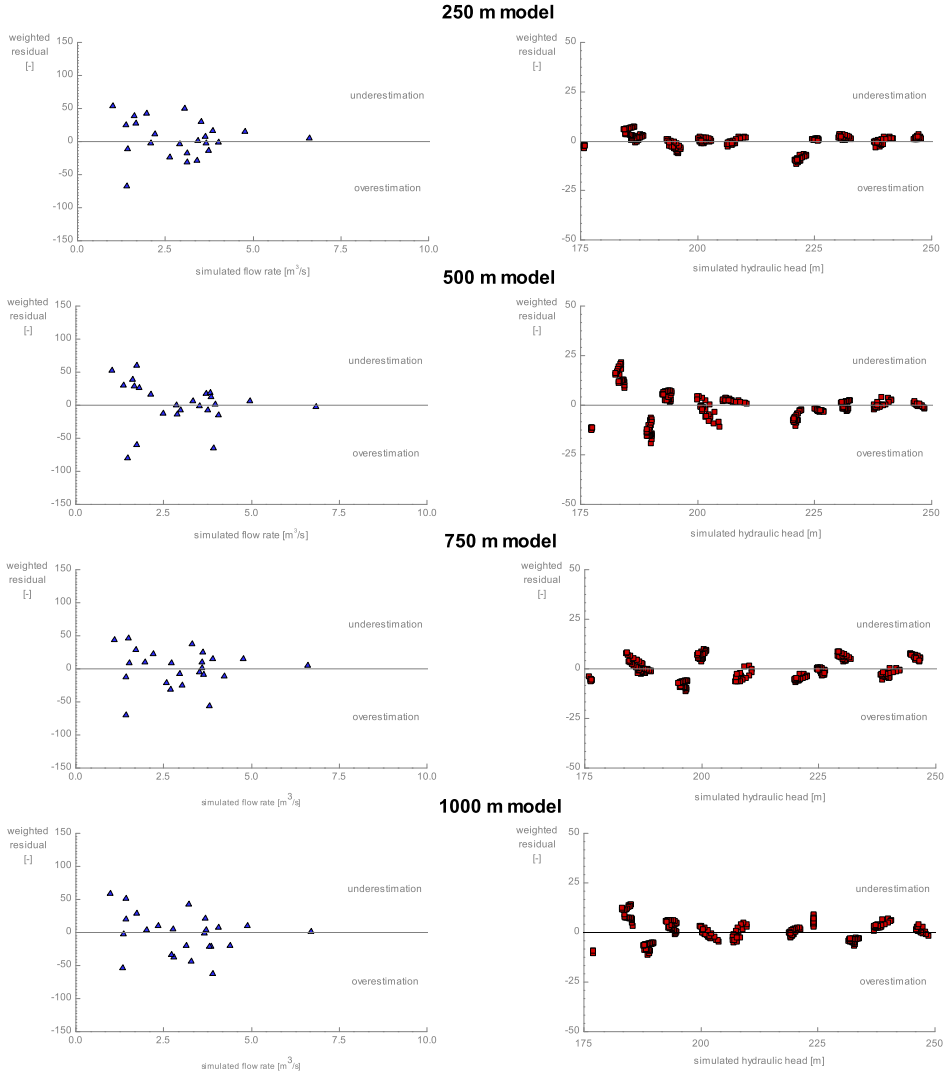

Figure 6.19: The weighted residuals of the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors are almost randomly distributed. This indicates that optimisation reduces model bias.

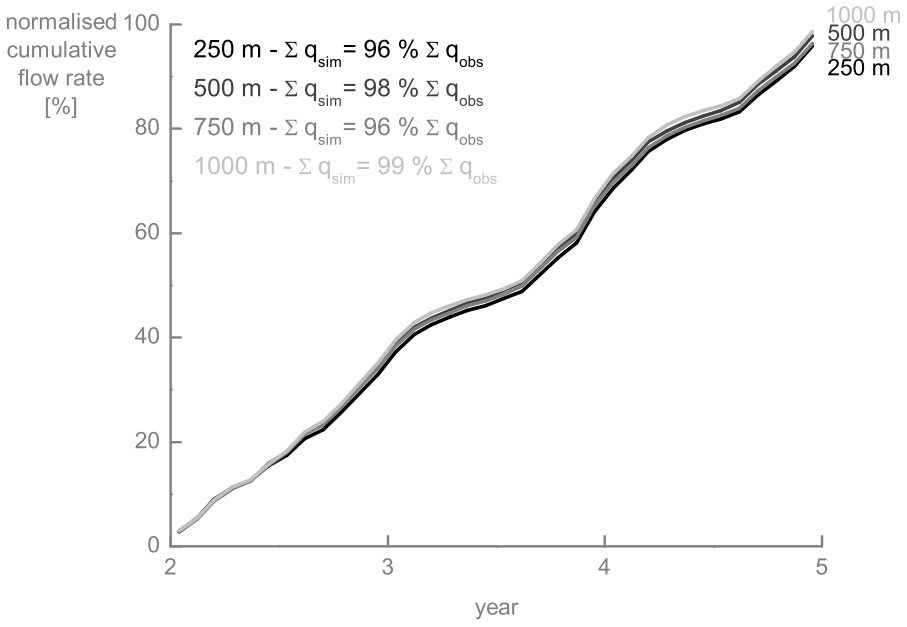

Figure 6.20: The mass conservation errors calculated for the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors are reduced to -4 %, -2 %, -4 %, and -1 %, respectively.

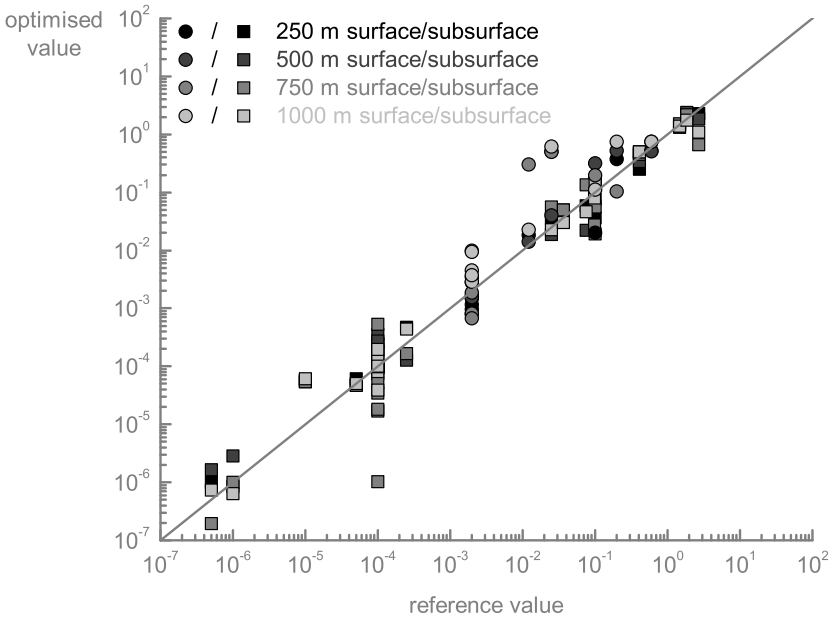

Figure 6.21: When horizontal spatial discretisation gets coarser, optimised parameter values are in general further from their reference values.

6.4.2 Influence of vertical spatial discretisation

The execution times of the *8 layers*, *6 layers*, and *3 layers* models for a 5-year simulation with daily and average monthly stress factors are compared in Table 6.8. The mean relative gain in execution time for these models is 85 % which higlights once more the tremendous time saved when working with average monthly stress factors instead of daily stress factors.

Table 6.8: Comparison of the execution times of the *8 layers*, *6 layers*, and *3 layers* models for a 5-year simulation with average monthly stress factors and daily stress factors.

	Execution time	
	Daily stress factors	Average monthly stress factors
8 layers	34.88 h	6.08 h
6 layers	24.22 h	3.68 h
3 layers	15.03 h	1.81 h

6.4.2.1 Forward model

The graphs comparing reference observations with their simulated equivalent indicate that flow rates are most often underestimated during low flow periods. Additionally, the model with the coarsest spatial discretisation sligthly overestimates flow rates during high flow period (Figure 6.22). These graph show also that hydraulic heads are poorly simulated (observed and simulated values shifted and false hydraulic head variations) by models with a coarse vertical spatial discretisation (see also Figure B.21 to Figure B.25 in Appendix B). These results, supported by the graph of unweighted simulated values versus weighted residuals plotted for the calibration period (Figure 6.23), are quite similar to those obtained for the same models with daily stress factors. Furthermore, the same mass conservation issues are observed (Figure 6.24).

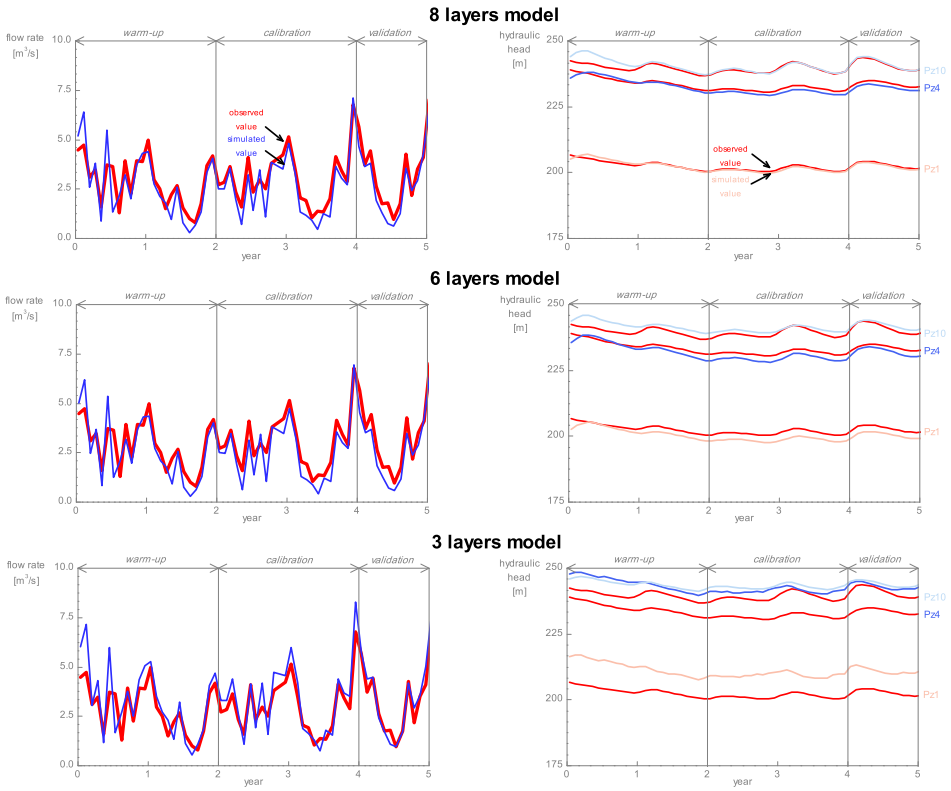

Figure 6.22: Flow rates and hydraulic heads simulated with the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors are compared with flow rate and hydraulic head observations produced with the 5-year reference simulation. These graphs indicate that hydraulic head simulation is progressively impaired when vertical spatial discretisation is coarsened.

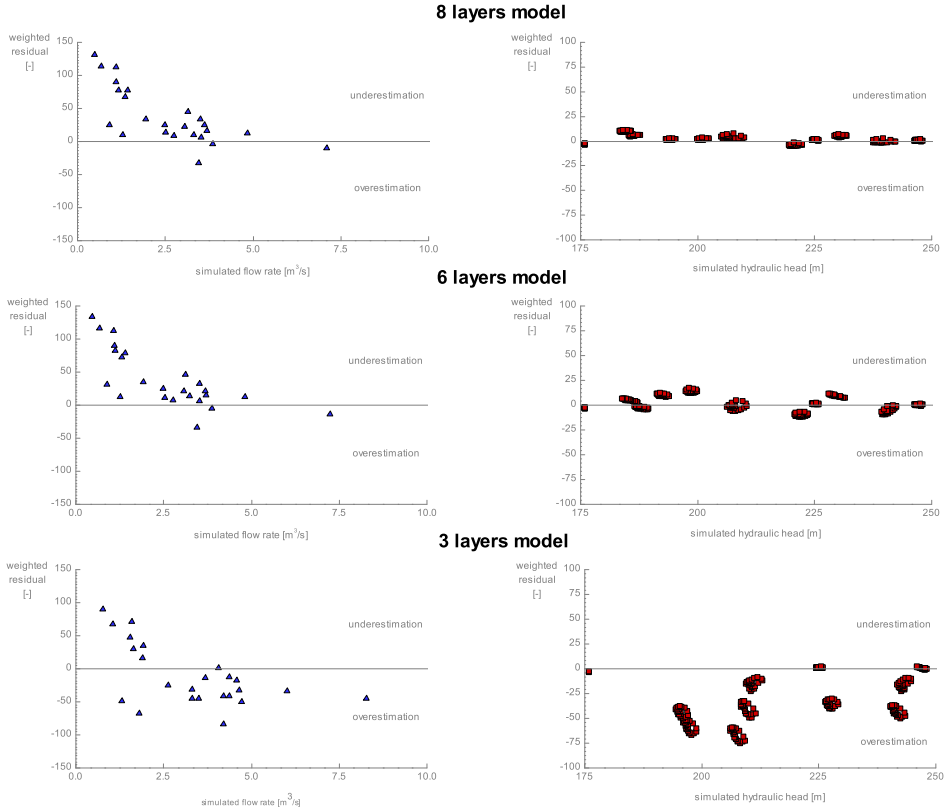

Figure 6.23: The graphs of unweighted simulated values versus weighted residuals for the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors show that the major issue of models with a coarse vertical spatial discretisation is the systematic overestimation of hydraulic heads.

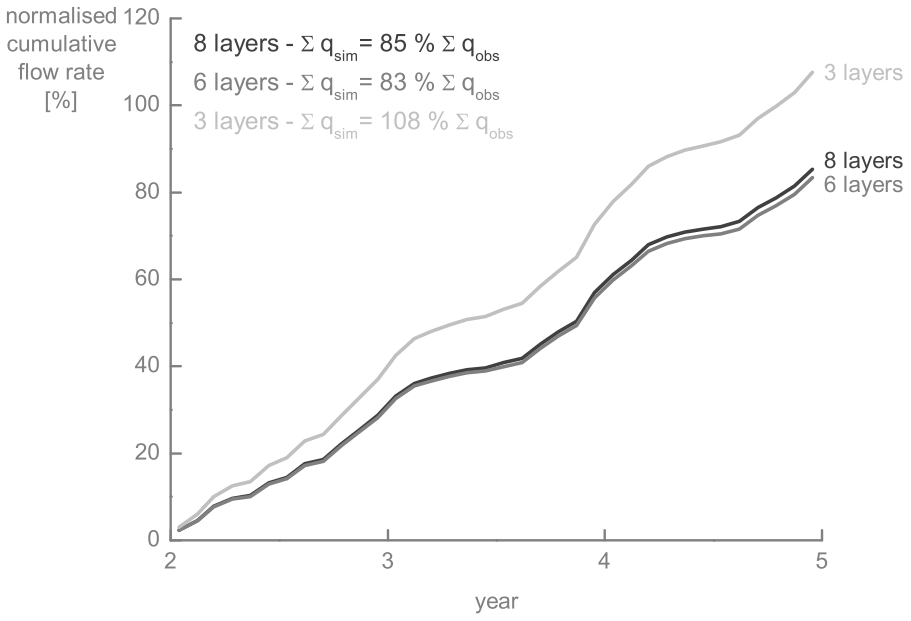

Figure 6.24: The mass conservation errors calculated for the *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors are -15 %, -17 %, and 8 %, respectively.

6.4.2.2 Sensitivity analysis

The sensitivity analysis is performed on the calibration period. The sets of observations and parameters are identical to those used for the models with different horizontal spatial discretisations and average monthly stress factors (24 flow rate observations, 288 hydraulic head observations, and 44 parameters).

The most sensitive parameters are the van Genuchten parameter β_{vG} of Mat I and Mat II, the evapotranspiration fitting parameters $\theta_{e2\,\&\,t2}$ of Mat 1, Mat 2, and Mat 3, and, to a lesser extent, the saturated hydraulic conductivity K of Mat IV (Figure 6.25). However, the values of css for $\beta_{vG-I-II}$ and $\theta_{e2\,\&\,t2-1,2,3}$ are far higher for the *3 layers* model. This suggests that the coarse representation of the unsaturated zone increases the relative sensitiv-

ity of parameters related to flow in the unsaturated zone and evapotranspiration processes. These results are similar to those obtained with daily stress factors. The ratios between maximum and minimum *css* values for the *8 layers*, *6 layers*, and *3 layers* are 3 %, 3 %, and 2 %, respectively. Apart from the *3 layers* model, this is lower than the values obtained with daily stress factors which suggests a loss of sensitivity for the less sensitive parameters of models with average monthly stress factors.

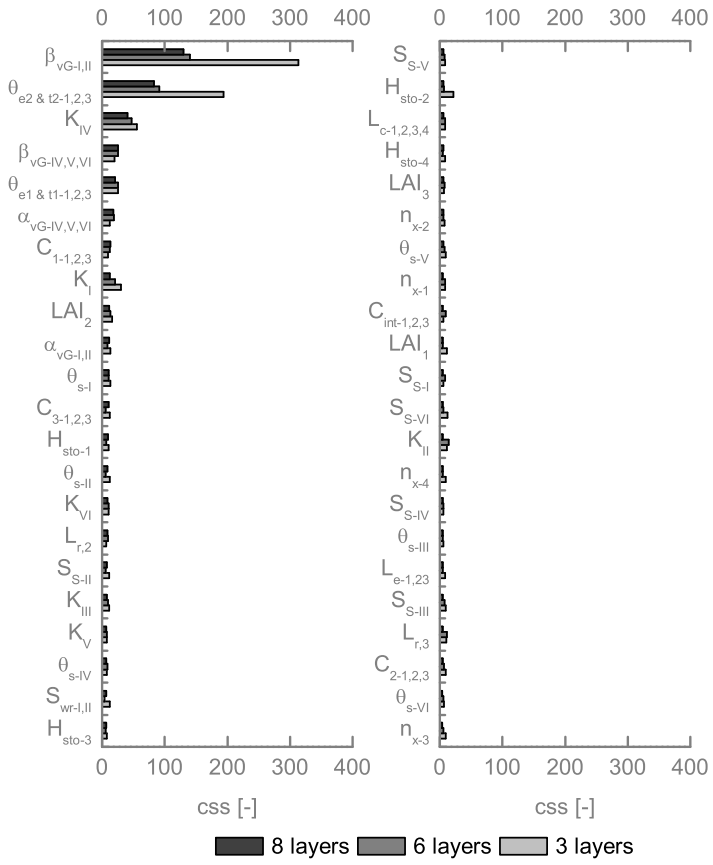

Figure 6.25: As for their equivalent with daily stress factors, composite scaled sensitivities calculated for the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors indicate that the most sensitive parameters are related to flow in the unsaturated zone ($\beta_{vG-I,II}$) and evapotranspiration processes ($\theta_{e2\ \&\ t2-1,2,3}$).

The pairs of parameters $C_{1-1,2,3}$–LAI_2 for the *6 layers* model and $\beta_{vG-I,II}$–$\theta_{e2\ \&\ t2-1,2,3}$ for the *3 layers* model have parameter correlation coefficient values of -0.97 and -0.96, respectively. This suggests that these parameters could be too correlated to be uniquely estimated.

6.4.2.3 Inverse model

The optimisation is performed on the calibration period. The sets of observations and parameters are identical to those used for models with different horizontal spatial discretisations (24 flow rate observations, 288 hydraulic head observations, and 32 parameters).

The graphical model fit analysis shows once more that optimisation significantly reduces misfit, model bias, and mass conservation errors, including for models with a coarse vertical spatial discretisation (Figures 6.26, 6.27, and 6.28 and Figure B.36 to Figure B.40 in Appendix B). However, model fit is not perfect and optimised parameter values are sometimes far from their reference values (Figure 6.29 and Table B.2 in Appendix B). This is especially the case for the coarsest models . The highest relative changes in parameters are calculated for friction (n_x) and unsaturated zone parameters (θ_s and α_{vG}).

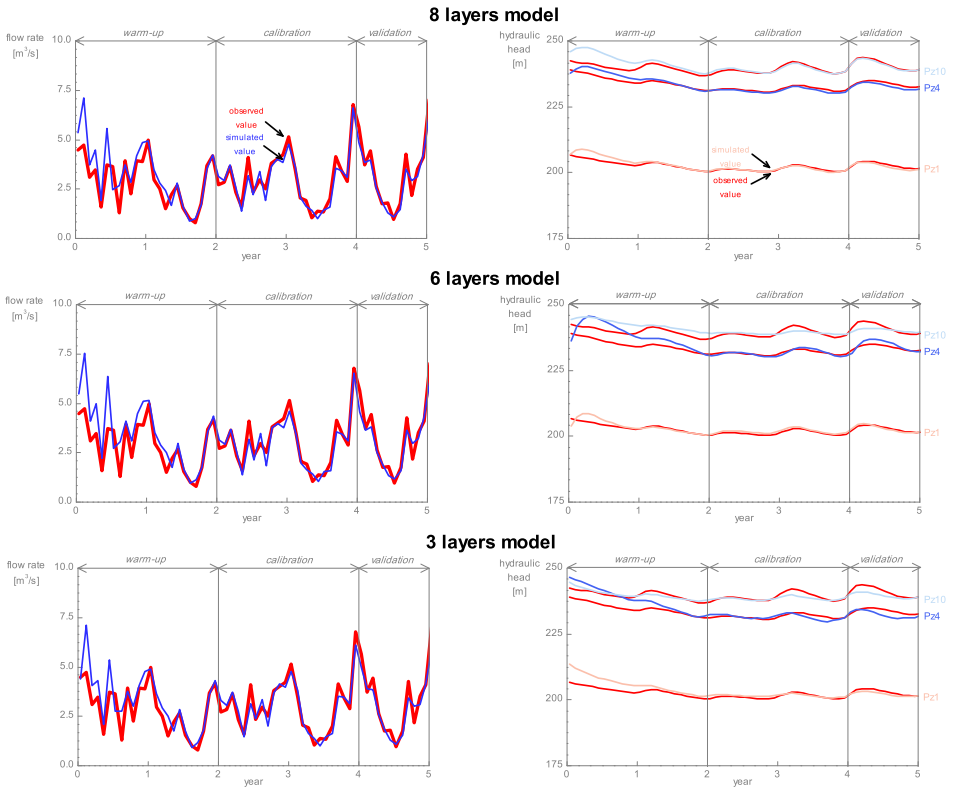

Figure 6.26: The optimisation significantly improves the simulation of hydraulic heads for the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors.

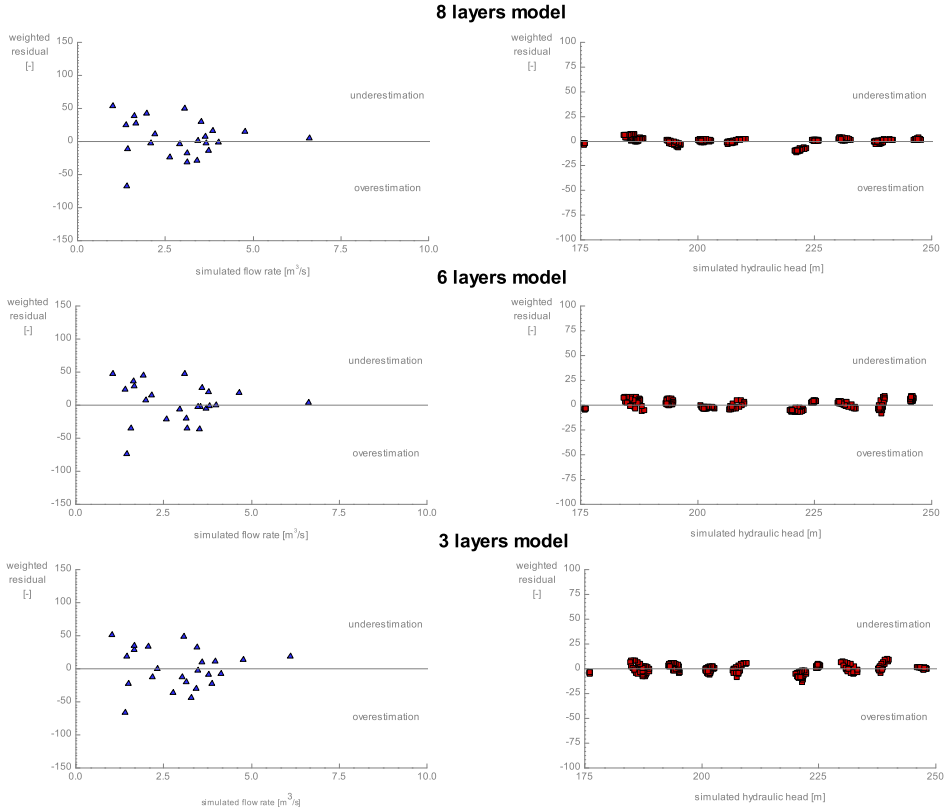

Figure 6.27: The weighted residuals of the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors are almost randomly distributed. This indicates that optimisation reduces model bias.

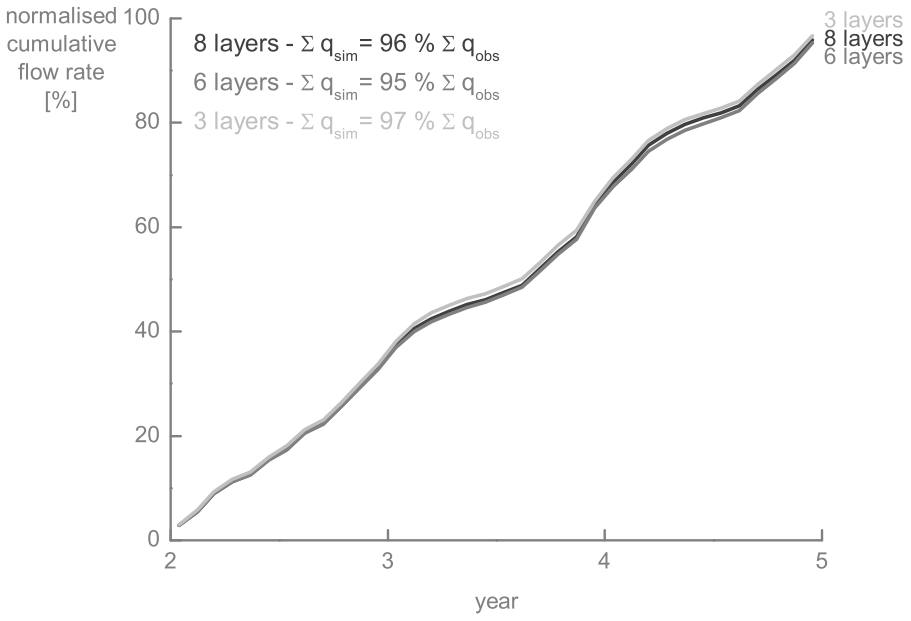

Figure 6.28: The mass conservation errors calculated for the optimised *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors are reduced to -4 %, -5 %, and -3 %, respectively.

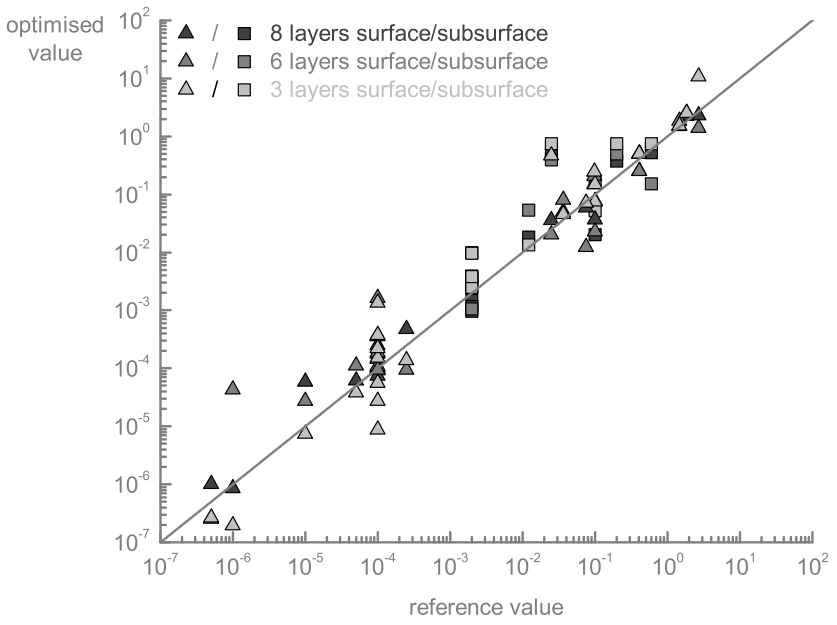

Figure 6.29: When vertical spatial discretisation gets coarser, optimised parameter values are in general further from their reference values.

6.4.3 Influence of the representation of the unsaturated zone

The execution times of the *van Genuchten* and *linear* models for a 5-year simulation with daily and average monthly stress factors are compared in Table 6.9. The mean relative gain in execution time for these models is 82 % which higlights once more the tremendous time saved when working with average monthly stress factors instead of daily stress factors.

Table 6.9: Comparison of the execution times of the *van Genuchten* and *linear* models for a 5-year simulation with average monthly stress factors and daily stress factors.

	Execution time	
	Daily stress factors	Average monthly stress factors
van Genuchten	34.88 h	6.08 h
linear	47.69 h	9.06 h

6.4.3.1 Forward model

The graphs comparing reference observations with their simulated equivalent indicate that both flow rates and hydraulic heads are poorly simulated by the *linear* model (Figure 6.30 and Figure B.26 to Figure B.30 in Appendix B). Flow rates are in general overestimated and hydraulic heads suffer from systematic errors and wrong variations. This is also visible on the graph of unweighted simulated values versus weighted residuals plotted for the calibration period (Figure 6.31). These findings are identical to those obtained for the same model with daily stress factors. Furthermore, the same severe problem of mass conservation occurs (Figure 6.32).

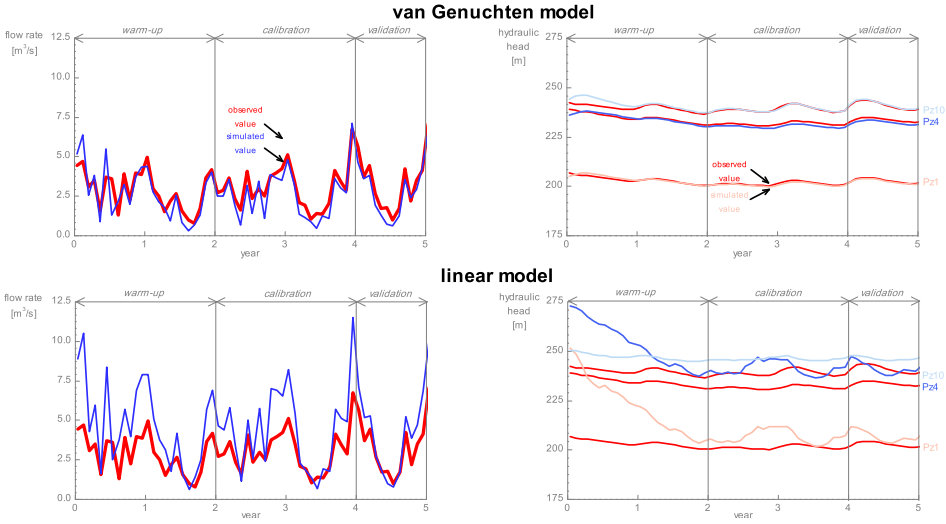

Figure 6.30: Flow rates and hydraulic heads simulated with the *van Genuchten* and *linear* models with monthly stress factors are compared with flow rate and hydraulic head observations produced with the 5-year reference simulation. These graphs indicate that both flow rates and hydraulic heads are poorly simulated by the model with linear saturation–pressure relations in the unsaturated zone.

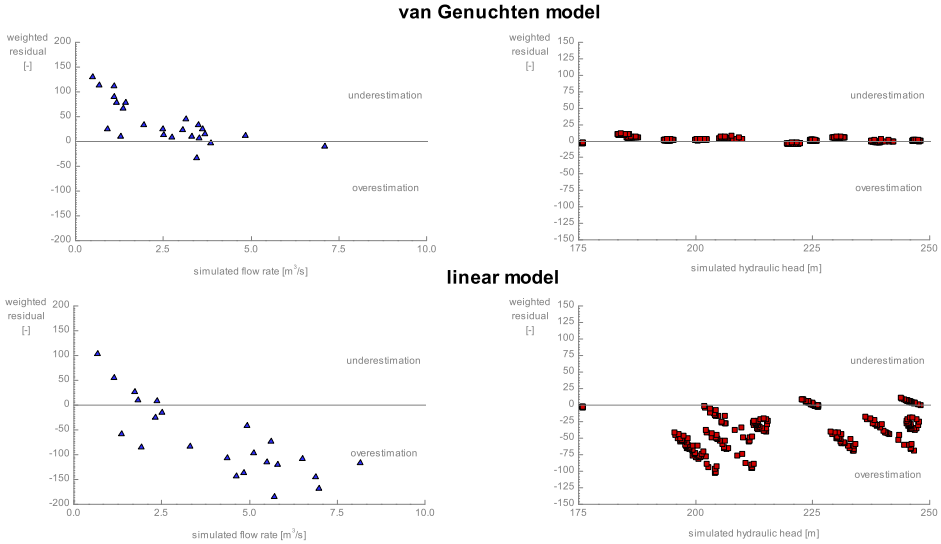

Figure 6.31: The graphs of unweighted simulated values versus weighted residuals for the *van Genuchten* and *linear* models with monthly stress factors show that both flow rates and hydraulic heads are poorly simulated by the model with linear saturation–pressure relations in the unsaturated zone.

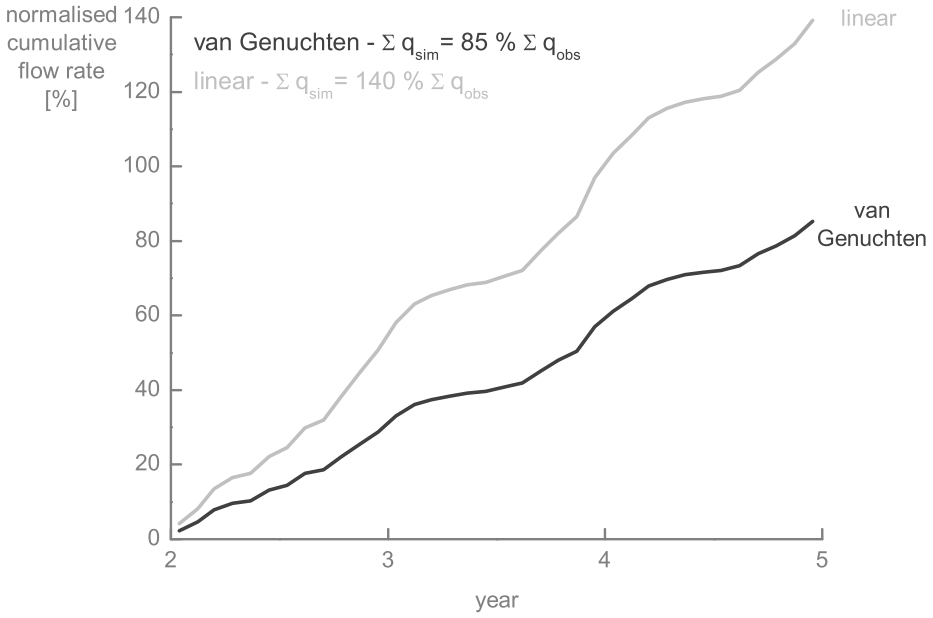

Figure 6.32: The mass conservation errors calculated for the *van Genuchten* and *linear* models with monthly stress factors are -15 % and 40 %, respectively.

6.4.3.2 Sensitivity analysis

The sensitivity analysis is performed on the calibration period. The sets of observations and parameters are identical to those used for the models with different horizontal spatial discretisations and average monthly stress factors, excepting van Genuchten parameters (α_{vG} and β_{vG}) which are replaced by their linear equivalent (h_a and h_b) for the *linear* model (24 flow rate observations, 288 hydraulic head observations, and 44 parameters).

The highest composite scaled sensitivities of the *van Genuchten* and *linear* models are in the same range of values, yet, their distribution is quite different (Figure 6.33). The css values for the *van Genuchten* model indicates that the van Genuchten parameter β_{vG} of Mat I and Mat II, the evapotranspiration fitting parameters $\theta_{e2\,\&\,t2}$ of Mat 1, Mat 2, and Mat 3, and, to a lesser extent, the saturated hydraulic conductivity K of Mat IV are comparatively more sensitive. The css values for the *linear* model suggests that no parameter is significantly more sensitive than the others. This is supported by the ratios of 3 % and 22% between maximum and minimum css values calculated for the *van Genuchten* and *linear* models, respectively. The higher value of this ratio for the *linear* model suggests that no parameter is predominant in terms of sensitivity. These results are similar to those obtained for the same models with daily stress factors.

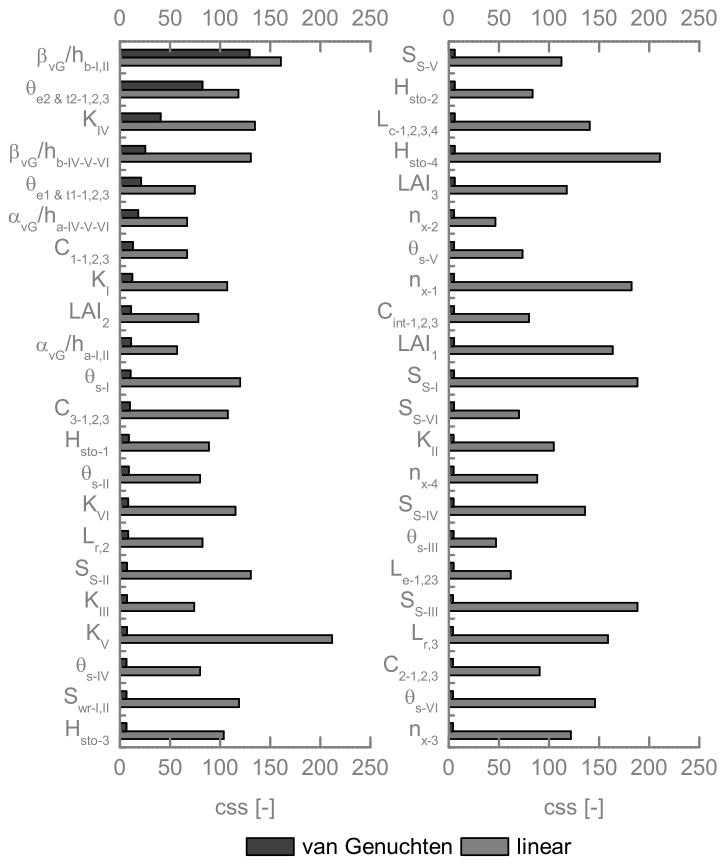

Figure 6.33: As for their equivalent with daily stress factors, composite scaled sensitivities calculated for the *van Genuchten* model with monthly stress factors indicate that the most sensitive parameters are related to flow in the unsaturated zone ($\beta_{vG-I,II}$) and evapotranspiration processes $\theta_{e2\ \&\ t2-1,2,3}$. This is not the case of the *linear* model for which no parameter is predominant in terms of sensitivity.

Whatever the representation of the unsaturated zone, van Genuchten or linear saturation–pressure relations, the maximum absolute value of *pcc* is lower than the critical value of 0.95.

6.4.3.3 Inverse model

The optimisation is performed on the calibration period. The sets of observations and parameters are identical to those used for the models with dif-

ferent horizontal spatial discretisations and average monthly stress factors, excepting once more van Genuchten parameters (α_{vG} and β_{vG}) which are replaced by their linear equivalent (h_a and h_b) for the *linear* model (24 flow rate observations, 288 hydraulic head observations, and 32 parameters).

The graphical model fit analysis indicates that optimisation significantly improves the simulation of flow rates and, to a lesser extent, hydraulic heads (Figures 6.34, 6.35, and 6.36 and Figure B.41 to Figure B.45 in Appendix B). However, flow rates and hydraulic heads are still poorly simulated by the *linear* model. Furthermore, given its very poor initial model fit, this model is improved uniquely through a significant modification of almost each parameter value (Figure 6.37 and Table B.3 in Appendix B).

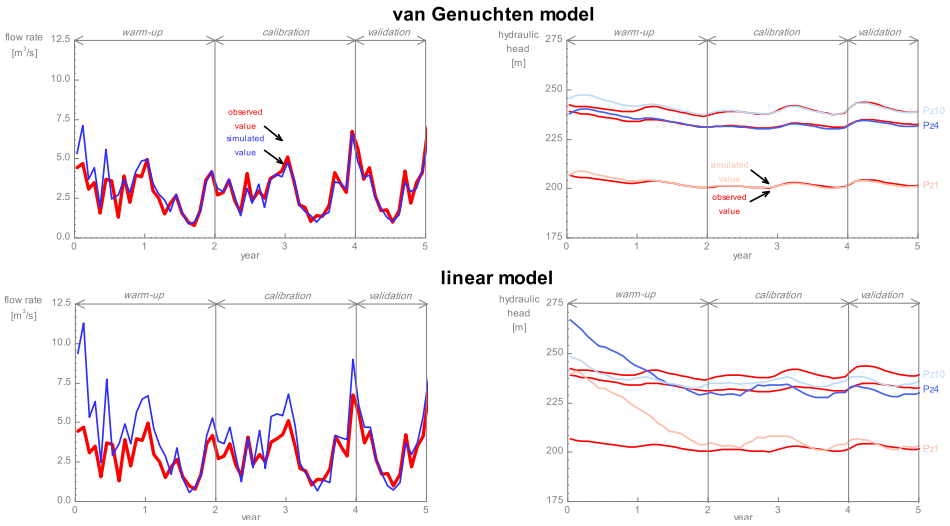

Figure 6.34: The optimisation significantly improves the simulation of flow rates and, to a lesser extent, hydraulic heads for the *van Genuchten* and *linear* models with monthly stress factors. However, the *linear* model is still poor.

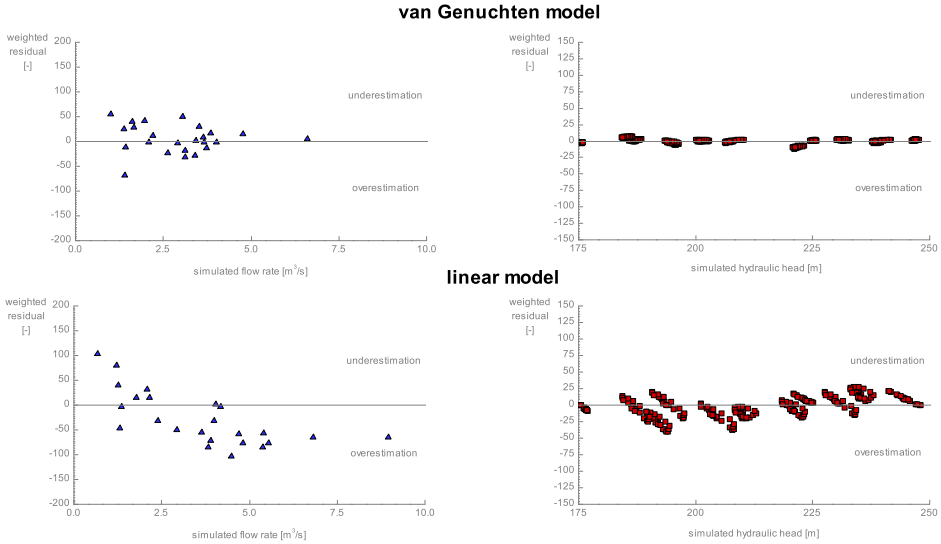

Figure 6.35: The optimisation strongly reduces model bias. This is higlighted by weighted residuals almost randomly distributed. However, the spread of weighted residuals for the *linear* model is still wide.

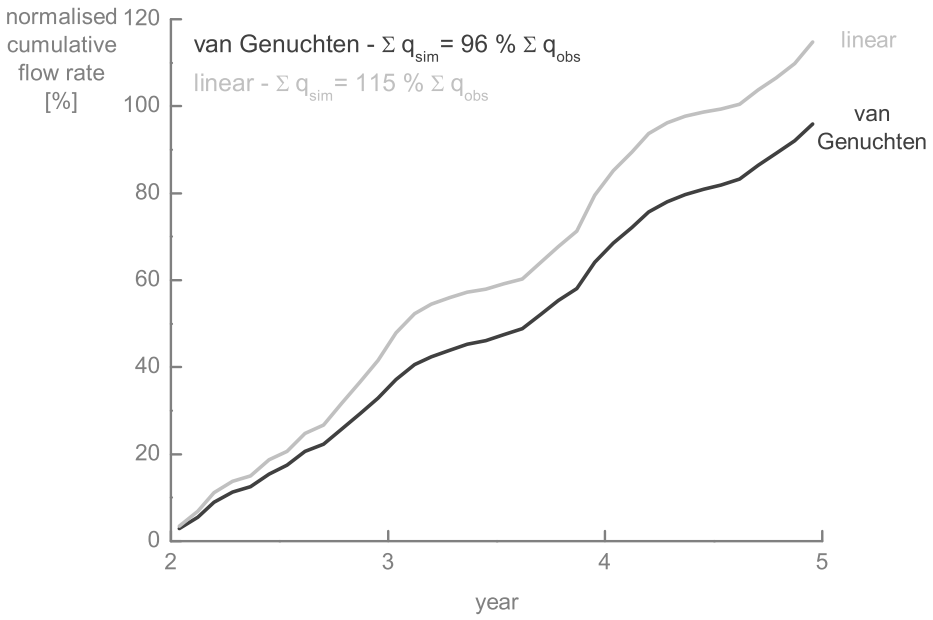

Figure 6.36: The mass conservation errors calculated for the optimised *van Genuchten* and *linear* models with monthly stress factors are reduced to -4 % and 15 %, respectively.

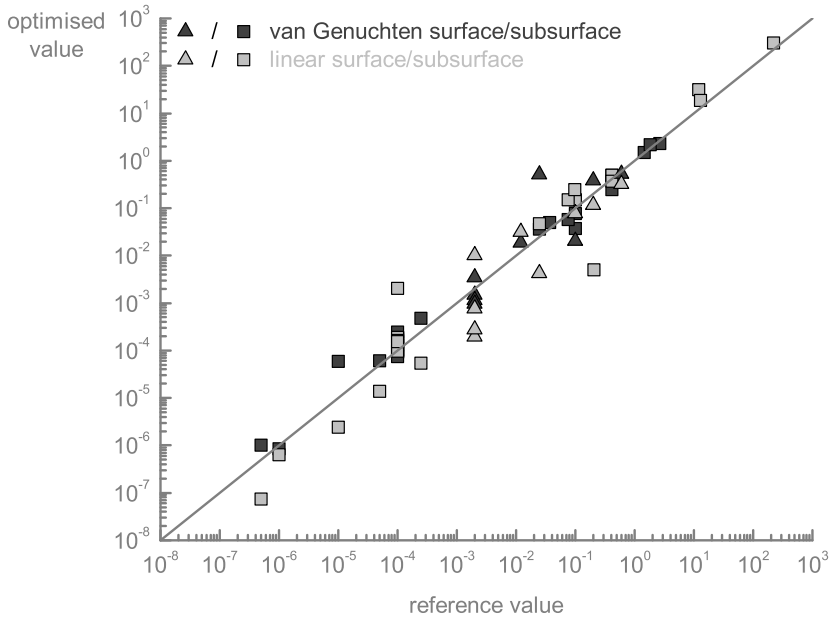

Figure 6.37: The optimisation of the *linear* model induces a significant modification of almost each parameter value.

6.5 Conclusions

A series of interesting findings concerning the influence of some model technical and structure uncertainties on model results is gained from this chapter. These findings also concern parameter sensitivities and execution times. When horizontal spatial discretisation gets coarser, flow rates are poorly simulated. This is highlighted in this work by the underestimation of low flow rates and the overestimation of high flow rates. Additionally, the likelihood that simulated hydraulic heads are shifted with respect to observed hydraulic heads is higher. When vertical spatial discretisation gets coarser, hydraulic heads are poorly simulated. This is visible in this work to the shift between observed and simulated hydraulic heads and to the wrong hydraulic

head variations simulated. Additionally, the overestimation of high flow rates is higher. When linear saturation–pressure relations are used instead of van Genuchten functions in the unsaturated zone, both flow rates and hydraulic heads are poorly simulated. These errors induced by model technical and structure uncertainties are partially compensated via optimisation. However, especially for the coarsest models, the improvement of model fit is limited and requires a significant modification of parameter values leading to unplausible parameter values. This shows the importance of systematically checking optimised parameter values against field data for making sure that they are plausible. All these findings are common to models with daily and average monthly stress factors.

Given the values of composite scaled sensitivity obtained, neither horizontal spatial discretisation nor vertical spatial discretisation influence significantly parameter sensititivies. The most sensitive parameters are always related to flow in the unsaturated zone and evapotranspiration processes. Additionally, excepting higher values for the most sensitive parameters of the *3 layers* model, composite scaled sensitivities are almost identical for each model. Furthermore, parameter correlation coefficient values obtained for each of model of this type are lower than the critical value of 0.95. The situation is quite different with linear saturation–pressure relations in the unsaturated zone. The linearisation of van Genuchten functions induces indeed an increase of sensitivity for the less sensitive parameters and a decrease of sensitivity for the most sensitive parameters. Consequently, there is no more predominant parameters in terms of sensitivity. However, parameter correlation coefficient values obtained for this model are still lower than the critical value of 0.95. All these observations, excepting higher parameter correlation coefficient for the *3 layers* model with monthly stress factors, are once more

common to models with daily and average monthly stress factors.

Whatever the time resolution of stress factors, similar observations are made concerning model results, sensitivity analysis, and optimisation performance. However, the gain in execution time is tremendous when working with average monthly stress factors instead of daily stress factors. Consequently, for studies which are satisfied with model results at such a time resolution, this is an efficient way of saving time.

Chapter 7

Quantitative evaluation of the simplified models

7.1 Key performance criteria selected for quantitative evaluation of model performance

A quantitative evaluation of model fit is essential for objectivising any qualitative evaluation performed with graphs. This quantitative evaluation requires to select key performance criteria for each variable of interest. The key performance criteria selected in this work are the Nash-Sutcliffe efficiency NSE_q for flow rates (Equation 2.42) and the root mean square error RMS_h for hydraulic heads (Equation 2.49). They are selected for their capacity of providing a global quantification of model performance for flow rates and hydraulic heads, respectively. These criteria are in fact closely related (Weglarczyk, 1998; Gupta et al., 2009):

$$NSE_y = 1 - \frac{\sum_{t=1}^{nt} \left(y_t^{sim} - y_t^{obs}\right)^2}{\sum_{t=1}^{nt} \left(y_t^{obs} - \mu^{obs}\right)^2} = 1 - \left(\frac{RMS_y}{\sigma^{obs}}\right)^2 \tag{7.1}$$

where nt is the total number of timesteps, y_t^{sim} is the simulated variable at timestep t, y_t^{obs} is the observed variable at timestep t, μ^{obs} is the mean of the observed values, RMS_y is the root mean square error calculated for the variable y, and σ^{obs} is the standard deviation of the observed values. As previously explained (see Section 2.3), NSE_y is a quite complete criterion often recommended for evaluating model performance in terms of flow rate simulation (Legates and McCabe, 1999; Moriasi et al., 2007). Consequently, some might want to use it for hydraulic heads. However, since hydraulic head variances are far lower than flow rate variances, normalising RMS_y by the variances of the observed values like in NSE_y could lead to the underestimation of model skill for hydraulic heads. Furthermore, calculating the mean value of RMS_y from the individual values obtained for each piezometer is a convenient and straightforward way of evaluating global model performance in terms of hydraulic head simulation. Therefore, NSE_q is preserved for flow rates and RMS_h is preferred for hydraulic heads. These key criteria are used in conjunction with other criteria for a thorough evaluation of the model performance. For flow rates, the other chosen criteria are the mass balance error MBE_q (Equation 2.45), the peak error PE_q (Equation 2.47), and the time peak error TPE_q (Equation 2.48). For hydraulic heads, the other criteria are the bias MBE_h and the hydraulic head variation error $HHVE_h$. Calculation of MBE_h is similar to the calculation of MBE_q, except that q is replaced by h. The calculation of $HHVE_h$ is based on the calculation of PE_q:

$$HHVE_h = \left(\frac{h_{max}^{sim} - h_{min}^{sim}}{h_{max}^{obs} - h_{min}^{obs}} - 1 \right) \times 100 \tag{7.2}$$

where h_{max}^{sim} is the maximum simulated hydraulic head value, h_{min}^{sim} is the minimum simulated hydraulic head value, h_{max}^{obs} is the maximum observed hy-

draulic head value, and h_{min}^{obs} is the minimum observed hydraulic head value. This criterion is the counterpart of the peak error criterion since it measures the capacity of the model to reproduce the magnitude of hydraulic head variations instead of measuring the capacity of the model to reproduce the peak in the hydrograph.

The key criteria NSE_q and RMS_h give a global evaluation of model performance in terms of flow rates and hydraulic heads, respectively. The other criteria are used for further investigating model issues in terms of mass balance/bias (MBE_q and MBE_h) and reproduction of peak flow rates (PE_q and TPE_q) and hydraulic head variations ($HHVE_h$). These criteria are calculated on the calibration period for the forward models. They are also calculated on the validation period for the inverse models in order to evaluate the performance of the models outside their calibration period.

7.2 Quantitative evaluation of models with daily stress factors

7.2.1 Influence of horizontal spatial discretisation

As horizontal spatial discretisation gets coarser, NSE_q values decrease and RMS_h values slightly increase (Figure 7.1). This indicates that simulation of both flow rates and hydraulic heads is impaired when using large elements. However, flow rate deterioration looks progressively increasing while hydraulic head deterioration looks rather stabilising when horizontal spatial discretisation is coarsened.

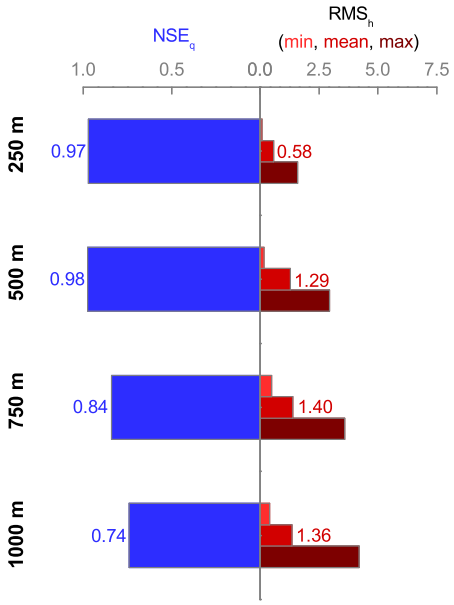

Figure 7.1: When horizontal spatial discretisation gets coarser, NSE_q values are lower and RMS_h values are slighlty higher. This indicates that flow rate and, to a lesser extent, hydraulic head simulation is progressively impaired when horizontal spatial discretisation is coarsened.

Thanks to Gupta's decomposition of NSE_q (Equation 2.44), it is possible to know which component of the hydrograph is poorly simulated (see Section 2.3). The progressive increase of Gupta's second term values for models with a coarse horizontal spatial discretisation suggests that they overestimate the standard deviation of flow rates (Table 7.1). This is supported by the increasing values of PE_q^1 and PE_q^2 indicating that models with large elements overestimates peak flow rates, so, the standard deviation of the hydrograph (Table 7.2). This is particularly the case for the *1000 m* model whose simulated peak flow rate is almost 100 % greater than the observed peak flow rate in the first year of the calibration period. The values of TPE_q^1 and TPE_q^2 show that observed and simulated peak flow rates are not shifted by more than one day. Although the absolute values of MBE_q are quite low, the transition from neg-

ative values to positive values suggests that the average magnitude of flow rates is underestimated by models with a fine horizontal spatial discretisation and overestimated by models with a coarse horizontal spatial discretisation. This is related to the fact that the underestimation of low flow rates is progressively compensated by the overestimation of high flow rates when the horizontal spatial discretisation gets coarser.

Table 7.1: Gupta's term of NSE_q calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors.

	250 m	*500 m*	*750 m*	*1000 m*
NSE_q [-]	0.97	0.98	0.84	0.74
Gupta's 1st term [-]	2.15	2.26	2.61	2.70
Gupta's 2nd term [-]	1.17	1.28	1.77	1.95
Gupta's 3rd term [-]	0.01	0.00	0.00	0.01

Table 7.2: Values of MBE_q, PE_q^1, TPE_q^1, PE_q^2, and TPE_q^2 calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors. PE_q^1 and TPE_q^1 correspond to PE_q and TPE_q for the 1st year of the calibration period. PE_q^2 and TPE_q^2 correspond to PE_q and TPE_q for the 2nd year of the calibration period.

	250 m	*500 m*	*750 m*	*1000 m*
MBE_q [%]	-7.87	-0.78	5.14	8.12
PE_q^1 [%]	-2.87	10.65	54.69	98.13
TPE_q^1 [day]	0	0	-1	-1
PE_q^2 [%]	12.26	11.26	36.34	39.76
TPE_q^2 [day]	-1	-1	-1	-1

Whatever the horizontal spatial discretisation, the absolute values of MBE_h are in general quite low (Table 7.3 and Table C.1 in Appendix C). This indicates that the models are not so biased in terms of hydraulic heads. However, the ranges of MBE_h values are in general wider for models with a coarse horizontal spatial discretisation. This suggests that the magnitudes of hydraulic heads are more likely to be poorly simulated when horizontal spatial

discretisation gets coarser. The absolute values of $HHVE_h^1$ and $HHVE_h^2$ are also in general quite low. This indicates that hydraulic head variations are fairly simulated.

Table 7.3: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors. $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	250 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	0.23	-0.03	0.41	-0.41	0.13	0.72
$HHVE_h^1$ [%]	3.36	19.69	-15.41	-45.26	-4.36	19.69
$HHVE_h^2$ [%]	-0.39	7.27	4.04	-43.33	-3.99	7.27
	500 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.24	-0.67	0.01	-1.32	0.15	1.57
$HHVE_h^1$ [%]	28.57	21.26	-7.89	-47.37	-0.22	28.57
$HHVE_h^2$ [%]	29.34	9.55	10.43	-48.33	29.34	-2.27
	750 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-1.78	-1.43	0.22	-1.78	-0.10	0.90
$HHVE_h^1$ [%]	0.84	15.75	-12.90	-29.41	-6.53	15.75
$HHVE_h^2$ [%]	-12.74	5.91	5.74	-36.36	-5.15	16.67
	1000 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	0.43	0.17	-0.20	-2.26	-0.37	0.94
$HHVE_h^1$ [%]	6.72	12.60	-3.23	-58.06	-9.17	12.84
$HHVE_h^2$ [%]	9.65	8.64	2.34	-63.07	-8.35	9.65

According to the quantitative evaluation performed, flow rates are more sensitive to horizontal spatial discretisation than hydraulic heads. This confirms the qualitative evaluation of model fit performed with graphs. While the underestimation of low flow rates is common to each model, the overestima-

tion of high flow rates is only a problem for the coarsest models. The underestimation of flow rates during low flow periods is probably related to a poor representation of the surface water network. This poor representation prevents from simulating properly groundwater–surface water interactions which constitute the key component of the hydrograph during dry seasons. This is already the case for the *250 m* model which suggests that refining the mesh along the surface water network is necessary for simulating accurately groundwater–surface water interactions. This problem of poor representation of the surface water network was previously mentioned by Refsgaard (1997) and Vázquez et al. (2002). The overestimation of flow rates during high flow periods is rather related to the smoothing of surface slopes when horizontal spatial discretisation gets coarser. This smoothing facilitates runoff during wet seasons. Additionally, surface and subsurface heterogeneities are poorly represented by the coarsest models which also influence runoff and infiltration.

7.2.2 Influence of vertical spatial discretisation

As vertical spatial discretisation gets coarser, RMS_h values strongly increase and NSE_q values slightly decrease (Figure 7.2). This indicates that hydraulic heads and, to a lesser extent, flow rates are impaired when the number of layers representing the unsaturated zone decreases.

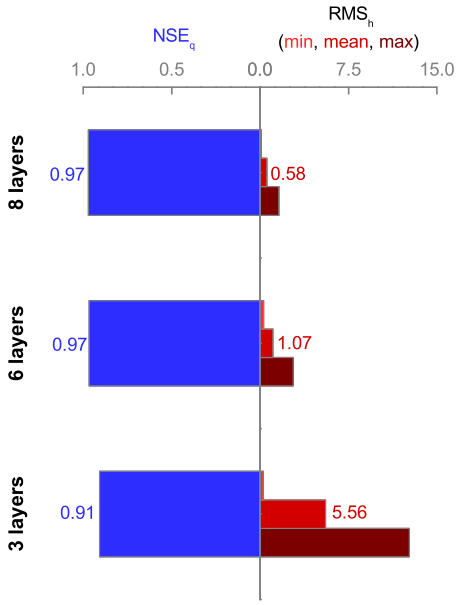

Figure 7.2: When vertical spatial discretisation gets coarser, NSE_q values are slightly lower and RMS_h values are significantly higher. This indicates that hydraulic head and, to a lesser extent, flow rate simulation is progressively impaired when vertical spatial discretisation is coarsened.

Gupta's decomposition of NSE_q indicates that the standard deviation of flow rates is slightly overestimated by the model with the coarsest vertical spatial discretisation (Table 7.4). This is visible to the greater value of Gupta's second term. This is also supported by the greater and positive values of PE_q^1 and PE_q^2 calculated for this model which suggests that peak flow rates, and so the standard deviation of the hydrograph, are overestimated (Table 7.5). However, NSE_q, PE_q^1, and PE_q^2 values calculated for the model with the coarsest vertical spatial discretisation are far better than those obtained for models with a coarse horizontal spatial discretisation. The values of TPE_q^1 and TPE_q^2 show that observed and simulated peak flow rates are not shifted by more than one day, except for the second year of the calibration period for the model with the coarsest vertical spatial discretisation. This is just a

minor issue since the global timing of the hydrograph is fairly reproduced (Figure 6.5). The absolute values of MBE_q are quite low. The transition from negative values to positive values is consistent with the fact that each model underestimates low flow rates while only the coarsest model overestimates high flow rates.

Table 7.4: Gupta's term of NSE_q calculated for the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors.

	8 layers	*6 layers*	*3 layers*
NSE_q [-]	0.97	0.97	0.91
Gupta's 1st term [-]	2.15	2.10	2.33
Gupta's 2nd term [-]	1.17	1.12	1.40
Gupta's 3rd term [-]	0.01	0.02	0.02

Table 7.5: Values of MBE_q, PE_q^1, TPE_q^1, PE_q^2, and TPE_q^2 calculated for the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors.

	8 layers	*6 layers*	*3 layers*
MBE_q [%]	-7.87	-10.84	11.60
PE_q^1 [%]	-2.87	-4.97	33.84
TPE_q^1 [day]	0	0	-1
PE_q^2 [%]	12.26	9.70	19.46
TPE_q^2 [day]	-1	-1	-344

When vertical spatial discretisation gets coarser, the absolute values of MBE_h and the ranges of MBE_h values significantly increase (Table 7.6 and Table C.2 in Appendix C). This indicates that the magnitudes of hydraulic heads are poorly simulated by models with a few number of thick layers in the unsaturated zone. The same observations are made for the absolute values of $HHVE_h^1$ and $HHVE_h^2$ and the ranges of $HHVE_h^1$ and $HHVE_h^2$ values. This suggests that hydraulic head variations are also poorly simulated by models representing the unsaturated zone by a few number of thick layers.

Compared to the values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for models with a coarse horizontal spatial discretisation, the values obtained for the model with the coarsest vertical spatial discretisation are far worse.

Table 7.6: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors.

	8 layers					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	0.23	-0.03	0.41	-0.41	0.13	0.72
$HHVE_h^1$ [%]	3.36	19.69	-15.41	-45.26	-4.36	19.69
$HHVE_h^2$ [%]	-0.39	7.27	4.04	-43.33	-3.99	7.27
	6 layers					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.96	-0.49	0.77	-0.96	0.14	1.28
$HHVE_h^1$ [%]	10.92	43.31	-46.24	-46.24	0.77	43.31
$HHVE_h^2$ [%]	-8.11	48.18	-33.19	-33.19	2.23	48.18
	3 layers					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	3.91	4.21	1.69	-0.12	2.68	6.49
$HHVE_h^1$ [%]	139.50	114.17	-38.35	-87.37	20.98	139.50
$HHVE_h^2$ [%]	51.35	79.09	-35.96	-85.00	5.80	82.35

Contrary to the horizontal spatial discretisation, the quantitative evaluation indicates that hydraulic heads are more sensitive to vertical spatial discretisation than flow rates. This is supported by the qualitative evaluation of model fit performed with graphs. When the unsaturated zone is coarsely discretised, both magnitudes and variations of hydraulic heads are poorly represented. This suggests that a fine vertical discretisation is required for simulating accurately infiltration through the unsaturated zone. Additionally, simulation of groundwater–surface water interactions is also impaired when the unsaturated zone is coarsely discretised.

7.2.3 Influence of the representation of the unsaturated zone

When linear saturation–pressure relations are used instead of van Genuchten relations in the unsaturated zone, the value of NSE_q decreases and the value of RMS_h increases significantly (Figure 7.3). This indicates that simulation of both flow rates and hydraulic heads is severely impaired.

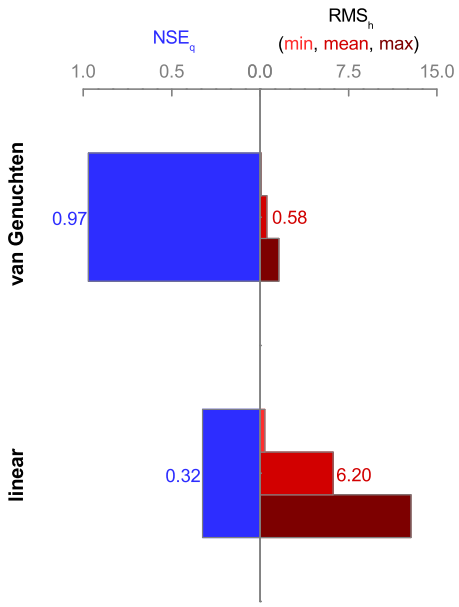

Figure 7.3: When linear saturation–pressure relations are used, NSE_q values are low and RMS_h values are high. This indicates that both flow rates and hydraulic heads are poorly simulated by models with linear saturation–pressure relations in the unsaturated zone.

Gupta's decomposition of NSE_q indicates that both the mean and the standard deviation of flow rates are overestimated by the *linear* model (Table 7.7). This is visible to the greater values of Gupta's second and third terms. This is also supported by the greater and positive values of MBE_q, PE_q^1 and PE_q^2 calculated for this model which suggests that both the mean and the peak flow rates are overestimated by this model (Table 7.8). The values of MBE_q is particularly high and shows that the average magnitude of flow rates

is overestimated by almost 50%. The values of TPE_q^1 and TPE_q^2 show that observed and simulated peak flow rates are not shifted by more than one day, except for the second year of the calibration period for the *linear* model. However, once more, this is just a local issue since the global timing of the hydrograph is fairly reproduced by this model (Figure 6.10).

Table 7.7: Gupta's term of NSE_q calculated for the *van Genuchten* and *linear* models with daily stress factors.

	van Genuchten	*linear*
NSE_q [-]	0.97	0.32
Gupta's 1st term [-]	2.15	3.15
Gupta's 2nd term [-]	1.17	2.57
Gupta's 3rd term [-]	0.01	0.25

Table 7.8: Values of MBE_q, PE_q^1, TPE_q^1, PE_q^2, and TPE_q^2 calculated for the *van Genuchten* and *linear* models with daily stress factors.

	van Genuchten	*linear*
MBE_q [%]	-7.87	42.00
PE_q^1 [%]	-2.87	66.88
TPE_q^1 [day]	0	-1
PE_q^2 [%]	12.26	57.67
TPE_q^2 [day]	-1	-344

According to the values of MBE_h and the ranges of MBE_h values, the *linear* model poorly simulates the magnitudes of hydraulic heads (Table 7.9 and Table C.3 in Appendix C). Additionally, the values of $HHVE_h^1$ and $HHVE_h^2$ and the ranges of $HHVE_h^1$ and $HHVE_h^2$ values indicates that hydraulic head variations are most often significantly overestimated by the *linear* model.

Table 7.9: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *van Genuchten* and *linear* models with daily stress factors.

	van Genuchten					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	0.23	-0.03	0.41	-0.41	0.13	0.72
$HHVE_h^1$ [%]	3.36	19.69	-15.41	-45.26	-4.36	19.69
$HHVE_h^2$ [%]	-0.39	7.27	4.04	-43.33	-3.99	7.27
	linear					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	1.54	3.23	2.85	-0.58	2.69	6.65
$HHVE_h^1$ [%]	635.29	579.53	-29.75	-29.75	261.31	635.29
$HHVE_h^2$ [%]	348.65	415.91	-43.40	-43.40	189.02	417.65

The quantitative evaluation indicates that both flow rates and hydraulic heads are poorly simulated when linear saturation–pressure relations are used instead of van Genuchten relations in the unsaturated zone. This is supported by the qualitative evaluation of model fit performed with graphs. This suggests that (over)simplifying saturation–pressure relations in the unsaturated zones can severely impair model results. Therefore, such simplifications should be taken with caution.

7.3 Quantitative evaluation of models with average monthly stress factors

7.3.1 Influence of horizontal spatial discretisation

7.3.1.1 Forward model

As horizontal spatial discretisation gets coarser, NSE_q values tend to decrease and RMS_h values tend to increase (Figure 7.4). This indicates that simulation of both flow rates and hydraulic heads is impaired when using large

elements. These findings are identical to those obtained for the same models with daily stress factors. However, monthly NSE_q values are lower than daily NSE_q values. This is probably inherent to the use of average monthly stress factors together with average monthly observations. Another explanation is proposed by Moriasi et al. (2007) who suppose that this is related to increased sample sizes for daily observations. The same reasons explain why monthly RMS_h values are slightly greater than daily RMS_h values.

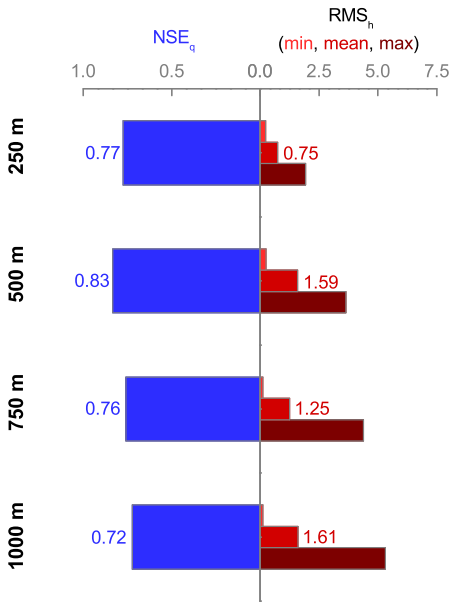

Figure 7.4: As for daily stress factors, NSE_q values are in general lower and RMS_h values are in general higher when horizontal spatial discretisation gets coarser. This indicates that the simulation of both flow rates and hydraulic heads is progressively deteriorated when horizontal spatial discretisation gets coarser.

Gupta's decomposition of NSE_q indicates that the standard deviation of flow rates is overestimated by models with a coarse horizontal spatial discretisation (Table 7.10). This is visible to the greater values of Gupta's second terms. This is also supported by the increasing positive values of PE_q^1 and PE_q^2 showing that models with large elements overestimate peak flow rates,

and so the standard deviation of the hydrograph (Table 7.11). The values of TPE_q^1 and TPE_q^2 indicate that observed and simulated peak flow rates are not shifted by more than two months. Gupta's decomposition shows also that the *250 m* model lacks to simulate properly the mean of flow rates. This is why NSE_q values for this model is lower than NSE_q values for the *500 m* model. This is confirmed by the values of MBE_q which shows that the *250 m* model underestimates the average magnitude of flow rates by almost 15 %. This is related to the fact that the underestimation of low flow rates is not compensated by the overestimation of high flow rates as it is the case for the other models.

Table 7.10: Gupta's term of NSE_q calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors.

	250 m	*500 m*	*750 m*	*1000 m*
NSE_q [-]	0.77	0.83	0.76	0.72
Gupta's 1st term [-]	2.26	2.33	2.53	2.59
Gupta's 2nd term [-]	1.37	1.47	1.77	1.87
Gupta's 3rd term [-]	0.11	0.03	0.00	0.00

Table 7.11: Values of MBE_q, PE_q^1, TPE_q^1, PE_q^2, and TPE_q^2 calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors.

	250 m	*500 m*	*750 m*	*1000 m*
MBE_q [%]	-14.32	-7.00	-1.32	1.23
PE_q^1 [%]	-8.65	-2.01	8.27	11.66
TPE_q^1 [month]	-2	-2	-2	-2
PE_q^2 [%]	5.04	11.82	20.27	21.83
TPE_q^2 [month]	0	0	0	0

Whatever the horizontal spatial discretisation, the absolute values of MBE_h are in general quite low (Table 7.12 and Table C.4 in Appendix C). This indicates that the models are not significantly biased in terms of hydraulic

heads. However, the ranges of MBE_h values are in general wider for models with a coarse horizontal spatial discretisation. This suggests that the magnitudes of hydraulic heads are more likely to be poorly simulated when horizontal spatial discretisation gets coarser. Although the absolute values of $HHVE_h^1$ and $HHVE_h^2$ are in general greater for models with a coarse horizontal spatial discretisation, the ranges of $HHVE_h^1$ and $HHVE_h^2$ values are similar for each model.

Table 7.12: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors.

	250 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.21	-0.54	0.01	-1.03	-0.17	0.50
$HHVE_h^1$ [%]	19.82	24.39	5.18	-91.67	-9.40	24.39
$HHVE_h^2$ [%]	-8.91	-2.35	-7.63	-89.96	-17.69	3.66
	500 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.66	-1.18	-0.36	-1.96	-0.15	1.18
$HHVE_h^1$ [%]	51.35	26.83	13.47	-91.67	-4.22	51.35
$HHVE_h^2$ [%]	24.70	-3.76	-0.22	-86.96	-12.62	24.70
	750 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-2.18	-1.89	-0.15	-2.18	-0.36	0.49
$HHVE_h^1$ [%]	12.16	19.51	3.63	-91.67	-12.46	19.51
$HHVE_h^2$ [%]	-19.84	-3.29	-3.92	-78.26	-19.83	2.76
	1000 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	0.04	-0.31	-0.49	-2.86	-0.62	0.96
$HHVE_h^1$ [%]	21.62	16.26	3.63	-98.04	-21.46	21.62
$HHVE_h^2$ [%]	0.00	-6.57	-1.53	-94.41	-24.97	10.98

All these observations are consistent with those made for the same mod-

els with daily stress factors. They are also consistent with the qualitative evaluation of model fit performed with graphs. However, particularly for NSE_q, the rate of deterioration when models gets coarser is greater for models with daily stress factors than for models with average monthly stress factors (Figure 7.5). This suggests that models with daily stress factors are more sensitive to horizontal spatial discretisation than models with average monthly stress factors.

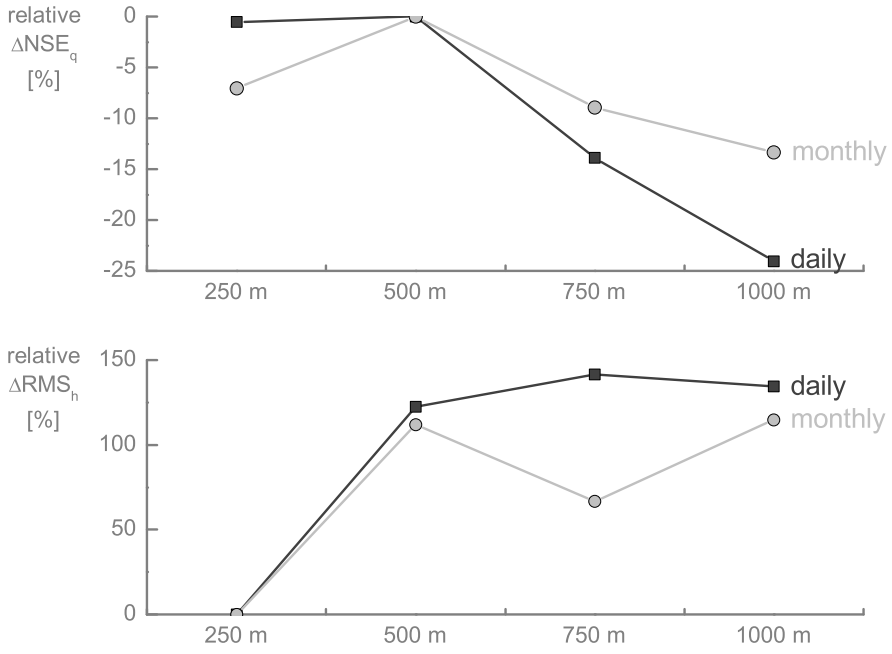

Figure 7.5: The rate of deterioration of NSE_q is greater for models with daily stress factors than for models with average monthly stress factors.

7.3.1.2 Inverse model

The optimisation significantly improves the simulation of both flow rates and hydraulic heads. This is visible to the greater values of NSE_q and the lower values of RMS_h (Figure 7.6). This indicates that optimisation can partially compensate for coarse horizontal spatial discretisation.

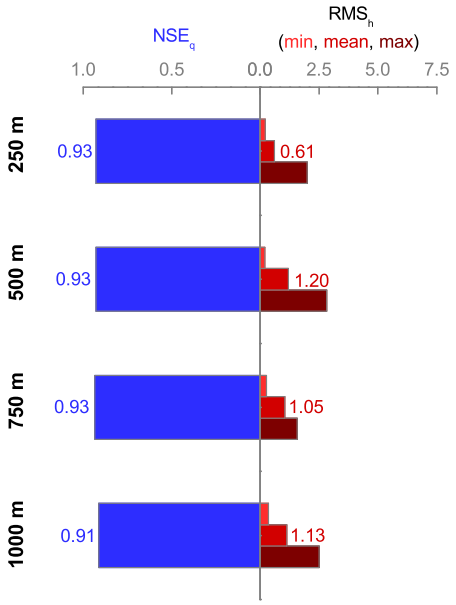

Figure 7.6: The higher values of NSE_q and the lower values of RMS_h indicate that the optimisation improves significantly the simulation of both flow rates and hydraulic heads.

The values of Gupta's terms together with the values of MBE_q, PE_q^1, and PE_q^2 (Tables 7.13 and 7.14) indicate that both the mean and the standard deviation of flow rates are better simulated.

Table 7.13: Gupta's term of NSE_q calculated for the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors.

	250 m	*500 m*	*750 m*	*1000 m*
NSE_q [-]	0.93	0.93	0.93	0.91
Gupta's 1st term [-]	1.87	2.01	1.87	1.91
Gupta's 2nd term [-]	0.94	1.08	0.94	1.00
Gupta's 3rd term [-]	0.01	0.00	0.00	0.00

Table 7.14: Values of MBE_q, PE_q^1, TPE_q^1, PE_q^2, and TPE_q^2 calculated for the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors.

	250 m	*500 m*	*750 m*	*1000 m*
MBE_q [%]	-3.44	-0.28	-1.82	0.26
PE_q^1 [%]	-4.81	-2.49	0.28	4.62
TPE_q^1 [month]	-1	-2	-1	-1
PE_q^2 [%]	-2.44	1.71	-2.42	-4.25
TPE_q^2 [month]	0	0	0	0

When observed and simulated hydraulic heads are shifted, the optimisation process strives for reducing this systematic error. This improvement of hydraulic head magnitudes is sometimes obtained to the detriment of the improvement of hydraulic head variations. This is why the absolute values of MBE_h and the range of MBE_h values are most often lower than those obtained with the forward models while the absolute values of $HHVE_h^1$ and $HHVE_h^2$ and the ranges of $HHVE_h^1$ and $HHVE_h^2$ values are most identical or even greater than those obtained with the forward models (Table 7.15 and Table C.7 in Appendix C). This highlights the limitations of the optimisation.

Table 7.15: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors.

	250 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.05	-0.20	0.00	-0.48	0.06	0.90
$HHVE_h^1$ [%]	3.60	20.33	-9.33	-95.24	-15.08	49.62
$HHVE_h^2$ [%]	8.91	17.37	-10.89	-88.46	-7.57	38.43
	500 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	0.38	0.29	0.00	-1.09	0.31	1.47
$HHVE_h^1$ [%]	98.20	16.26	-19.69	-85.71	-7.83	98.20
$HHVE_h^2$ [%]	93.93	-11.27	-23.75	-80.77	-3.78	93.93
	750 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.68	-0.62	0.25	-0.96	-0.04	0.79
$HHVE_h^1$ [%]	-5.41	49.59	-10.88	-36.11	5.14	52.74
$HHVE_h^2$ [%]	-23.08	35.21	-14.38	-28.45	18.92	96.15
	1000 m					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	0.05	0.39	-0.38	-1.72	-0.16	0.93
$HHVE_h^1$ [%]	61.26	14.63	-24.35	-98.04	-21.40	61.26
$HHVE_h^2$ [%]	63.16	-6.10	-19.83	-96.50	-16.89	63.16

As expected, NSE_q and RMS_h values calculated for the validation period are slightly lower and greater, respectively, than those obtained for the calibration period (Figure 7.7 and Table C.10 in Appendix C). This indicates that model performances are poorer outside their calibration period even when stress factors are almost identical. However, without optimisation, NSE_q and RMS_h values calculated for the validation period are even poorer. This suggests that optimisation also improves model performances outside their calibration period.

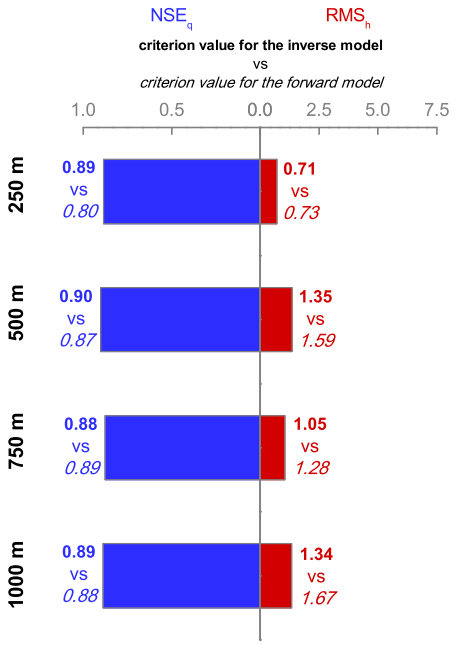

Figure 7.7: The values of NSE_q and RMS_h calculated for the validation period are slightly poorer than those obtained for the calibration period.

7.3.2 Influence of vertical spatial discretisation

7.3.2.1 Forward model

As vertical spatial discretisation gets coarser, RMS_h values strongly increase and NSE_q values slightly decrease (Figure 7.8). This indicates that hydraulic heads and, to a lesser extent, flow rates are impaired when the number of layers representing the unsaturated zone decreases. These findings are identical to those obtained for the same models with daily stress factors.

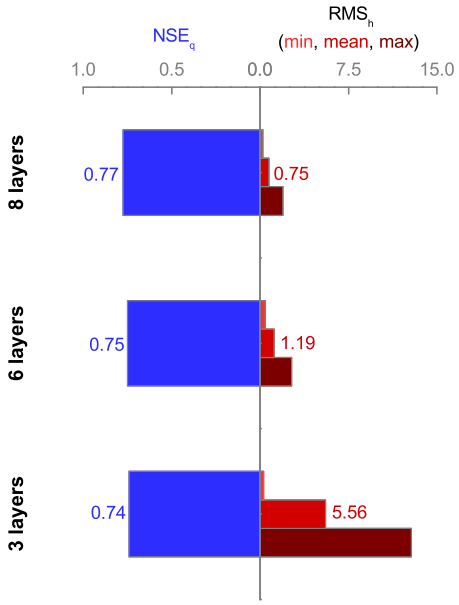

Figure 7.8: As for daily stress factors, when vertical spatial discretisation gets coarser, NSE_q values are slightly lower and RMS_h values are significantly higher when vertical spatial discretisation gets coarser. This indicates that the simulation of hydraulic heads, to a lesser extent, flow rates is progressively deteriorated when vertical spatial discretisation gets coarser.

Gupta's decomposition of NSE_q indicates that the standard deviation of flow rates is slightly overestimated by the model with the coarsest vertical spatial discretisation (Table 7.16). This is visible to the greater value of Gupta's second term. This is also supported by the greater and positive values of PE_q^1 and PE_q^2 calculated for this model which suggests that peak flow rates, and so the standard deviation of the hydrograph are overestimated (Table 7.17). The values of TPE_q^1 and TPE_q^2 show that observed and simulated peak flow rates are not shifted by more than two months. The absolute values of MBE_q are quite low. The transition from negative values to positive values is consistent with the fact that each model underestimates low flow rates while only the coarsest model overestimates high flow rates.

Table 7.16: Gupta's term of NSE_q calculated for the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors.

	8 layers	*6 layers*	*3 layers*
NSE_q [-]	0.77	0.75	0.74
Gupta's 1st term [-]	2.26	2.23	2.60
Gupta's 2nd term [-]	1.37	1.34	1.81
Gupta's 3rd term [-]	0.11	0.13	0.05

Table 7.17: Values of MBE_q, PE_q^1, TPE_q^1, PE_q^2, and TPE_q^2 calculated for the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors.

	8 layers	*6 layers*	*3 layers*
MBE_q [%]	-14.32	-16.04	9.31
PE_q^1 [%]	-8.65	-9.34	12.04
TPE_q^1 [month]	-2	-2	-2
PE_q^2 [%]	5.04	2.88	22.33
TPE_q^2 [month]	0	0	0

When vertical spatial discretisation gets coarser, the absolute values of MBE_h and the ranges of MBE_h values significantly increase (Table 7.18 and Table C.5 in Appendix C). This indicates that the magnitudes of hydraulic heads are poorly simulated by models with a few number of thick layers in the unsaturated zone. The same observations are made for the ranges of $HHVE_h^1$ and $HHVE_h^2$ values. This suggests that hydraulic head variations are also poorly simulated by models with a few number of thick layers in the unsaturated zone.

Table 7.18: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors.

	8 layers					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.21	-0.54	0.01	-1.03	-0.17	0.50
$HHVE_h^1$ [%]	19.82	24.39	5.18	-91.67	-9.40	24.39
$HHVE_h^2$ [%]	-8.91	-2.35	-7.63	-89.96	-17.69	3.66
	6 layers					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-1.32	-0.92	0.52	-1.32	-0.13	0.99
$HHVE_h^1$ [%]	15.32	41.46	-34.20	-88.89	-9.52	41.46
$HHVE_h^2$ [%]	-10.53	37.09	-39.43	-82.61	-12.14	37.09
	3 layers					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	4.06	4.23	1.61	-0.13	2.68	6.57
$HHVE_h^1$ [%]	66.67	43.09	-37.31	-97.22	-10.93	66.67
$HHVE_h^2$ [%]	25.51	49.30	-38.56	-92.31	-6.67	49.30

All these observations are once more consistent with those made for the same models with daily stress factors. They are also consistent with the qualitative evaluation of model fit performed with graphs. However, the rates of deterioration of both NSE_q and RMS_h when models gets coarser is greater for models with daily stress factors than for models with average monthly stress factors (Figure 7.9). This suggests that models with daily stress factors are more sensitive to vertical spatial discretisation than models with average monthly stress factors.

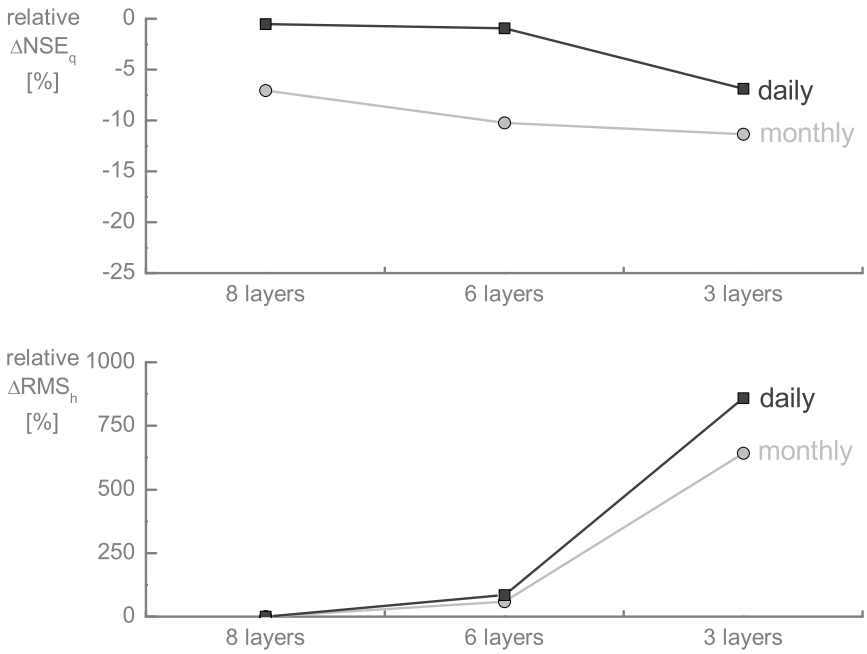

Figure 7.9: The rate of deterioration of both NSE_q and RMS_h is greater for models with daily stress factors than for models with average monthly stress factors.

7.3.2.2 Inverse model

The optimisation increases NSE_q values and decreases RMS_h values significantly (Figure 7.10). This indicates that optimisation can partially compensate for coarse vertical discretisation.

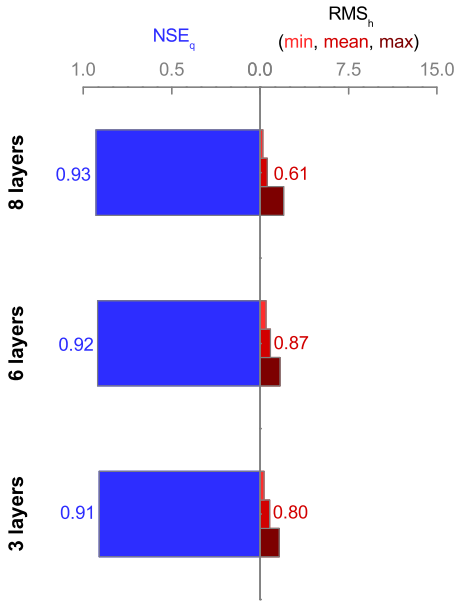

Figure 7.10: The higher values of NSE_q and the lower values of RMS_h indicate that the optimisation improves significantly the simulation of both flow rates and hydraulic heads.

Compared to those obtained with the forward models, the values of Gupta's terms together with the absolute values of MBE_q, PE_q^1, and PE_q^2 (Tables 7.19 and 7.20) are lower. This indicates that simulation of both the mean and the standard deviation are improved through optimisation.

Table 7.19: Gupta's term of NSE_q calculated for the optimised *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors.

	8 layers	*6 layers*	*3 layers*
NSE_q [-]	0.93	0.92	0.91
Gupta's 1st term [-]	1.87	1.83	1.76
Gupta's 2nd term [-]	0.94	0.90	0.85
Gupta's 3rd term [-]	0.01	0.01	0.00

Table 7.20: Values of MBE_q, PE_q^1, TPE_q^1, PE_q^2, and TPE_q^2 calculated for the optimised *8 layers*, *6 layers*, and emph3 layers models with average monthly stress factors.

	8 layers	*6 layers*	*3 layers*
MBE_q [%]	-3.44	-4.20	-2.23
PE_q^1 [%]	-4.81	-6.02	-2.04
TPE_q^1 [month]	-1	-1	-1
PE_q^2 [%]	-2.44	-3.83	-9.97
TPE_q^2 [month]	0	0	0

Compared to the those obtained with the forward models, the values of MBE_h, $HHVE_h^1$ and $HHVE_h^2$ show that the improvement of hydraulic head magnitudes is obtained to the detriment of the improvement of hydraulic head variations (Table 7.21 and Table C.8 in Appendix C). This highlights the limitations of the optimisation.

Table 7.21: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors.

	8 layers					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.05	-0.20	0.00	-0.48	0.06	0.90
$HHVE_h^1$ [%]	3.60	20.33	-9.33	-95.24	-15.08	49.62
$HHVE_h^2$ [%]	8.91	17.37	-10.89	-88.46	-7.57	38.43
	6 layers					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	0.28	-0.02	0.06	-0.67	-0.02	0.56
$HHVE_h^1$ [%]	29.73	45.53	-52.85	-80.56	-19.07	51.37
$HHVE_h^2$ [%]	12.15	64.32	-71.02	-71.02	-13.48	79.34
	3 layers					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	0.15	-0.06	-0.25	-0.25	0.09	0.67
$HHVE_h^1$ [%]	-36.04	43.09	-48.19	-94.44	-24.99	43.09
$HHVE_h^2$ [%]	-23.89	61.03	-54.03	-91.30	-22.15	61.03

The values of NSE_q and RMS_h calculated for the validation period are slightly lower and greater, respectively, than the values obtained for the calibration period (Figure 7.11 and Table C.11 in Appendix C)). This indicates that model performances are poorer outside their calibration period. However, NSE_q and RMS_h values calculated for the validation period are even poorer without optimisation. This suggests that optimisation also improves model performances outside their calibration period.

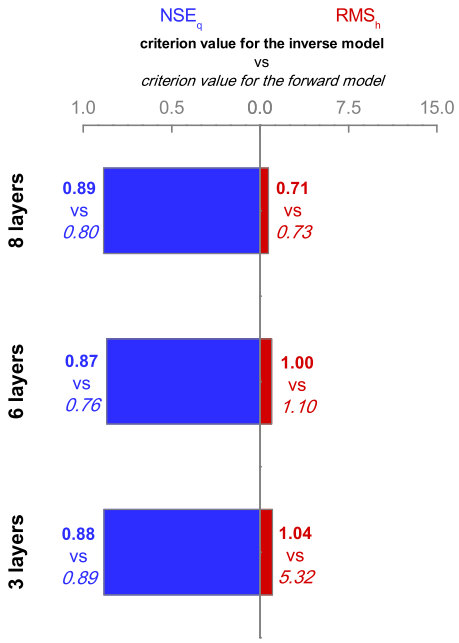

Figure 7.11: The values of NSE_q and RMS_h values calculated for the validation period are slightly poorer than those obtained for the calibration period.

7.3.3 Influence of the representation of the unsaturated zone

7.3.3.1 Forward model

When linear saturation–pressure relations are used instead of van Genuchten relations in the unsaturated zone, the value of NSE_q decreases and the value of RMS_h increases significantly (Figure 7.12). This indicates

that simulation of both flow rates and hydraulic heads is severely impaired. Furthermore, the negative NSE_q value for the *linear* model indicates that the mean of the observed values is a better predictor than the simulated values. These findings are identical to those obtained for the same model with daily stress factors.

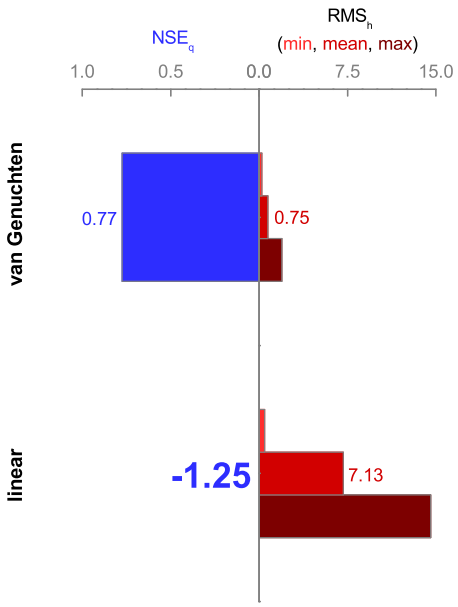

Figure 7.12: As for daily stress factors, NSE_q values are low and RMS_h values are high when linear saturation–pressure relations are used. This indicates that both flow rates and hydraulic heads are poorly simulated by models with linear relations in the unsaturated zone.

Gupta's decomposition of NSE_q indicates that both the mean and the standard deviation of flow rates are overestimated by the *linear* model (Table 7.22). This is visible to the greater values of Gupta's second and third terms. This is also supported by the greater and positive values of MBE_q, PE_q^1 and PE_q^2 calculated for this model which suggests that both the mean and the peak flow rates are overestimated by this model (Table 7.23). The values of TPE_q^1 and TPE_q^2 show that observed and simulated peak flow rates are not

shifted by more than two months.

Table 7.22: Gupta's term of NSE_q calculated for the *van Genuchten* and *linear* models with average monthly stress factors.

	van Genuchten	*linear*
NSE_q [-]	0.77	-1.25
Gupta's 1st term [-]	2.26	3.78
Gupta's 2nd term [-]	1.37	3.88
Gupta's 3rd term [-]	0.11	1.11

Table 7.23: Values of MBE_q, PE_q^1, TPE_q^1, PE_q^2, and TPE_q^2 calculated for the *van Genuchten* and *linear* models with average monthly stress factors.

	van Genuchten	*linear*
MBE_q [%]	-14.32	46.07
PE_q^1 [%]	-8.65	66.54
TPE_q^1 [day]	-2	-2
PE_q^2 [%]	5.04	70.04
TPE_q^2 [day]	0	0

According to the values of MBE_h and the ranges of MBE_h values, the *linear* model poorly simulates the magnitudes of hydraulic heads (Table 7.24 and Table C.6 in Appendix C). Additionally, the values of $HHVE_h^1$ and $HHVE_h^2$ and the ranges of $HHVE_h^1$ and $HHVE_h^2$ values indicate that hydraulic head variations are most often significantly overestimated by the *linear* model.

Table 7.24: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *van Genuchten* and *linear* models with average monthly stress factors.

	van Genuchten					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.21	-0.54	0.01	-1.03	-0.17	0.50
$HHVE_h^1$ [%]	19.82	24.39	5.18	-91.67	-9.40	24.39
$HHVE_h^2$ [%]	-8.91	-2.35	-7.63	-89.96	-17.69	3.66
	linear					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	2.77	4.18	2.95	-0.57	3.21	7.65
$HHVE_h^1$ [%]	600.90	593.50	-14.51	-94.44	215.44	600.90
$HHVE_h^2$ [%]	302.02	350.23	-45.10	-91.30	127.05	359.26

All these observations are once more consistent with those made for the same model with daily stress factors. They are also consistent with the qualitative evaluation of model fit performed with graphs. However, while the rate of deterioration of RMS_h is greater for models with daily stress factors, the rate of deterioration of NSE_q is greater for models with average monthly stress factors (Figure 7.13).

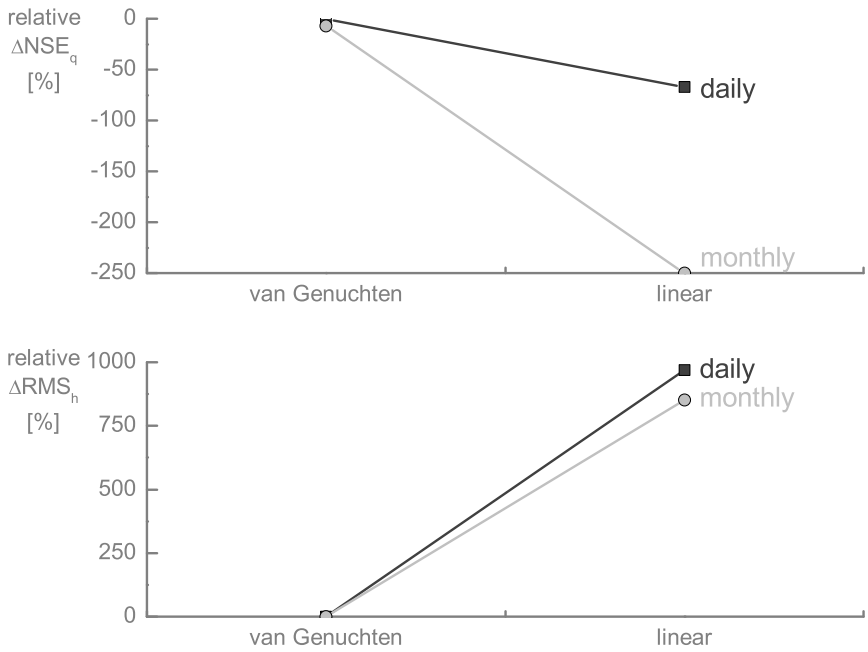

Figure 7.13: Contrarily to the rate of deterioration of RMS_h, the rate of deterioration of NSE_q is greater for models with average monthly stress factors than for models with monthly stress factors.

7.3.3.2 Inverse model

Although the optimisation significantly improves the simulation of both flow rates and hydraulic heads, the values of NSE_q and RMS_h values calculated for the *linear* model are still far worse than those obtained for the *van Genuchten* model (Figure 7.14). This indicates that optimisation can only compensate to a limited extent for this kind of simplification.

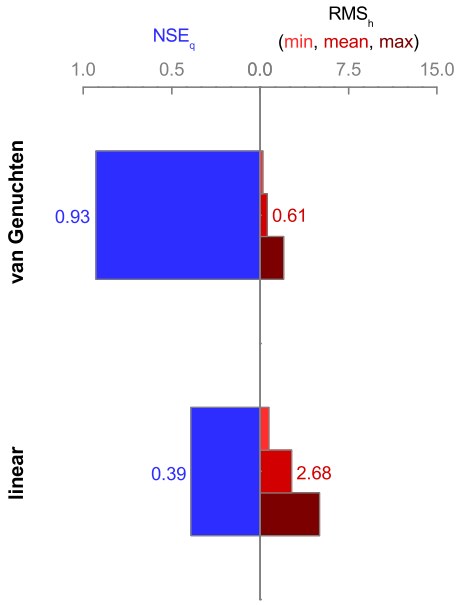

Figure 7.14: The simulation of both flow rates and hydraulic heads is significantly improved through optimisation. However, NSE_q and RMS_h values calculated for the *linear* model are still worse than those obtained for the *van Genuchten* model.

Compared to those obtained with the forward models, the values of Gupta's terms together with the absolute values of MBE_q, PE_q^1, and PE_q^2 (Tables 7.25 and 7.26) are lower. This indicates that simulation of both the mean and the standard deviation are improved through optimisation. However, the improvement is not sufficient for the *linear* model to equal the *van Genuchten* model.

Table 7.25: Gupta's term of NSE_q calculated for the optimised *van Genuchten* and *linear* models with average monthly stress factors.

	van Genuchten	*linear*
NSE_q [-]	0.93	0.39
Gupta's 1st term [-]	1.87	2.91
Gupta's 2nd term [-]	0.94	2.31
Gupta's 3rd term [-]	0.01	0.20

Table 7.26: Values of MBE_q, PE_q^1, TPE_q^1, PE_q^2, and TPE_q^2 calculated for the optimised *van Genuchten* and *linear* models with average monthly stress factors.

	van Genuchten	*linear*
MBE_q [%]	-3.44	19.68
PE_q^1 [%]	-4.81	31.42
TPE_q^1 [day]	-1	-1
PE_q^2 [%]	-2.44	32.53
TPE_q^2 [day]	0	0

Compared to those obtained with the forward models, the values of MBE_h, $HHVE_h^1$ and $HHVE_h^2$ show that the improvement of hydraulic head magnitudes is obtained to the detriment of the improvement of hydraulic head variations (Table 7.27 and Table C.9 in Appendix C). Furthermore, like for flow rates, the improvement of hydraulic heads is not sufficient for the *linear* model to equal the *van Genuchten* model.

Table 7.27: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the optimised *van Genuchten* and *linear* models with average monthly stress factors.

	van Genuchten					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	-0.05	-0.20	0.00	-0.48	0.06	0.90
$HHVE_h^1$ [%]	3.60	20.33	-9.33	-95.24	-15.08	49.62
$HHVE_h^2$ [%]	8.91	17.37	-10.89	-88.46	-7.57	38.43
	linear					
	Pz1	Pz4	Pz10	min	mean	max
MBE_h [%]	1.66	-0.35	-1.61	-1.61	0.31	2.55
$HHVE_h^1$ [%]	420.72	204.47	-10.45	-65.86	156.63	546.82
$HHVE_h^2$ [%]	190.69	214.08	6.32	-60.87	162.16	546.15

Contrary to the *van Genuchten* model, the *linear* model obtained greater values of NSE_q and lower values of RMS_h for the validation period than for the calibration period (Figure 7.15 and Table C.12 in Appendix C). This is prob-

ably related to the fact that stress factors of the validation period are more suited to the *linear* model than those of the validation period. However, without optimisation, NSE_q and RMS_h values calculated for the validation period are poorer for both the *van Genuchten* and the *linear* models.

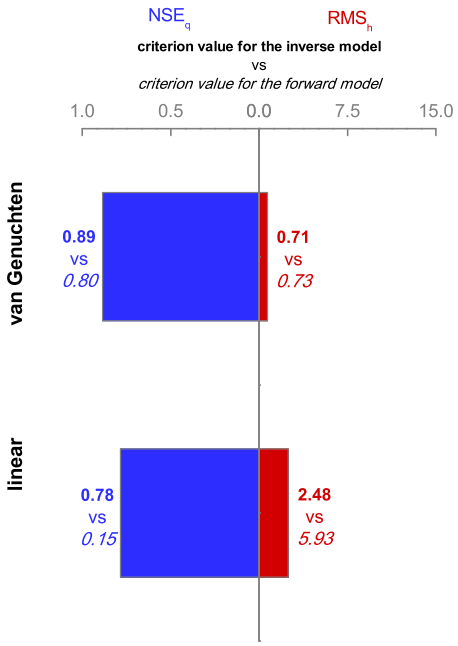

Figure 7.15: Contrary to the other models tested, NSE_q and RMS_h values calculated for the validation period with the *linear* model are better than those obtained for the calibration period. This is probably related to the fact that stress factors of the validation period are more suited to the *linear* model than those of the validation period.

7.4 Conclusions

As expected, the quantitative evaluation carried out in this chapter confirms the findings of the qualitative evaluation previously performed. Furthermore, in spite of slightly poorer performance for models with average monthly stress factors, observations made for models with daily stress factors are once more identical to those made for models with average monthly stress factors. However, the rate of deterioration of model performance when model gets

coarser is in general greater for models with daily stress factors than for models with average monthly stress factors. This suggests that models with daily stress factors are more sensitive to technical and structure uncertainties than models with average monthly stress factors.

When horizontal spatial discretisation is coarsened, the deterioration of flow rate simulation is predominant over the deterioration of hydraulic head simulation. Low flow rates are underestimated and high flow rates are overestimated. This leads to a poor representation of the variance of the hydrograph. Common to each model tested, the underestimation of low flow rates is probably due to a poor representation of the surface water network which prevents from simulating properly groundwater–surface water interactions. Proper to the coarsest models tested, the overestimation of high flow rates is rather related to the smoothing of surface slopes which facilitates runoff.

When vertical spatial discretisation is coarsened, the deterioration of hydraulic head simulation is predominant over the deterioration of flow rate simulation. Both magnitudes and variations of hydraulic heads are poorly simulated. This is probably due to the coarse discretisation of the unsaturated zone which prevents from simulating properly infiltration and groundwater–surface water interactions.

When linear saturation–pressure relations are used instead of van Genuchten functions in the unsaturated zone, the simulation of both flow rates and hydraulic heads is significantly deteriorated.

All the errors due to model technical and structure uncertainties are partially compensated via optimisation. However, in addition to the risk of unplausible optimised parameter values, performance criteria values obtained for the optimised models highlight that the improvement of simulated hydraulic head magnitudes potentially induce a deterioration of simulated hy-

draulic head variations. This is especially the case for models whose simulated hydraulic heads are significantly shifted with respect to observed hydraulic heads. This points out the limitations of the optimisation. Additionally, in spite of quite similar ranges of stress factors, performance criteria calculated for the validation period are in general slightly poorer than those calculated for the calibration period. This proves the necessity of calibrating models with ad hoc stress factors with respect to the objective of the study. However, without optimisation, model performances for the validation period are even poorer. Therefore, in spite of its limitations, optimisation is essential for improving model performance, either inside or outside the calibration period.

Chapter 8

Model ranking

8.1 Model ranking based on information theory

8.1.1 Posterior model probability issue

According to the information theory, posterior model probabilities p_i decrease exponentially with delta values Δ_i (Equation 2.58). This way of calculating posterior model probabilities quickly produces extremely low values and potentially leads to reject models that were looking quite good. This is quite usual with groundwater models (Poeter E.P., personal communication). Therefore, given this issue, linear posterior model probabilities are preferred in this work (Equation 2.60). This way of calculating posterior model probabilities produces inverted evidence ratios increasing linearly from 0 % for the worst model to 100 % for the best model in the set. Additionally, given that the number of observations $nobs$ and the number of parameters $npar$ are the same for each model, the objective function Φ is used instead of AIC or other criteria for calculating posterior model probabilities and associated statistics. This is supported by the fact that these criteria are only useful for models with different number of observations and/or number of parameters.

Therefore, the inverted evidence ratio IER is calculated as:

$$IER = \frac{\Phi_{max} - \Phi_j}{\Phi_{max} - \Phi_i} \tag{8.1}$$

As previously mentioned and provided that i is the best model in the set, this statistic is used to make statement such as “the evidence supporting model j is only IER percent of the evidence supporting the best model”.

8.1.2 Ranking of models with daily stress factors

Based on linear posterior model probabilities and the values of the objective function, the best model is the *500 m* model and the poorest model is the *linear* model (Figure 8.1). However, with evidence ratios greater than 95 %, the *250 m*, *750 m*, and *6 layers* models are almost equivalent in term of model fit since the evidence supporting these models are greater than 95 % of the evidence supporting the *500 m* model.

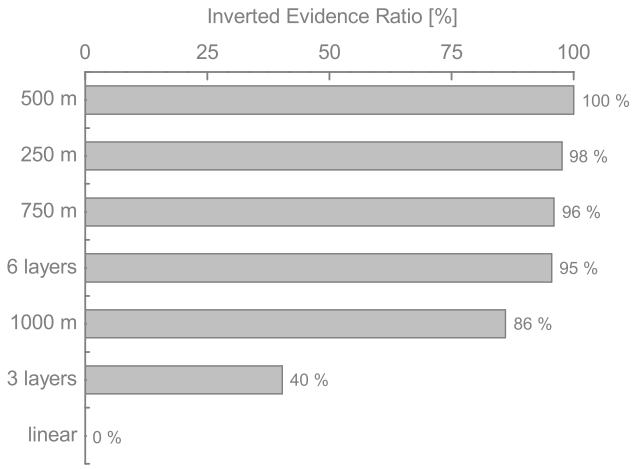

Figure 8.1: According to linear posterior model probabilities, the best model with daily stress factors is the *500 m* model and the poorest model is the *linear* model.

8.1.3 Ranking of models with average monthly stress factors

Compared to the ranking of models with daily stress factors, the ranking of models with average monthly stress factors is almost identical. The best model is the *250 m* model and the poorest model is the *linear* model (Figure 8.2). However, excepting the *linear* and the *3 layers*, the other models have inverted evidence ratios greater than 95 % which suggests that they are almost equivalent to the *250 m* model in term of model fit.

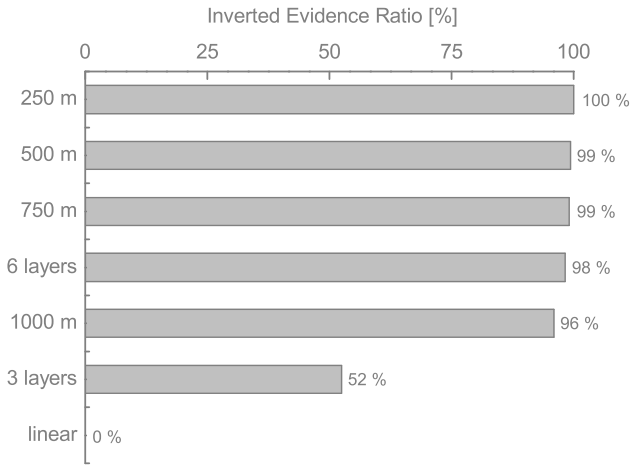

Figure 8.2: According to linear posterior model probabilities, the best model with average monthly stress factors is the *250 m* model and the poorest model is the *linear* model.

When optimised models are included in the ranking, the best models are the optimised *250 m*, *6 layers*, and *3 layers* models (Figure 8.3). Given the inverted evidence ratios, the other optimised models are almost equivalent in term of model fit, except the *linear* model which remains comparatively worse. This is quite surprising since some optimised models are clearly poorer than others when examining graphs of model fit and values of perfomance criteria. This is related to the sole use of the objective function for establishing this ranking. The strength of the objective function lies in its capacity of tak-

ing into account measurement error through weighting. The weakness of the objective function is related to its measure of the overall model fit regardless of the reproduction of key features such as measured by performance criteria. Therefore, a ranking system based on performance criteria is proposed for determining which models are capable of reproducing satisfactorily flow rates and hydraulic heads.

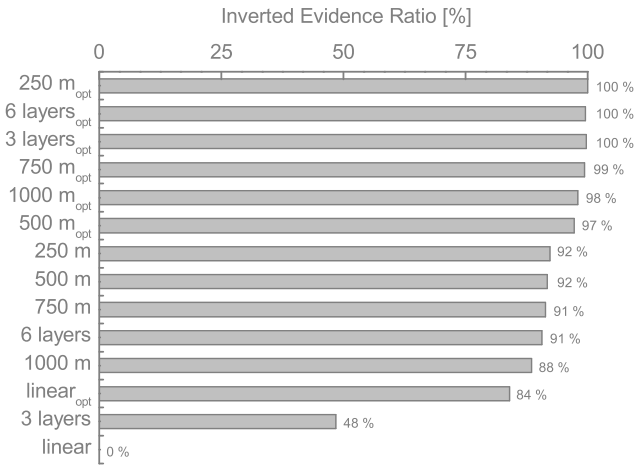

Figure 8.3: When optimised models are included in the ranking, they are the best ones. Furthermore, the inverted evidence ratios suggest that they are almost equivalent in term of model fit.

8.2 Model ranking based on performance criteria

8.2.1 Performance intervals and ranking system

Considering the shortcomings of model ranking based on information theory, a combined use of flow rate and hydraulic head performance criteria is proposed for ranking models. This ranking implies to define performance intervals for each criterion. A total of five intervals, from *very poor* to *very good*, is used in this work and scores, from 1 to 5 points, are associated with

each of them. These intervals are defined based on ranges of values found in the literature and typical expectations of large-scale models (for example, see Henriksen et al., 2003; Moriasi et al., 2007; Foglia et al., 2009) (Table 8.1). For flow rates, model simulation is judged as satisfactory if NSE_q is greater than 0.50, $\left|MBE_q\right|$ is less than 20 %, $\left|PE_q\right|$ is less than 25 %, and $\left|TPE_q\right|$ is less than 5 days for daily stress factors and 2 months for average monthly stress factors. For hydraulic heads, model simulation is judged as satisfactory if RMS_h is less than 1 m, $|MBE_h|$ is less than 0.50 %, and $|HHVE_h|$ is less than 25 %. This system results in scores from 1 to 5 points for each performance criteria and for each observation point (1 gauging station and 12 piezometers). These scores are averaged by type of observations for evaluating separately model performance in terms of flow rates and hydraulic heads. A number of stars, from 1 to 5, is used for symbolising the performance of each model obtained in this way. When a model obtains 3 stars for flow rates and 3 stars for hydraulic heads, it means that it simulates satisfactorily both flow rates and hydraulic heads.

Table 8.1: Performance intervals defined for ranking models from very poor to very good.

	Very poor	*Poor*	*Satisfactory*	*Good*	*Very good*
	1 point	2 points	3 points	4 points	5 points
NSE_q [-]	< 0.20	0.20-0.50	0.50-0.65	0.65-0.85	> 0.85
MBE_q [%]	> 40	20-40	10-20	5-10	< 5
PE_q [%]	> 50	25-50	10-25	5-10	< 5
TPE_q [day] or [month]	> 10 or > 3	5-10 or 3	3-5 or 2	1-3 or 1	< 1 or 0
RMS_h [m]	> 1.50	1.00-1.50	0.50-1.00	0.25-0.50	< 0.25
MBE_h [%]	> 1.00	0.50-1.00	0.25-0.50	0.10-0.25	< 0.10
$HHVE_h$ [%]	> 50	25-50	10-25	5-10	< 5

8.2.2 Ranking of models with daily stress factors

According to the performance intervals defined, only the *linear* model fails to simulate flow rates satisfactorily (Table 8.2). The other models get scores ranging from *satisfactory* for the *3 layers* model to *very good* for the *250 m* and the *6 layers* models. This ranking is consistent with the qualitative and quantitative analyses performed which highlighted a progressive deterioration of flow rate simulation when spatial discretisation gets coarser. The scores are slightly worse for hydraulic heads. This is probably related to more stringent performance thresholds. However, the ranking is quite similar. The *linear* and the *3 layers* models fail to simulate hydraulic heads satisfactorily. The other models obtain scores ranging from *satisfactory* for the *500 m*, *750 m*, *1000 m* and *6 layers* models to *good* for the *250 m* model. Once more, this is consistent with the progressive deterioration of hydraulic head simulation highlighted by the qualitative and quantitative analyses performed. Considering both flow rates and hydraulic heads, only the *linear* and the *3 layers* models are incapable of satisfying the performance levels defined. This indicates that horizontal spatial discretisation is less important than vertical spatial discretisation for obtaining satisfying results in terms of both flow rates and hydraulic heads. Although the scores obtained are highly dependent on the performance intervals defined, the fact that flow rates obtain in general slightly higher scores than hydraulic heads could suggest that variables representative of the integrated response of a model (e.g. flow rate at the outlet) are less sensitive to technical and structure uncertainties than variables representative of the distributed response of a model (e.g. hydraulic heads in several points).

Table 8.2: Ranking of models with daily stress factors based on performance intervals.

	Flow rates	Hydraulic heads
250 m	★★★★★	★★★★
500 m	★★★★	★★★
750 m	★★★★	★★★
1000 m	★★★★	★★★
6 layers	★★★★★	★★★
3 layers	★★★	★★
linear	★★	★

Considering simultaneously the number of stars obtained for flow rates and hydraulic heads, model performance is *good* for the *250 m* model and *satisfactory* for the *500 m*, *750 m*, and *1000 m* models. This indicates that model results are not significantly deteriorated from the *500 m* to the *1000 m* models. Converting these element sizes into a number of elements per square kilometer, this suggests that model results are not especially poorer with 2 elements per square kilometer than with 10 elements per square kilometer and that model results are only significantly improved from 40 elements per square kilometer. Vertical spatial discretisation require more caution since model performance are poor for the *3 layers* model and only *satisfactory* for the *6 layers* model. This indicates that model results are severely deteriorated when the unsaturated zone is coarsely discretised. The same severe deterioration of model results is obtained when (over)simplifying saturation–pressure relations in the unsaturated zone.

8.2.3 Ranking of models with average monthly stress factors

Compared to the ranking of models with daily stress factors, the ranking of models with average monthly stress factors is almost identical though less marked (Table 8.3). Flow rates are poorly simulated by the *linear* model. All

the other models get a score of *good*. Hydraulic heads are poorly simulated by the *3 layers* model and very poorly simulated by the *linear* model. The other models obtain a score of *satisfactory*. These scores indicate that models with monthly stress factors are closer from each others in term of model performance than models with daily stress factors. However, considering both flow rates and hydraulic heads, the *linear* and the *3 layers* models still fail to satisfy the performance levels defined.

Table 8.3: Ranking of models with monthly stress factors based on performance intervals.

	Flow rates	Hydraulic heads
250 m	★★★★	★★★
500 m	★★★★	★★★
750 m	★★★★	★★★
1000 m	★★★★	★★★
6 layers	★★★★	★★★
3 layers	★★★★	★★
linear	★★	★

Considering simultaneously the number of stars obtained for flow rates and hydraulic heads, model performance is *satisfactory* for the *250 m*, *500 m*, *750 m*, and *1000 m* models. Therefore, contrary to results obtained for models with daily stress factors, model results are not significantly improved when increasing the number of elements per square kilometer from 2 to 40. However, a fine discretisation of the unsaturated zone and representative saturation–pressure relations are still required for expecting *satisfactory* model performance.

As expected, optimisation increases model scores (Table 8.4). All the optimised models gain one star in terms of flow rates. The scores obtained range from *satisfactory* for the *linear* model to *very good* for the other models. This indicates that every optimised model satisfies the performance levels defined

in terms of flow rates. Apart for the *3 layers* model, the improvement of hydraulic head simulation is not sufficient for the other models to gain one star. Furthermore, the *6 layers* model loses one rank. This is due to the fact that the improvement of hydraulic head magnitudes is sometimes obtained to the detriment of the improvement of hydraulic head variations. The scores obtained ranges from *very poor* for the *linear* model and *poor* for the *6 layers* model to *satisfactory* for the other models. This highlights the limitations of optimisation which is capable of compensating for technical and structure uncertainties only to some extent. This is particularly the case for hydraulic heads.

Table 8.4: Ratings of optimised models with monthly stress factors based on performance intervals.

	Flow rates	Hydraulic heads
250 m	★★★★★	★★★
500 m	★★★★★	★★★
750 m	★★★★★	★★★
1000 m	★★★★★	★★★
6 layers	★★★★★	★★
3 layers	★★★★★	★★★
linear	★★★	★

8.3 Conclusions

According to linear posterior model probabilities, only the *3 layers* and the *linear* models are significantly poorer than the other models. The same ranking is obtained for models with daily stress factors and for models with monthly stress factors. When the optimised models are included, they take the first places. However, with the sole use of the objective function, specific model weaknesses highlighted by performance criteria are easily overlooked.

Therefore, a ranking system based on performance criteria is proposed as a complement to this ranking based on information theory.

The ranking system based on performance criteria highlights that horizontal spatial discretisation is less important than vertical spatial discretisation and saturation–pressure relations in the unsaturated zone for obtaining satisfactory model results. Considering simultaneously the number of stars obtained for flow rates and hydraulic heads, satisfactory model results are expected until an element size of 1000 m which is equivalent to 2 elements per square kilometer. The model results are more sensitive to vertical spatial discretisation and representation of the unsaturated zone since poor model results are obtained when the unsaturated zone is coarsely discretised and when (over)simplified saturation–pressure relations are used in the unsaturated zone. The optimisation can partially compensate for such model technical and structure uncertainties. However, the improvement is limited, especially for hydraulic heads. All these findings are common to models with daily and average monthly stress factors.

All together, for gaining time while ensuring satisfactory model results, these findings suggest that coarsening horizontal spatial discretisation is preferable than coarsening vertical spatial discretisation or (over)simplifying saturation–pressure relations in the unsaturated zone. Given the complexity of the synthetic case, these findings are likely to be powerful in practice for similar type of large-scale physically-based and spatially-distributed groundwater flow model (i.e. several hundreds of square kilometers).

Chapter 9

General conclusions and perspectives

9.1 Key findings on the influence of model technical and structure uncertainties

According to the qualitative and quantitative analyses performed, models with daily stress factors and models with average monthly stress factors are influenced identically by model technical and structure uncertainties. However, the rate of deterioration of model performance when model gets coarser is in general slower for models with average monthly stress factors. Additionally, execution times of models with daily stress factors are far longer. Therefore, provided that the objectives of the study are met at such a time resolution, working with average monthly stress factors is an efficient way of saving time.

The investigations performed indicate that flow rates are more sensitive to horizontal spatial discretisation than hydraulic heads. Common to each model tested, the underestimation of low flow rates is due to a poor represen-

tation of the surface water network. This prevents from simulating properly groundwater–surface water interactions which constitute the key component of the hydrograph during dry seasons. Proper to the coarsest models tested, the overestimation of high flow rates is rather related to the smoothing of surface slopes. This facilitates runoff during wet periods. Furthermore, surface and subsurface heterogeneities are poorly represented when horizontal spatial discretisation gets coarser. This also influences model results.

Conversely, hydraulic heads are more sensitive to vertical spatial discretisation than flow rates. The coarsest model tested poorly simulates both hydraulic head magnitudes and hydraulic head variations. This is due to the coarse discretisation of the unsaturated zone which prevents from simulating properly infiltration and groundwater–surface water interactions.

Considering the poor simulation of both flow rates and hydraulic heads obtained with linear saturation–pressure relations instead of van Genuchten functions in the unsaturated zone, it is clear that (over)simplifying saturation–pressure relations in the unsaturated zone can significantly deteriorate model results. Furthermore, this increases the execution times.

Optimisation can partially compensate for errors induced by model technical and model structure uncertainties. However, the improvement of model fit is limited and potentially obtained at the cost of unplausible parameter values, especially for the coarsest models. This highlights the necessity of systematically checking optimised parameter values against field data. Furthermore, in spite of quite similar ranges of stress factors, model fit is in general slightly poorer outside the calibration period. This highlights the necessity of calibrating model with ad hoc stress factors with respect to the objectives of the study. However, compared to non-optimised models, performance of optimised models outside the calibration period is greater. This shows the

importance of optimisation in spite of its limitations.

Whatever the spatial discretisation, composite scaled parameter sensitivities are almost identical. The most sensitive parameters are always related to flow in the unsaturated zone and evapotranspiration processes. This is not the case of models with linear saturation–pressure relations. These models show no predominant parameter in terms of sensitivity.

All together, considering simultaneously flow rates and hydraulic heads, model performance is less sensitive to horizontal spatial discretisation than to vertical spatial discretisation and saturation–pressure relations in the unsaturated zone. Therefore, for gaining time while ensuring satisfactory model results, coarsening horizontal spatial discretisation is preferable to coarsening vertical spatial discretisation or (over)simplifying saturation–pressure relations in the unsaturated zone. Given the model ranking obtained, satisfactory model results are expected until an element size of 1000 m.

9.2 Guidelines for large-scale groundwater flow model development

A series of guidelines is proposed in order to further orient large-scale groundwater flow model development towards end-users' expectations. Given the investigation of performance criteria carried out, it is clear that some of them are better suited than others for evaluating model performance with respect to the specific objectives of a study. Typically, end-users are interested in one or some of these model outputs:

- average magnitude, maximum value, or general evolution of baseflow rates,

- average magnitude, seasonal variations, or general evolution of hydraulic heads,
- hydraulic head maps.

Given these typical end-users' expectations, it is possible to specify which performance criteria to privilege for evaluating model performance with respect to the specific objective(s) of the study.

If the study focuses on the general evolution of baseflow rates without any particular interest for a specific feature of the hydrograph, the Nash-Sutcliffe efficiency criterion NSE_q is to be privileged for evaluating model performance. This criterion takes implicitly into account the capacity of the model to reproduce the mean flow, the spread of flows, and the timing and shape of the hydrograph. Therefore, it does not focus on a specific feature of the hydrograph. Conversely, given its formulation, the peak error criterion PE_q is to be privileged for studies focusing on the maximum values of baseflow rates. When the objective of the study is only to estimate the average magnitude of baseflow rates, the mass balance error criterion MBE_q is sufficient. The same kind of observations are made for hydraulic heads. If the study focuses on the general evolution of hydraulic heads, the root mean square error criterion RMS_h is to be privileged for evaluating model performance since it provides a quantification of overall model fit by measuring the discrepancies between observed values and their simulated equivalent. This criterion is also suited for studies whose objective is to produce hydraulic head maps. When the study focuses on hydraulic head variations, the hydraulic head variations error criterion $HHVE_h$ is to be privileged. If the objective of the study is only to estimate the average magnitude of hydraulic heads, the bias criterion MBE_h is sufficient. These observations are summarised in Table

9.1.

Table 9.1: List of performance criteria to privilege with respect to typical end-users' expectations from large-scale groundwater flow models.

Objective of the study	Performance criteria to privilege
General evolution of baseflow rates	$NSE_q = 1 - \frac{\sum_{t=1}^{nt}\left(q_t^{sim} - q_t^{obs}\right)^2}{\sum_{t=1}^{nt}\left(q_t^{obs} - \mu^{obs}\right)^2}$
Average magnitude of baseflow rates	$MBE_q = \frac{\sum_{t=1}^{nt}\left(q_t^{sim} - q_t^{obs}\right)}{\sum_{t=1}^{nt} q_t^{obs}} \times 100$
Maximum value of baseflow rates	$PE_q = \left(\frac{h_{peak}^{sim}}{h_{peak}^{obs}} - 1\right) \times 100$
General evolution of hydraulic heads	$RMS_h = \sqrt{\frac{1}{nt} \times \sum_{t=1}^{nt}\left(h_t^{sim} - h_t^{obs}\right)^2}$
Average magnitude of hydraulic heads	$MBE_h = \frac{\sum_{t=1}^{nt}\left(h_t^{sim} - h_t^{obs}\right)}{\sum_{t=1}^{nt} h_t^{obs}} \times 100$
Hydraulic head variations	$HHVE_h = \left(\frac{h_{max}^{sim} - h_{min}^{sim}}{h_{max}^{obs} - h_{min}^{obs}} - 1\right) \times 100$
Hydraulic head maps	$RMS_h = \sqrt{\frac{1}{nt} \times \sum_{t=1}^{nt}\left(h_t^{sim} - h_t^{obs}\right)^2}$

Considering the striking capacity of such performance criteria to provide a specific quantification of model performance, they should be calculated systematically depending on the objective(s) of the study and values to reach for each of them should be discussed with end-users prior to model development. The ideal would be to include these performance criteria into the optimisation process via an end-users objective function. Furthermore, for saving time, optimisation could be stopped once performance criteria included in this end-users objective function would have reached the values specified by the end-users with respect to the objectives of the study. The general form of this end-users objective function could be:

$$\Phi_{end-users} = \sum_{i=1}^{nobj} \omega_i \times \Omega_i \tag{9.1}$$

where ω_i would be the weight corresponding to the performance criteria Ω_i and $nobj$ would depend on the number of objective(s) of the study. Therefore, this end-users objective function would include from 1 to $nobj$ performance criteria. This end-users objective function $\Phi_{end-users}$ could constitute

a substitute to the weighted least-squares objective function Φ. However, for problems focusing on NSE_q and/or RMS_h, optimising $\Phi_{end-users}$ instead of Φ should lead to the same results since these performance criteria are closely related to the weighted least-squares objective function through their formulation. Therefore, under certain circumstances, using $\Phi_{end-users}$ instead of Φ would be equivalent.

Given the influence of model technical and structure uncertainties on model results, choices relative to spatial discretisation should also be made in function of the objectives of the study. When focusing on baseflow rates, horizontal spatial discretisation should be refined along the surface water network for simulating properly groundwater–surface water interactions. Conversely, when focusing on hydraulic heads, vertical spatial discretisation should be refined in the unsaturated zone for simulating properly infiltration.

9.3 Perspectives for future works

Future works could consist in implementing the end-users objective function together with a specific optimisation method in an inverse modelling code such as PEST or UCODE_2005. If successful on a set of simple cases, the implementation could be further tested on real cases.

Future works could also consist in performing further tests with the synthetic case. These tests could focus on the influence on model results of simulating subsurface flow instead of integrated surface/subsurface flow. These tests could also focus on the influence of model technical and structure uncertainties on transport simulation.

Appendix A

Complement to Chapter 5

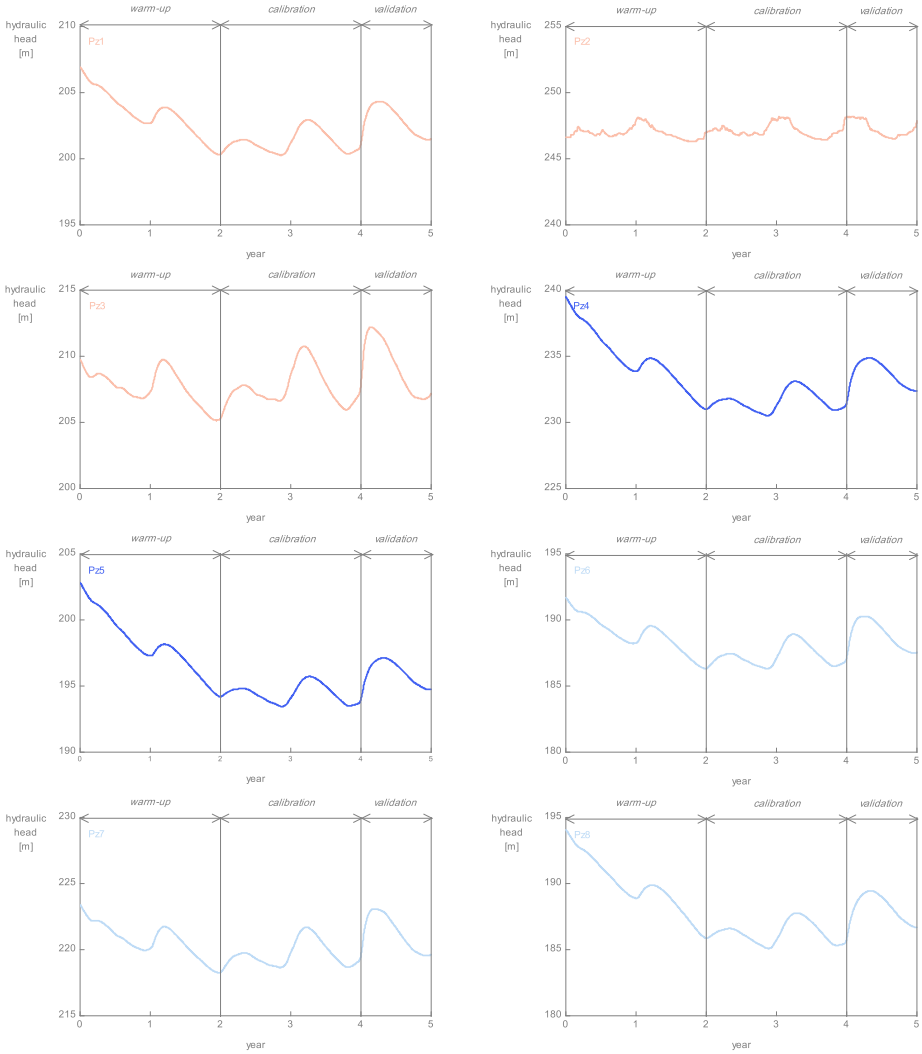

Figure A.1: Transient reference hydraulic head observations for Pz1 to Pz8.

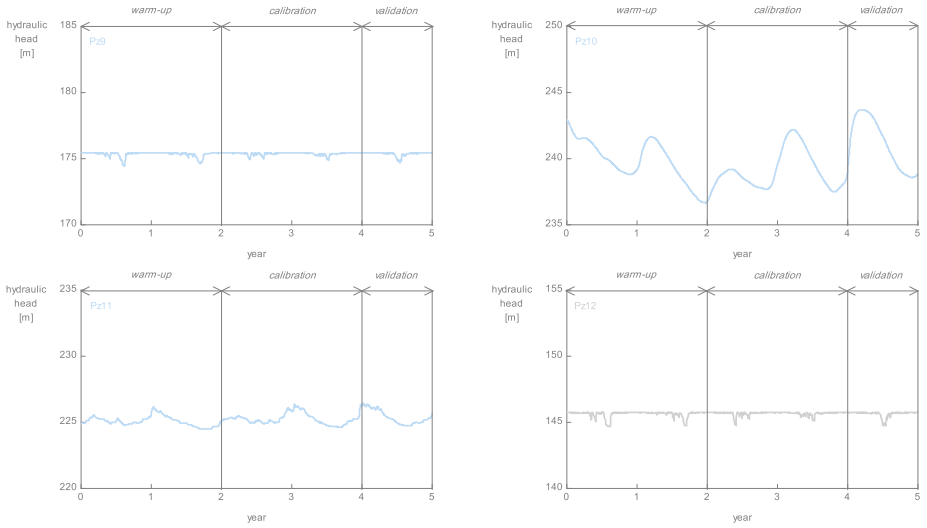

Figure A.2: Transient reference hydraulic head observations for Pz9 to Pz12.

Appendix B

Complement to Chapter 6

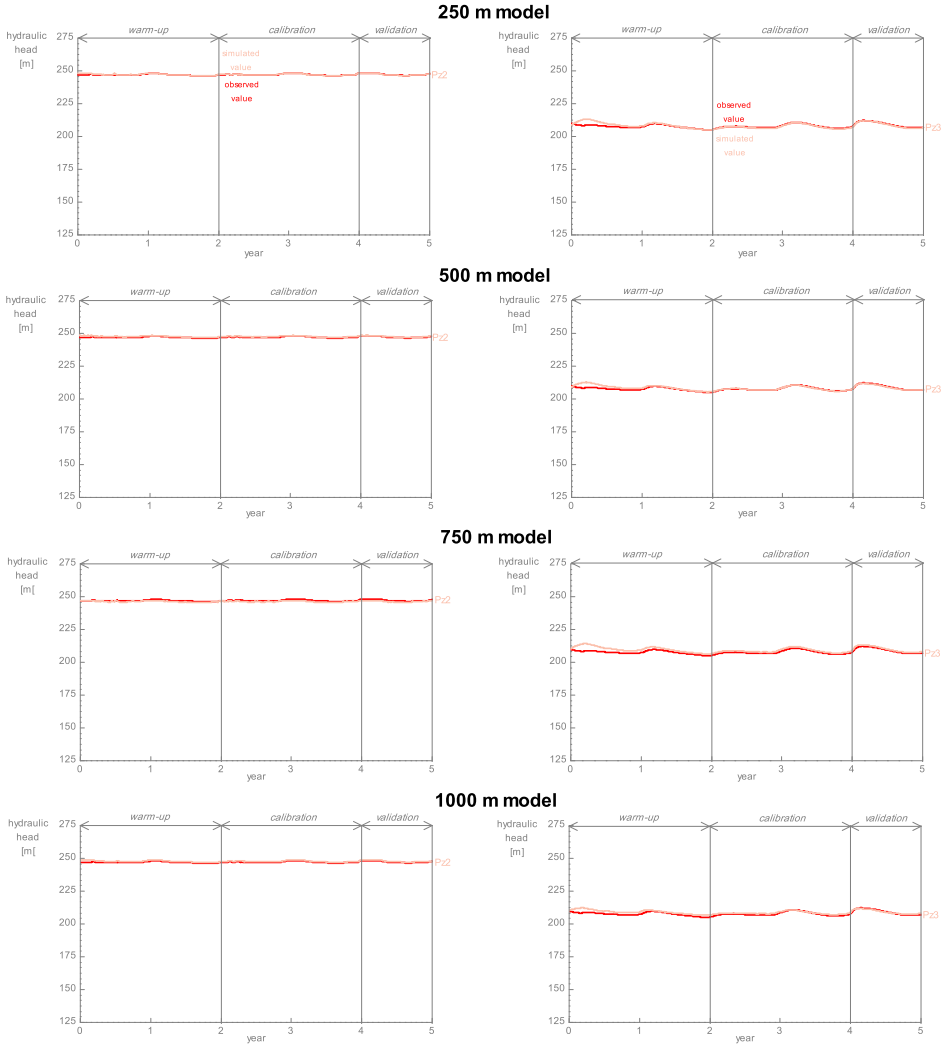

Figure B.1: Comparison of observed and simulated hydraulic heads for Pz2 and Pz3 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors.

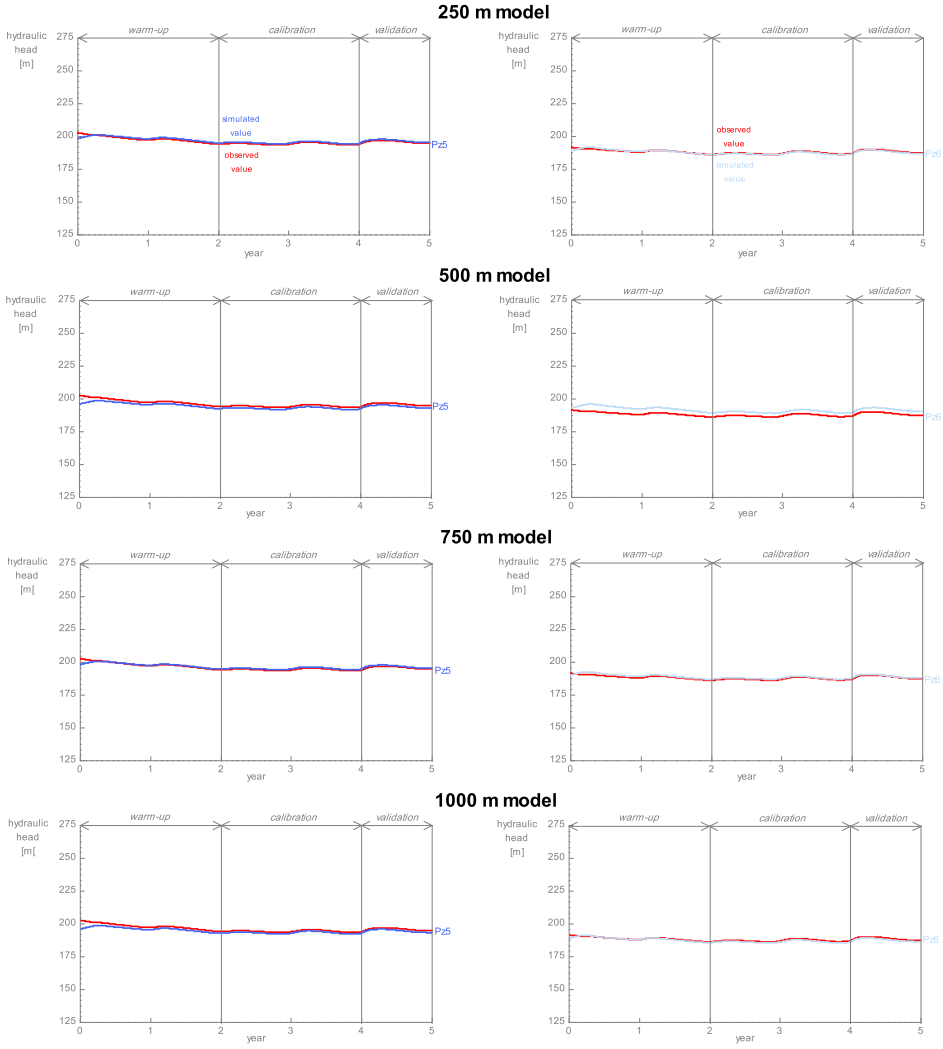

Figure B.2: Comparison of observed and simulated hydraulic heads for Pz5 and Pz6 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors.

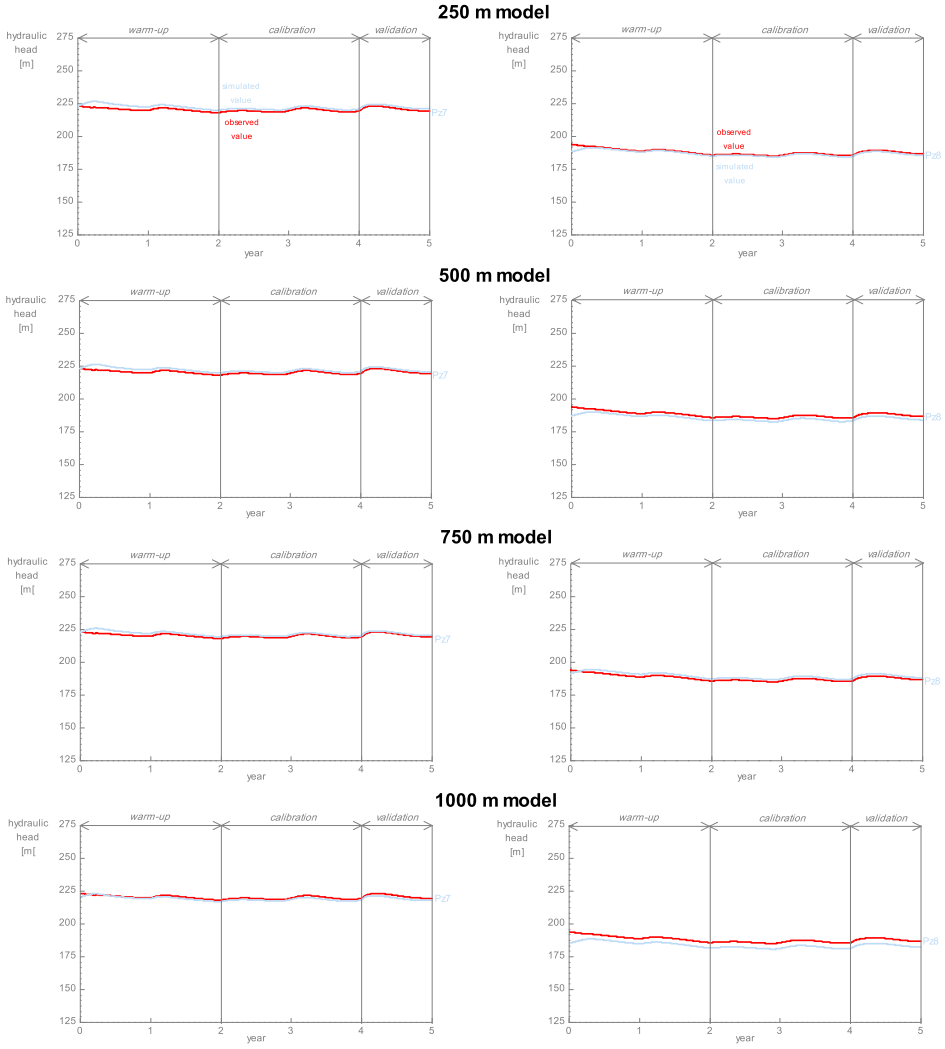

Figure B.3: Comparison of observed and simulated hydraulic heads for Pz7 and Pz8 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors.

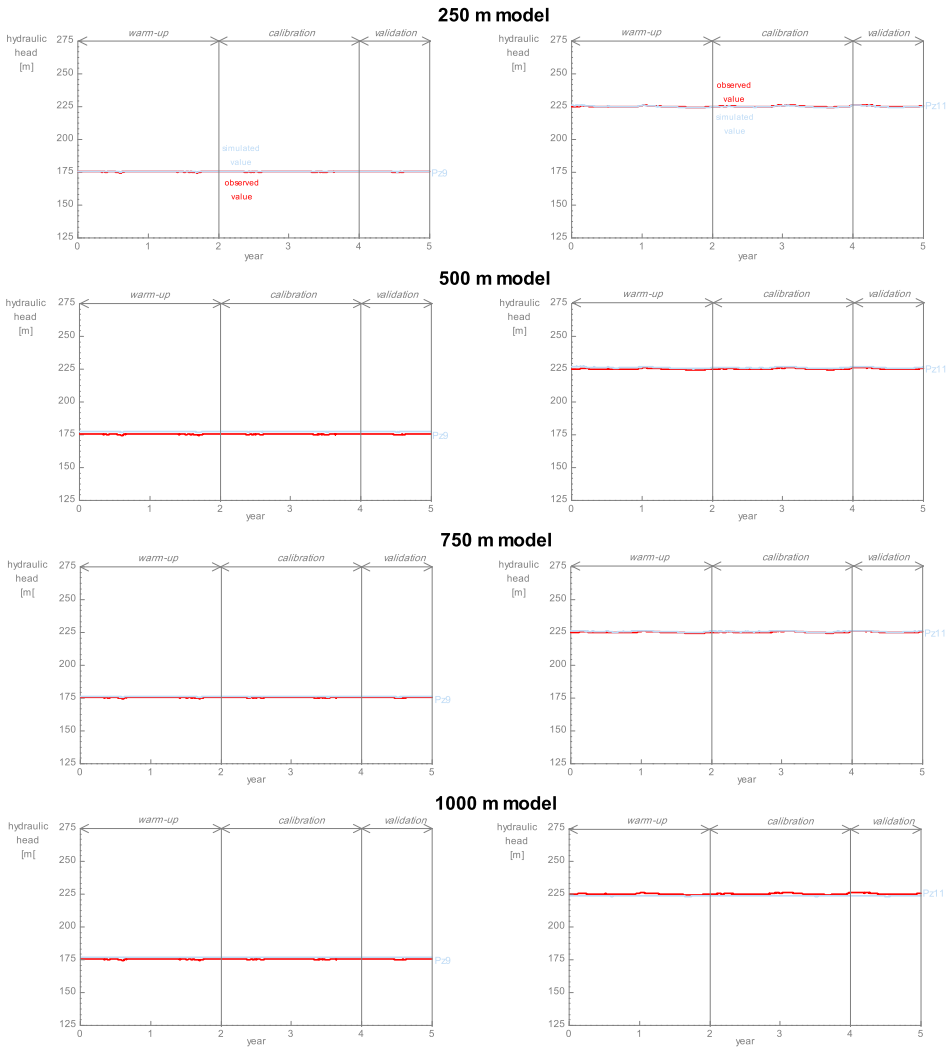

Figure B.4: Comparison of observed and simulated hydraulic heads for Pz9 and Pz11 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors.

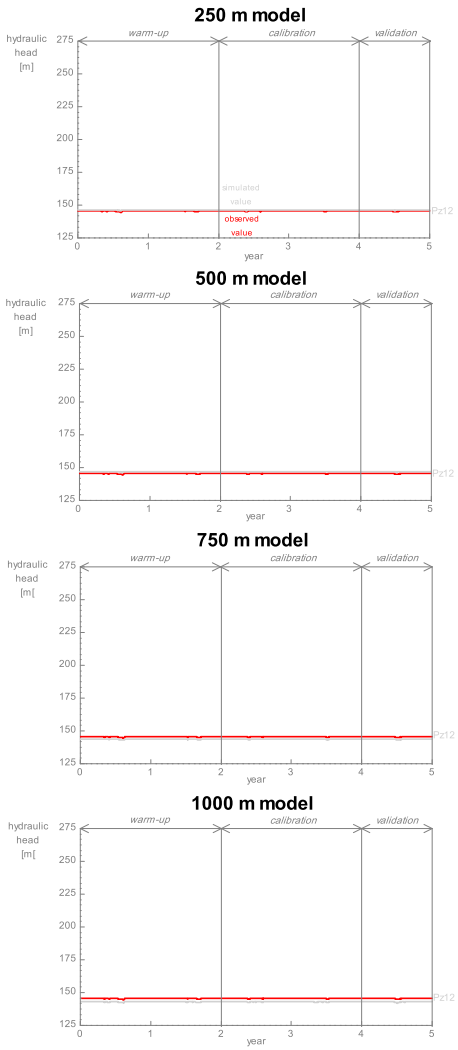

Figure B.5: Comparison of observed and simulated hydraulic heads for Pz12 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors.

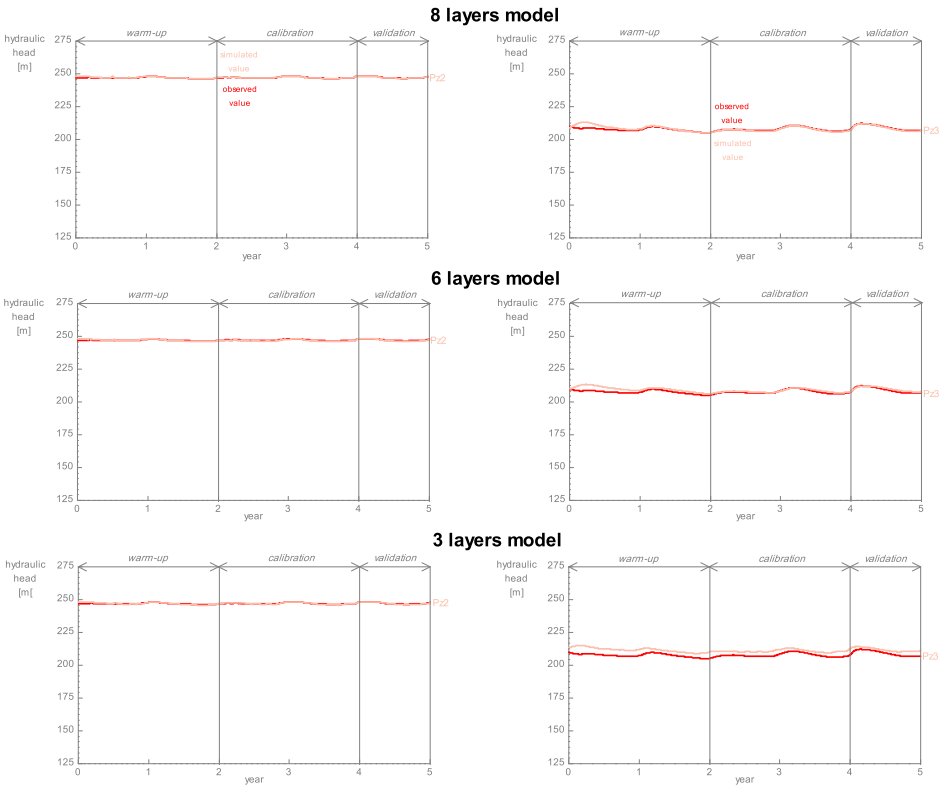

Figure B.6: Comparison of observed and simulated hydraulic heads for Pz2 and Pz3 produced with the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors.

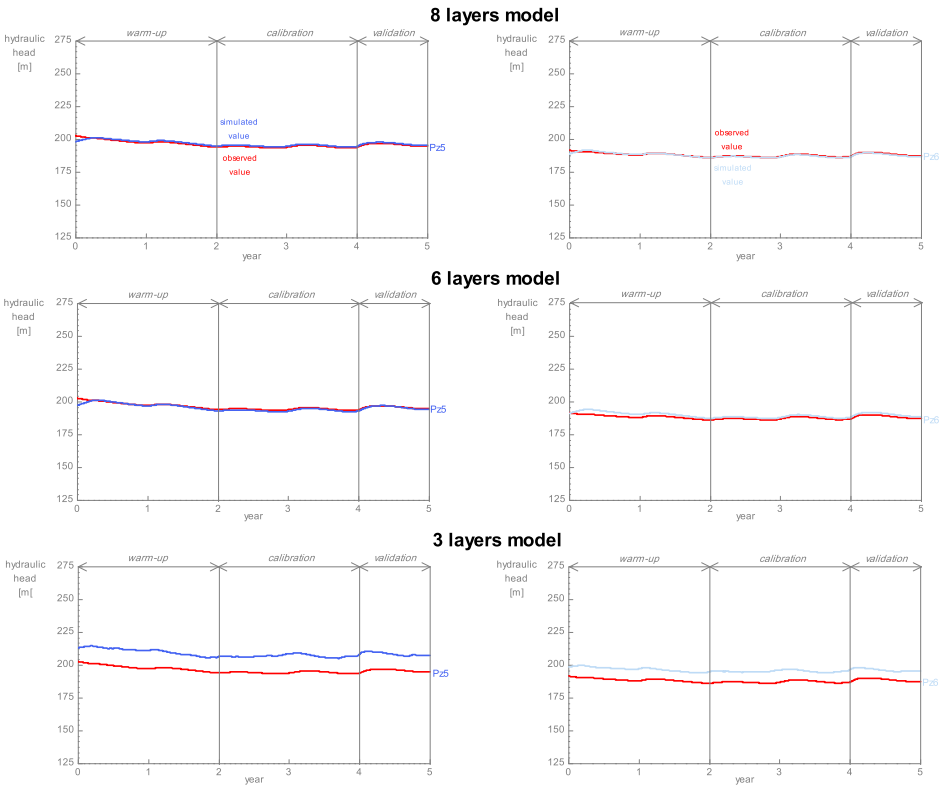

Figure B.7: Comparison of observed and simulated hydraulic heads for Pz5 and Pz6 produced with the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors.

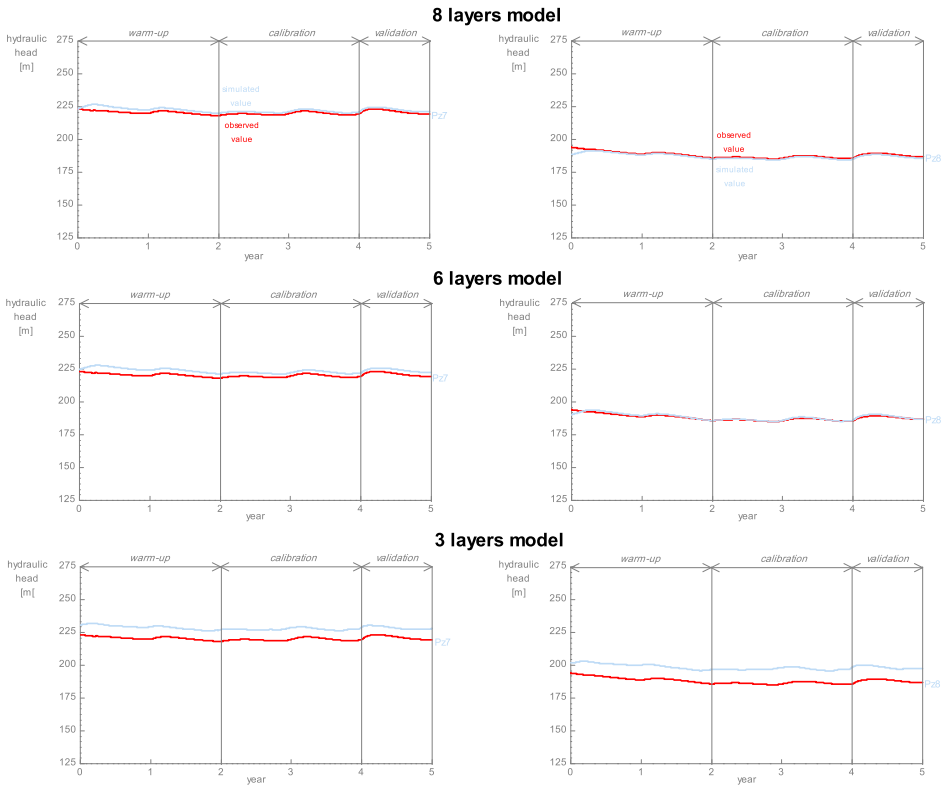

Figure B.8: Comparison of observed and simulated hydraulic heads for Pz7 and Pz8 produced with the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors.

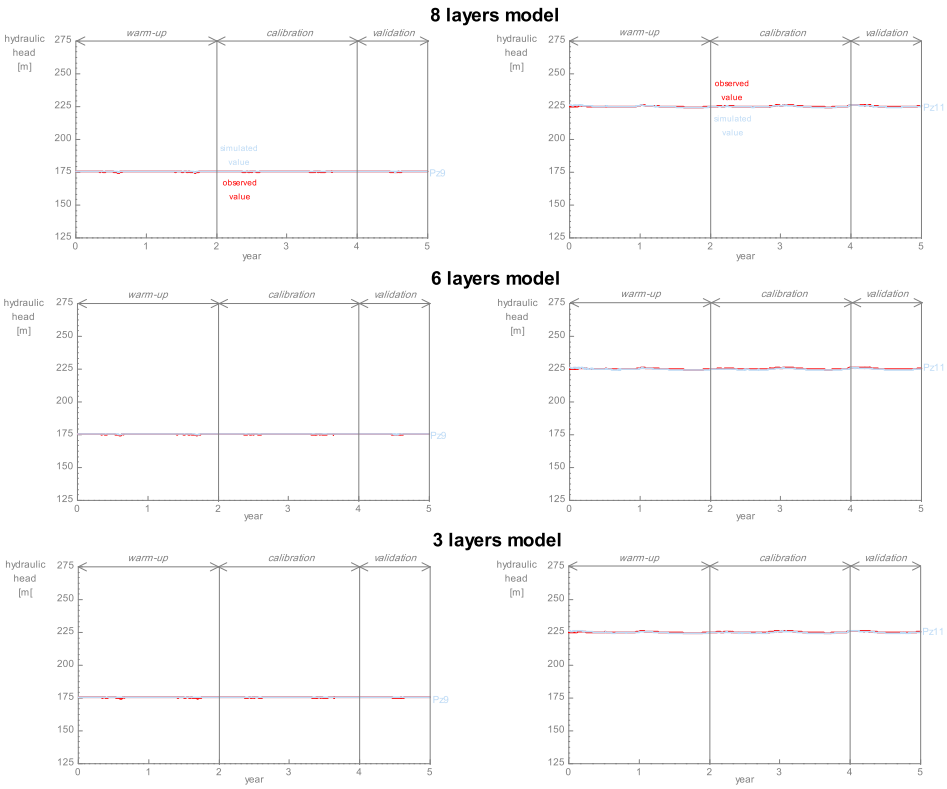

Figure B.9: Comparison of observed and simulated hydraulic heads for Pz9 and Pz11 produced with the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors.

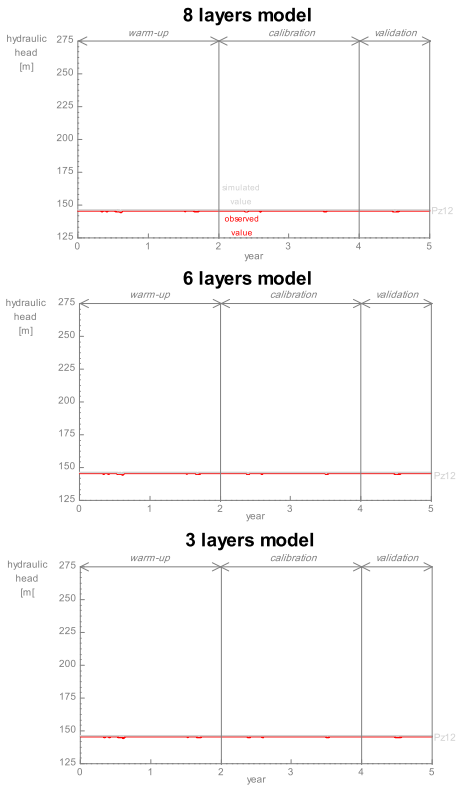

Figure B.10: Comparison of observed and simulated hydraulic heads for Pz12 produced with the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors.

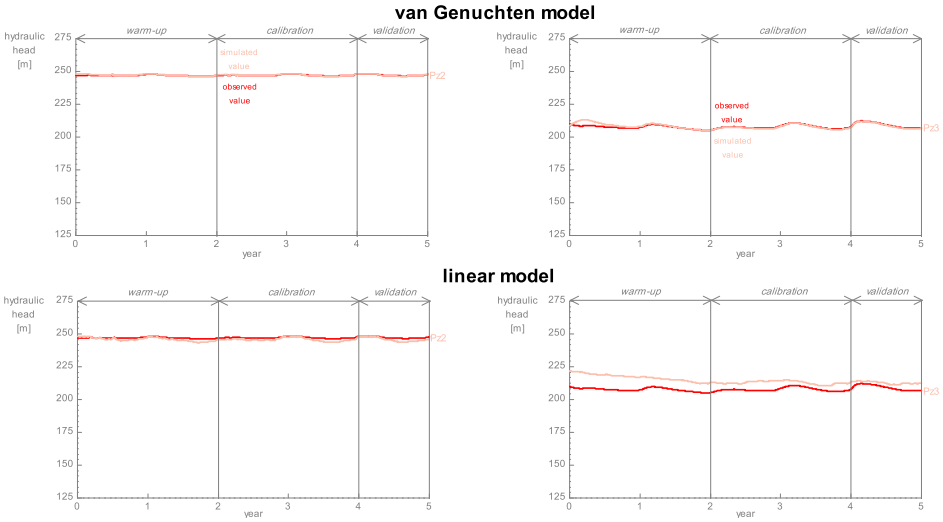

Figure B.11: Comparison of observed and simulated hydraulic heads for Pz2 and Pz3 produced with the *van Genuchten* and *linear* models with daily stress factors.

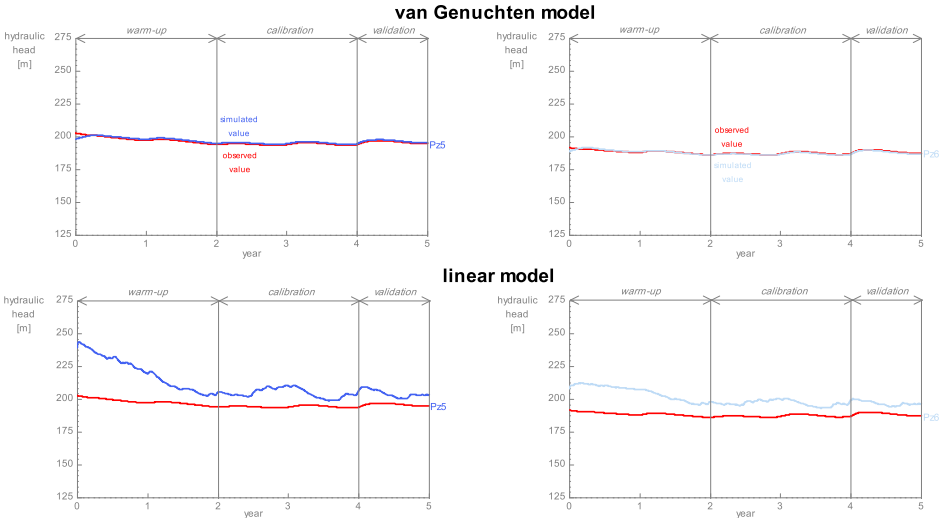

Figure B.12: Comparison of observed and simulated hydraulic heads for Pz5 and Pz6 produced with the *van Genuchten* and *linear* models with daily stress factors.

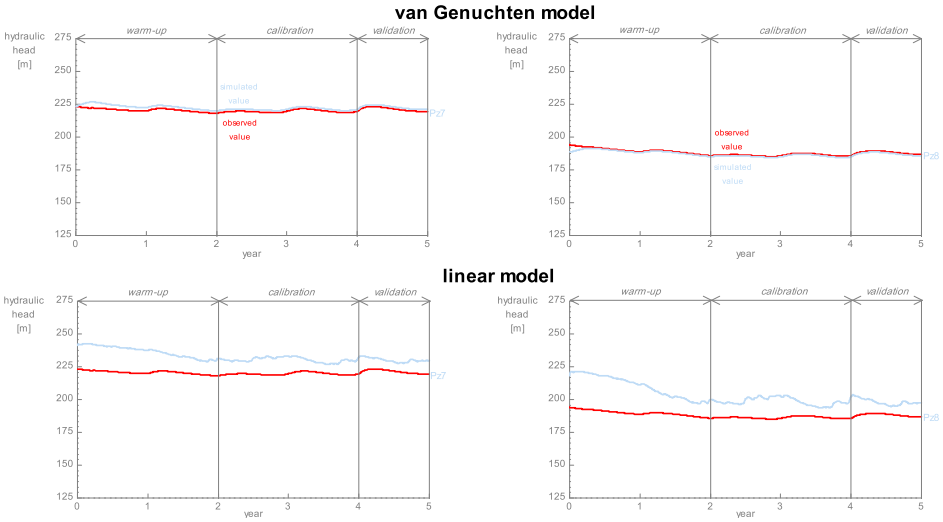

Figure B.13: Comparison of observed and simulated hydraulic heads for Pz7 and Pz8 produced with the *van Genuchten* and *linear* models with daily stress factors.

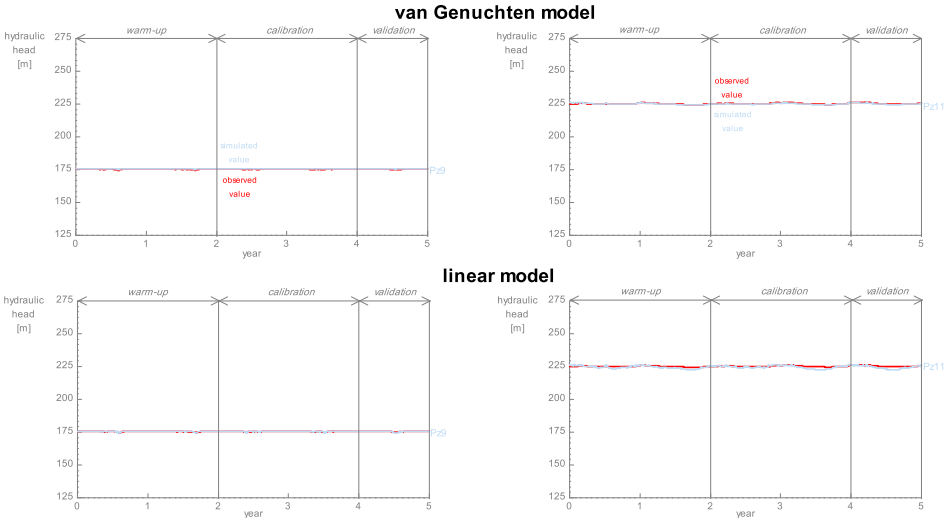

Figure B.14: Comparison of observed and simulated hydraulic heads for Pz9 and Pz11 produced with the *van Genuchten* and *linear* models with daily stress factors.

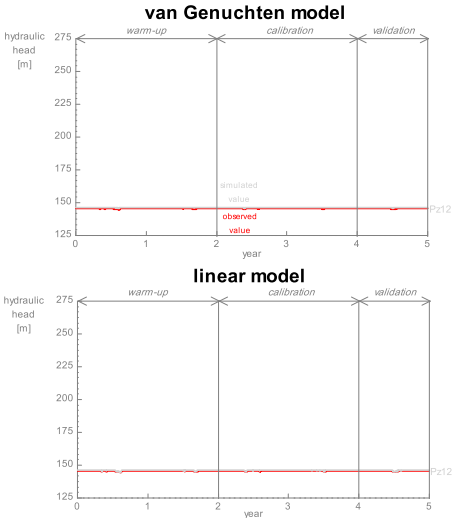

Figure B.15: Comparison of observed and simulated hydraulic heads for Pz12 produced with the *van Genuchten* and *linear* models with daily stress factors.

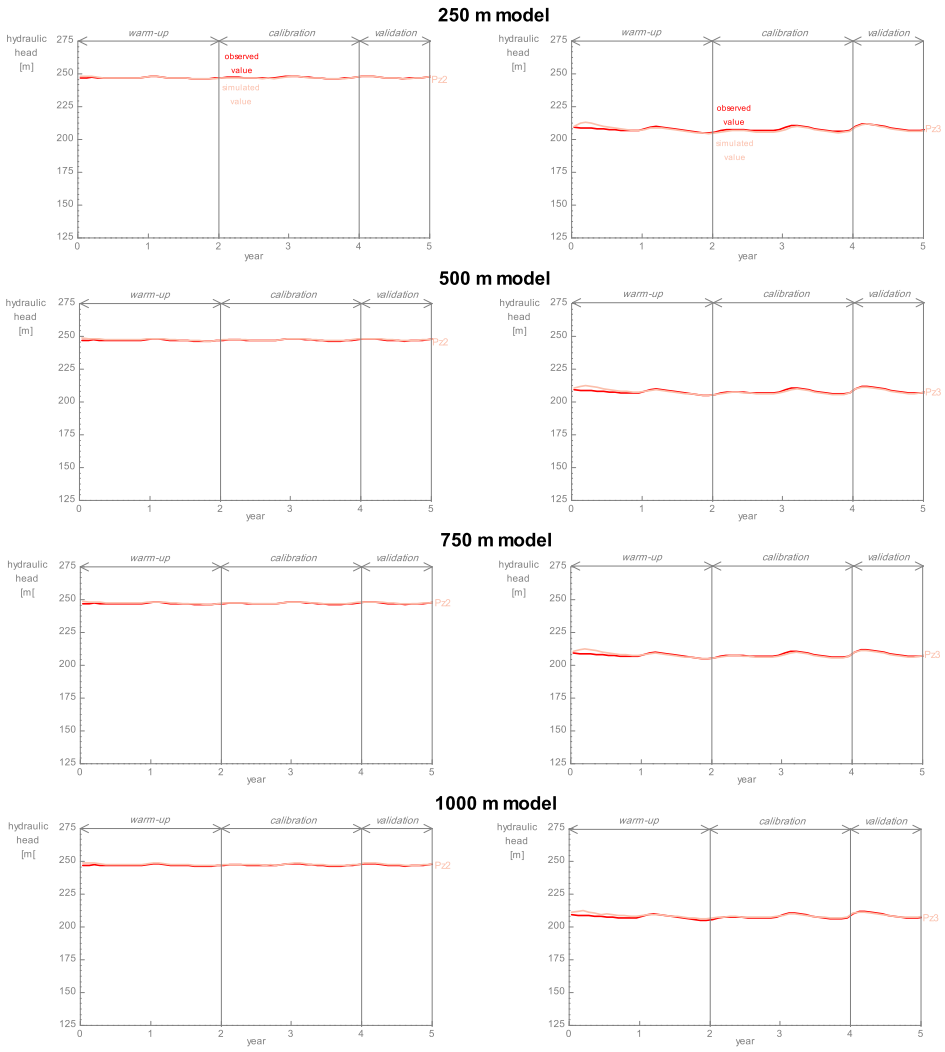

Figure B.16: Comparison of observed and simulated hydraulic heads for Pz2 and Pz3 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

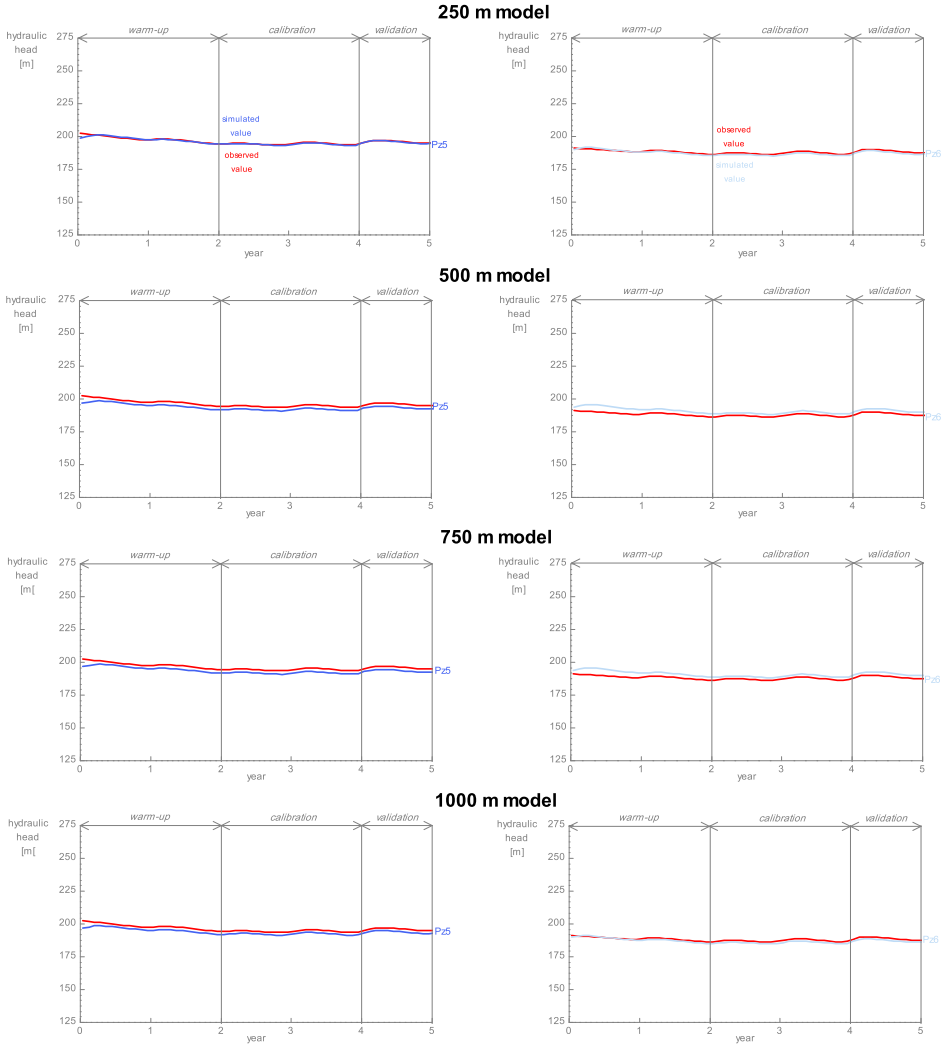

Figure B.17: Comparison of observed and simulated hydraulic heads for Pz5 and Pz6 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

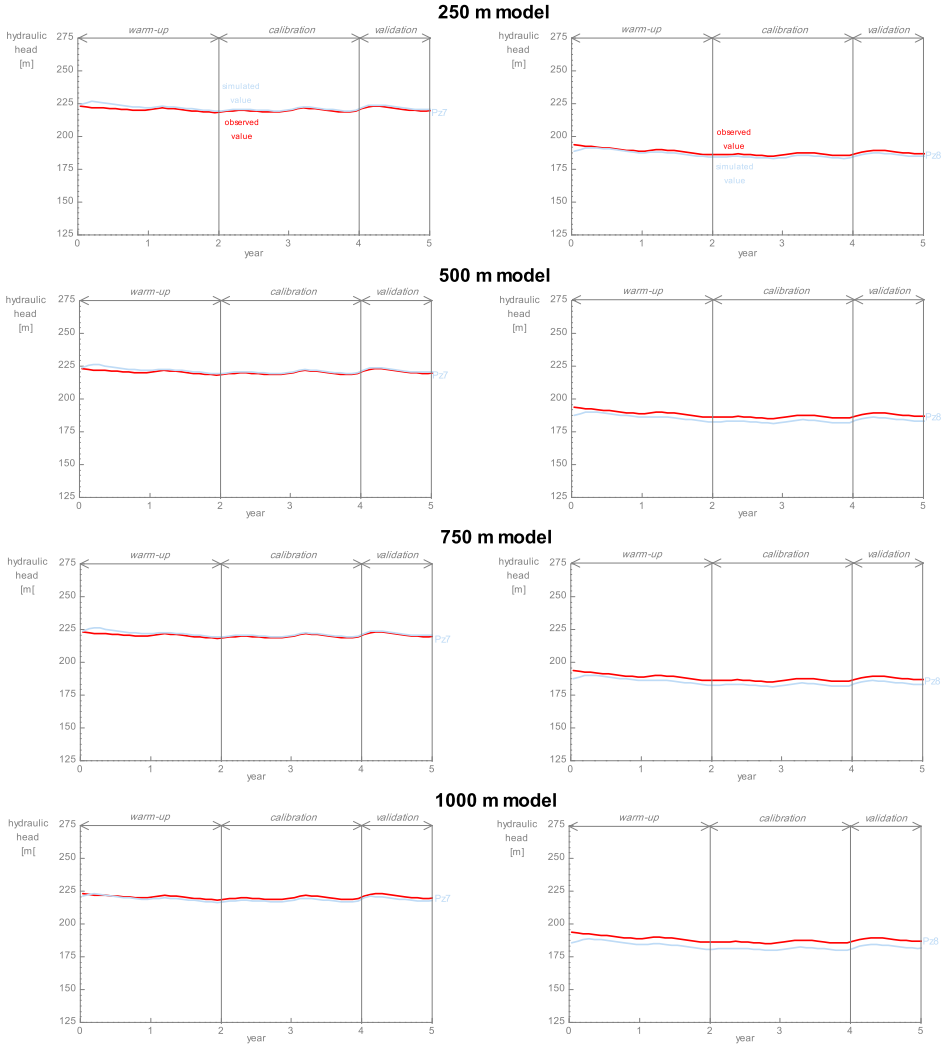

Figure B.18: Comparison of observed and simulated hydraulic heads for Pz7 and Pz8 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

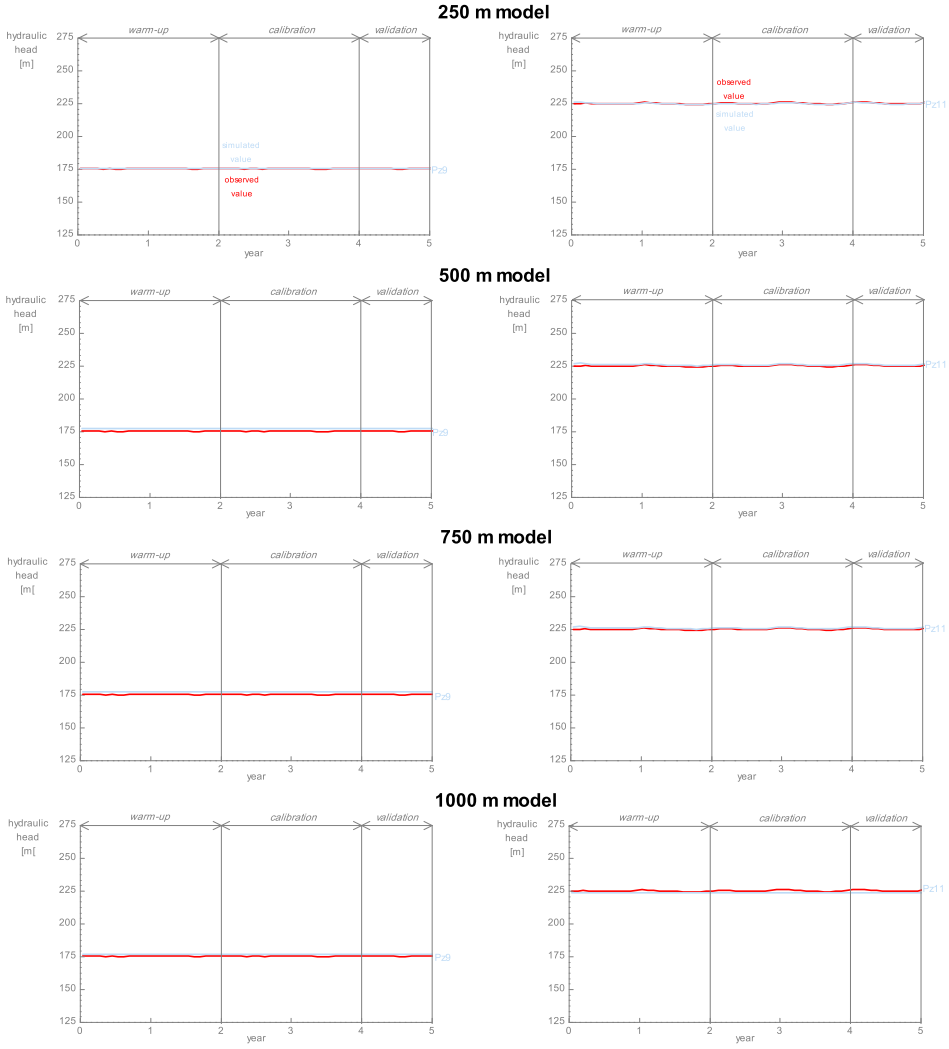

Figure B.19: Comparison of observed and simulated hydraulic heads for Pz9 and Pz11 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

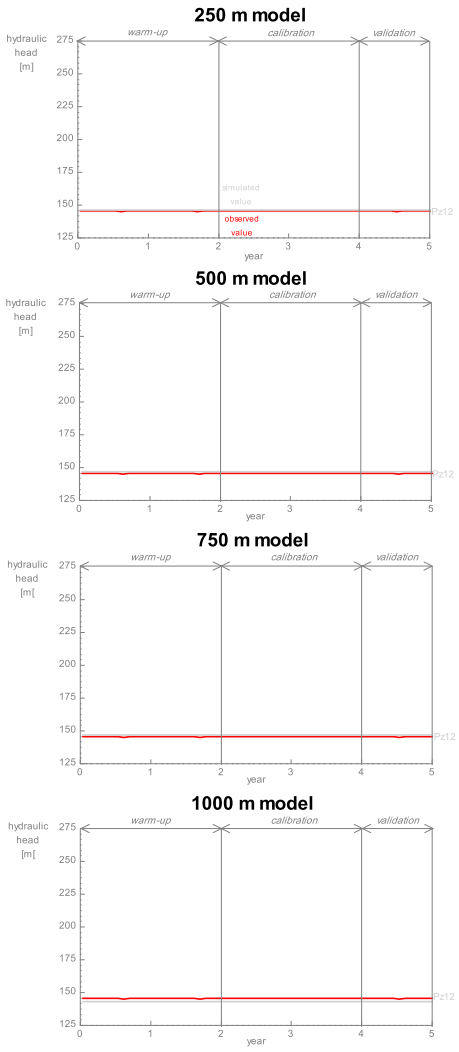

Figure B.20: Comparison of observed and simulated hydraulic heads for Pz12 produced with the *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

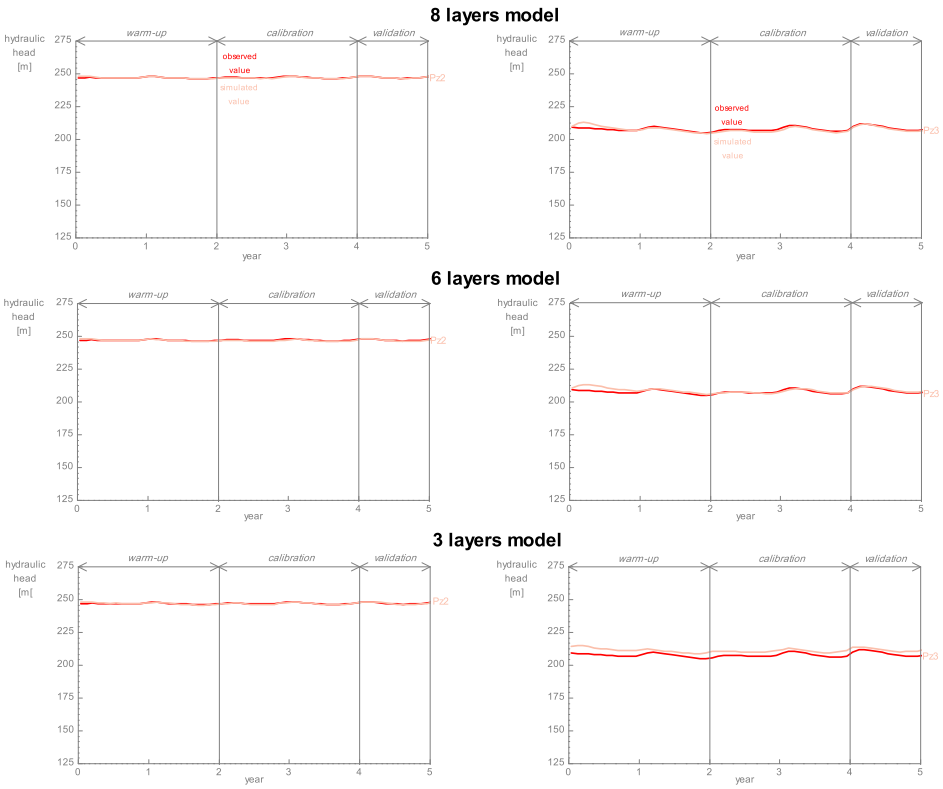

Figure B.21: Comparison of observed and simulated hydraulic heads for Pz2 and Pz3 produced with the *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

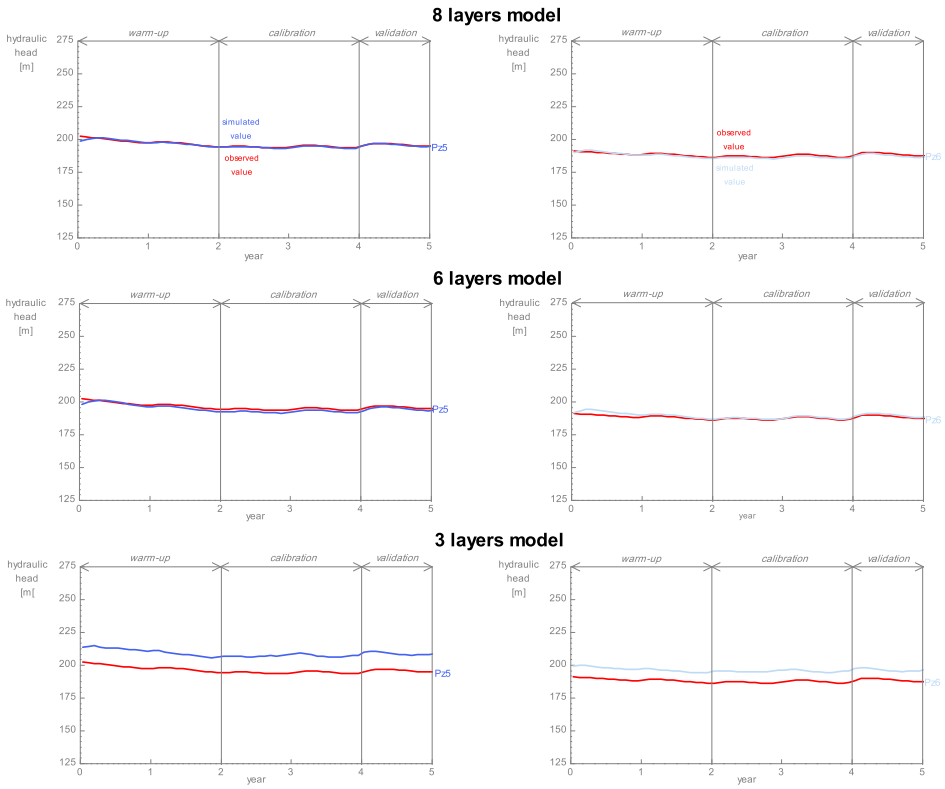

Figure B.22: Comparison of observed and simulated hydraulic heads for Pz5 and Pz6 produced with the *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

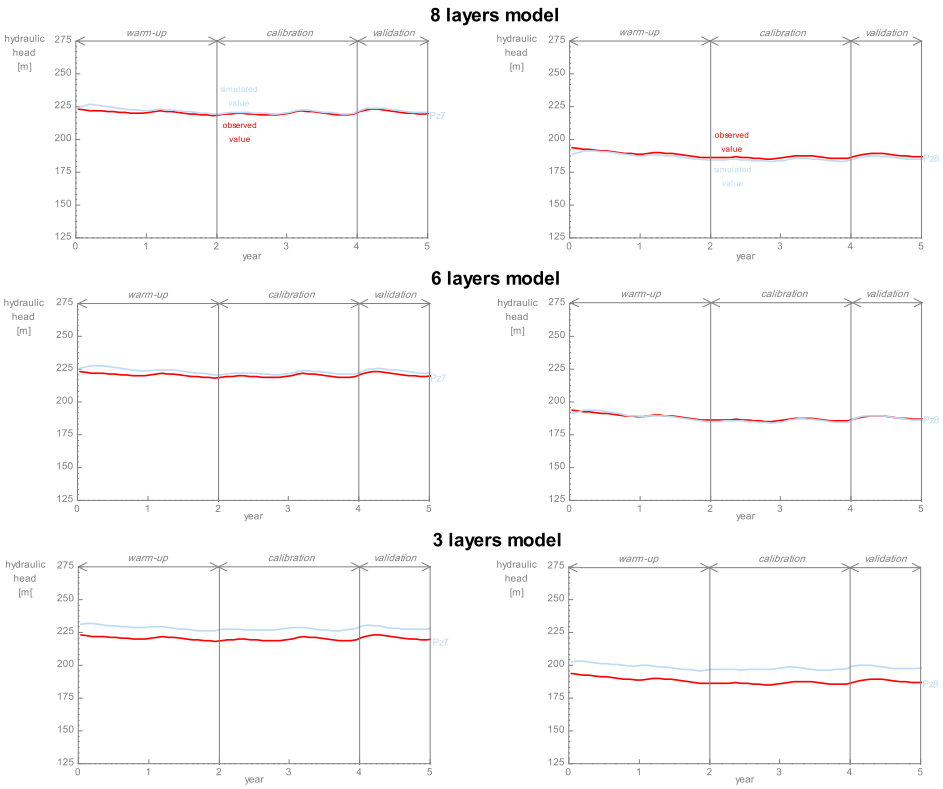

Figure B.23: Comparison of observed and simulated hydraulic heads for Pz7 and Pz8 produced with the *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

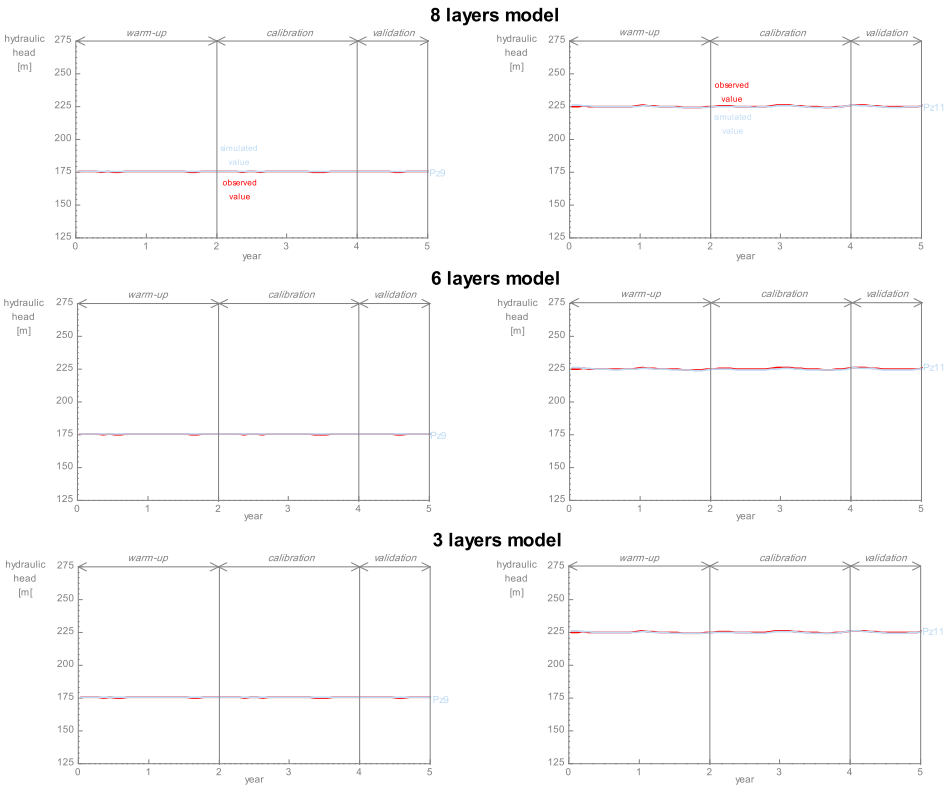

Figure B.24: Comparison of observed and simulated hydraulic heads for Pz9 and Pz11 produced with the *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

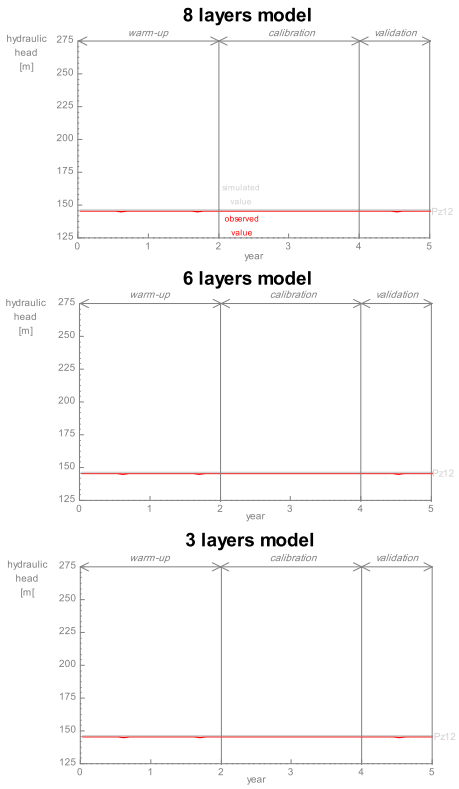

Figure B.25: Comparison of observed and simulated hydraulic heads for Pz12 produced with the *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

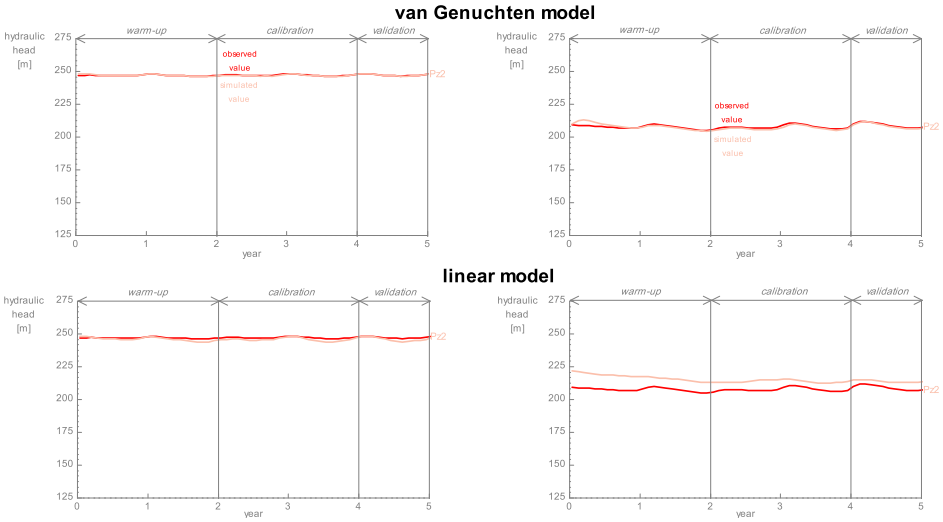

Figure B.26: Comparison of observed and simulated hydraulic heads for Pz2 and Pz3 produced with the *van Genuchten* and *linear* models with monthly stress factors.

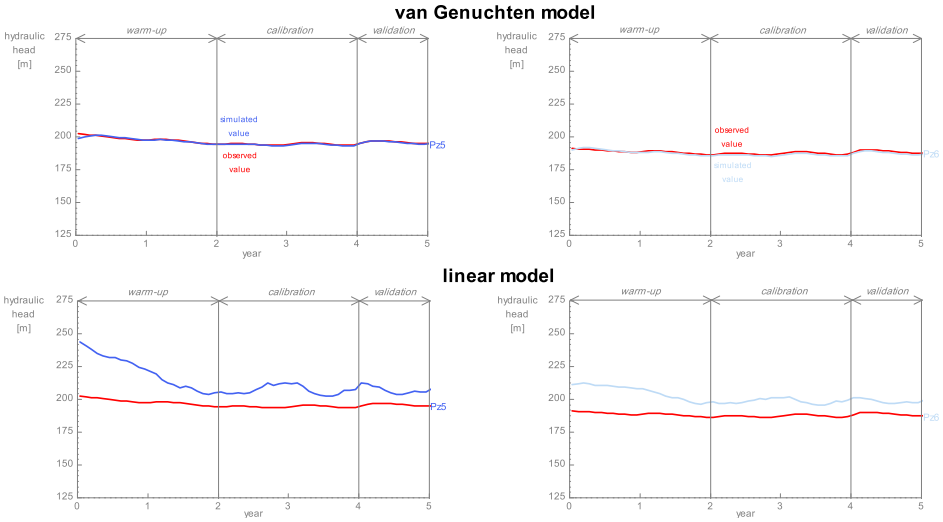

Figure B.27: Comparison of observed and simulated hydraulic heads for Pz5 and Pz6 produced with the *van Genuchten* and *linear* models with monthly stress factors.

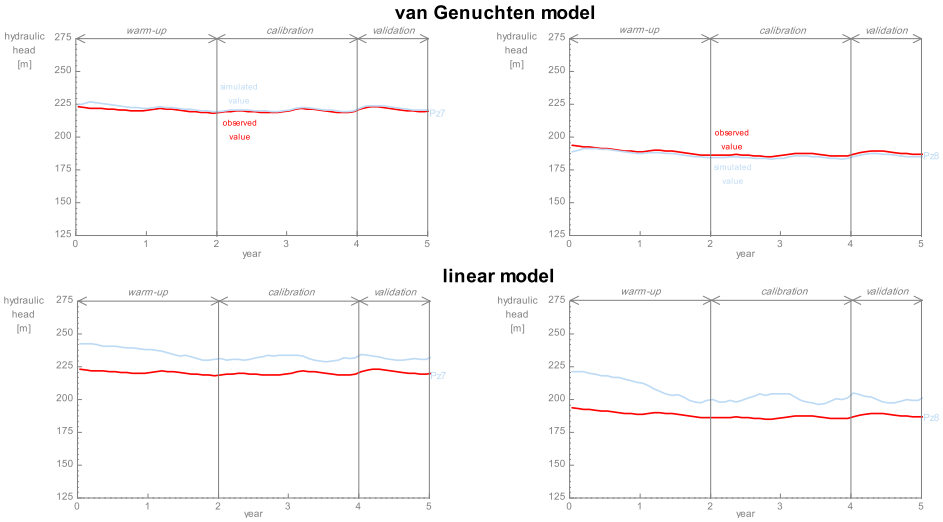

Figure B.28: Comparison of observed and simulated hydraulic heads for Pz7 and Pz8 produced with the *van Genuchten* and *linear* models with monthly stress factors.

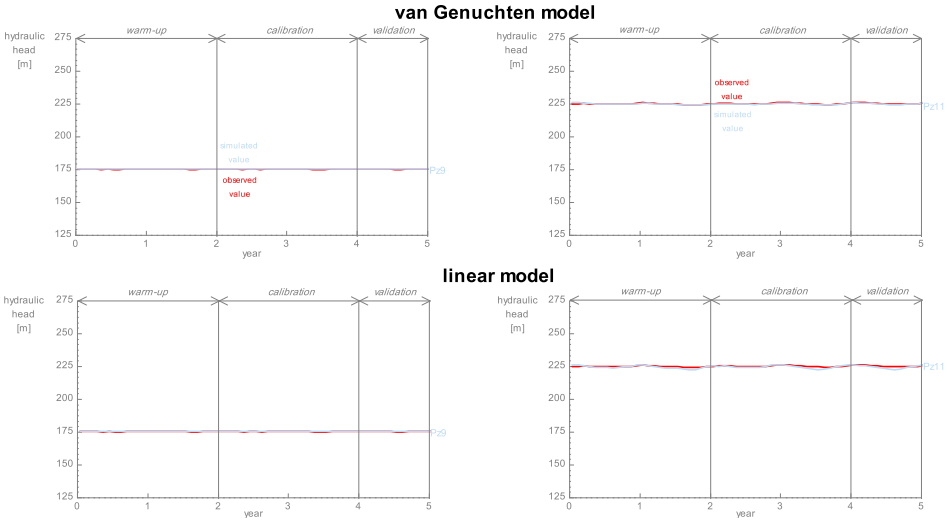

Figure B.29: Comparison of observed and simulated hydraulic heads for Pz9 and Pz11 produced with the *van Genuchten* and *linear* models with monthly stress factors.

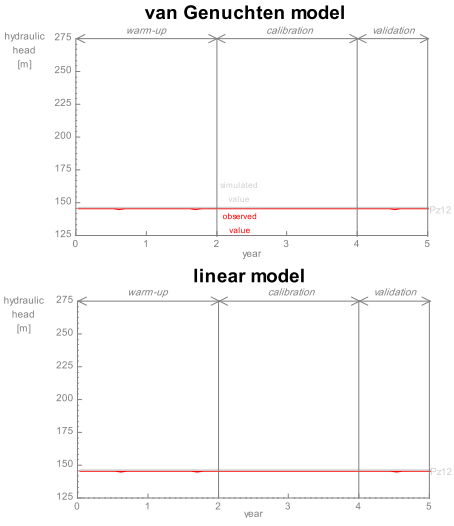

Figure B.30: Comparison of observed and simulated hydraulic heads for Pz12 produced with the *van Genuchten* and *linear* models with monthly stress factors.

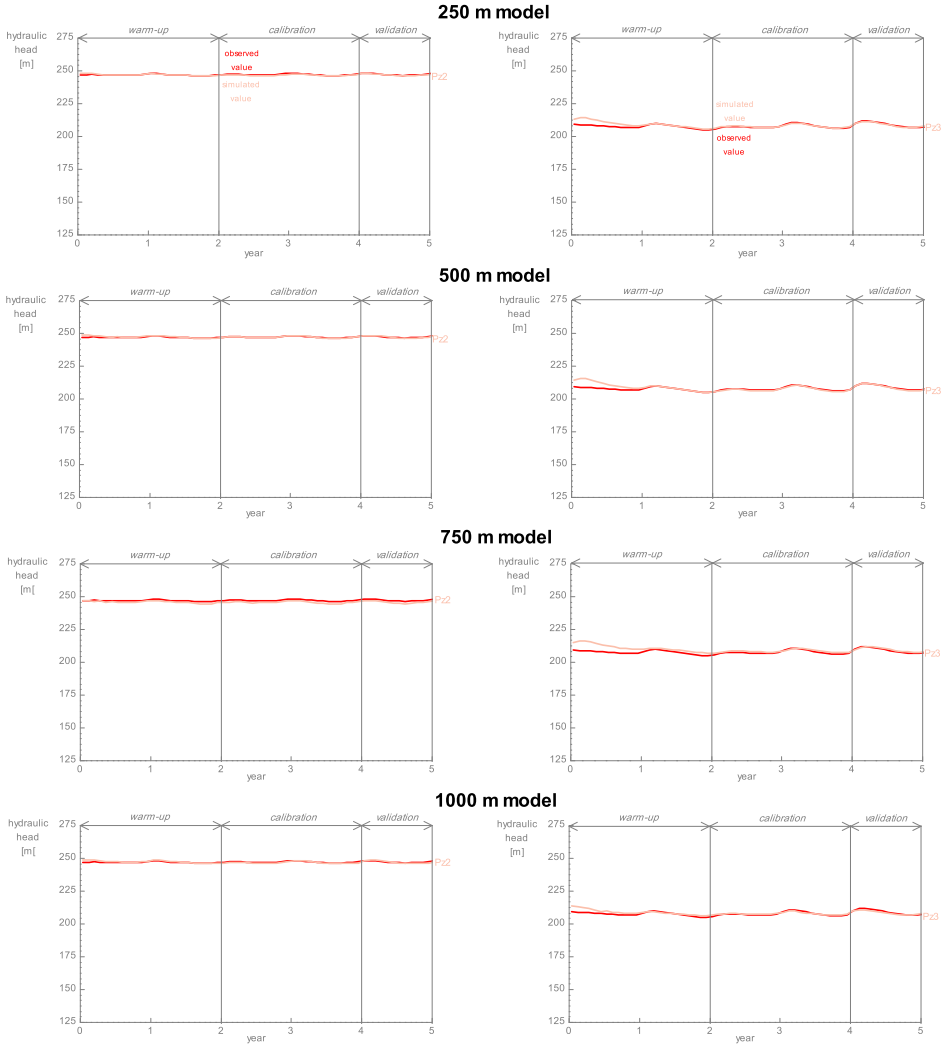

Figure B.31: Comparison of observed and simulated hydraulic heads for Pz2 and Pz3 produced with the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

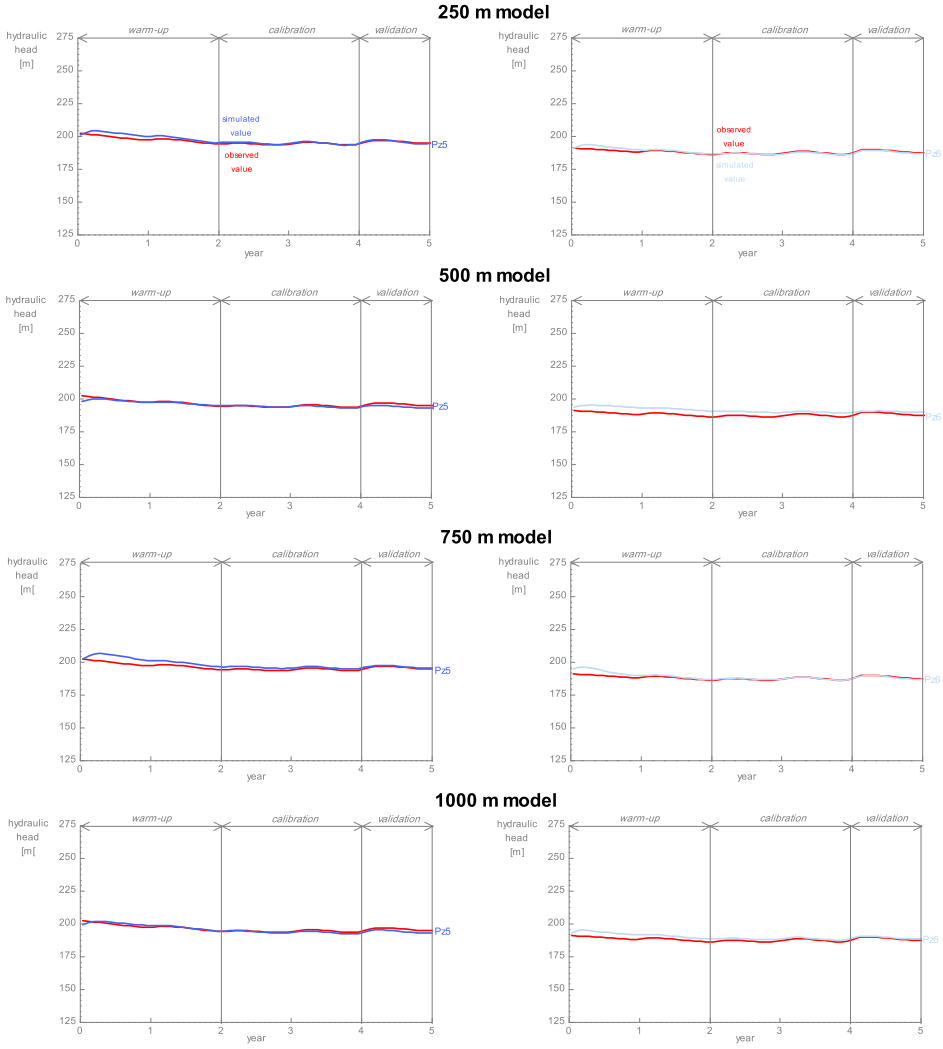

Figure B.32: Comparison of observed and simulated hydraulic heads for Pz5 and Pz6 produced with the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

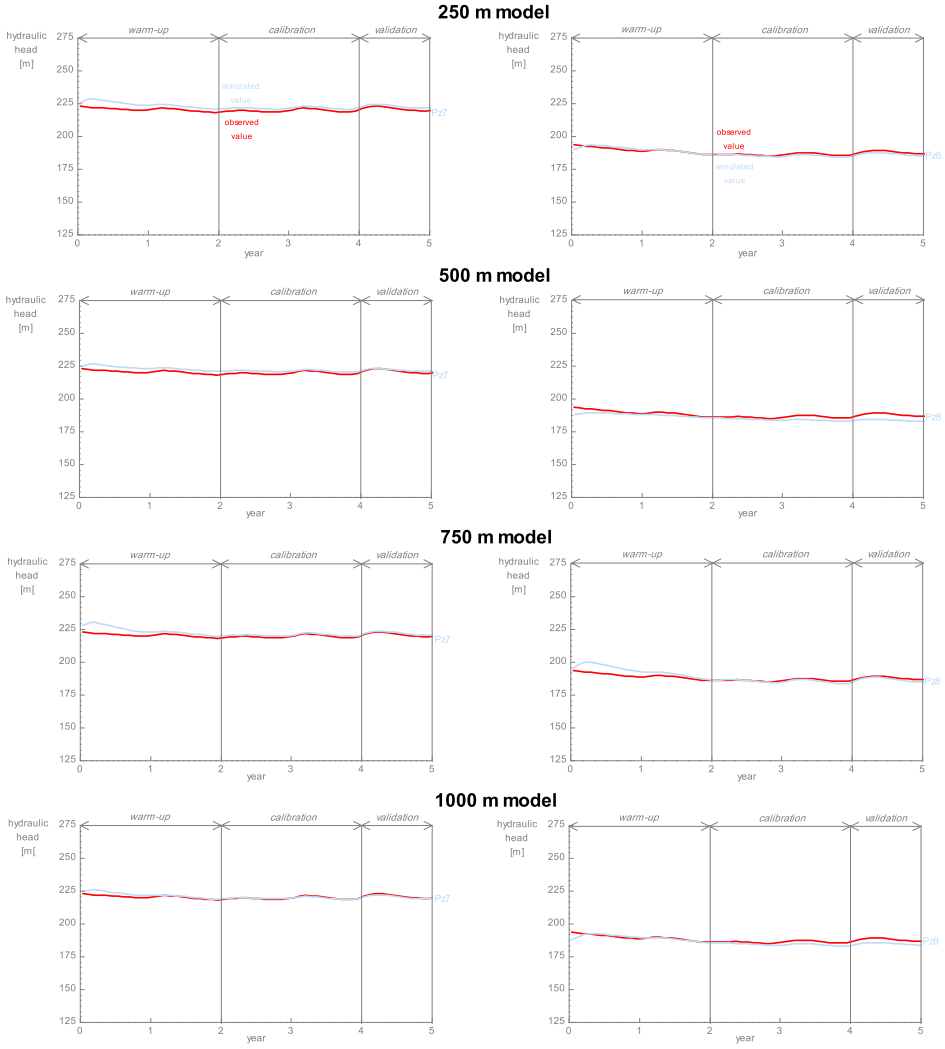

Figure B.33: Comparison of observed and simulated hydraulic heads for Pz7 and Pz8 produced with the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

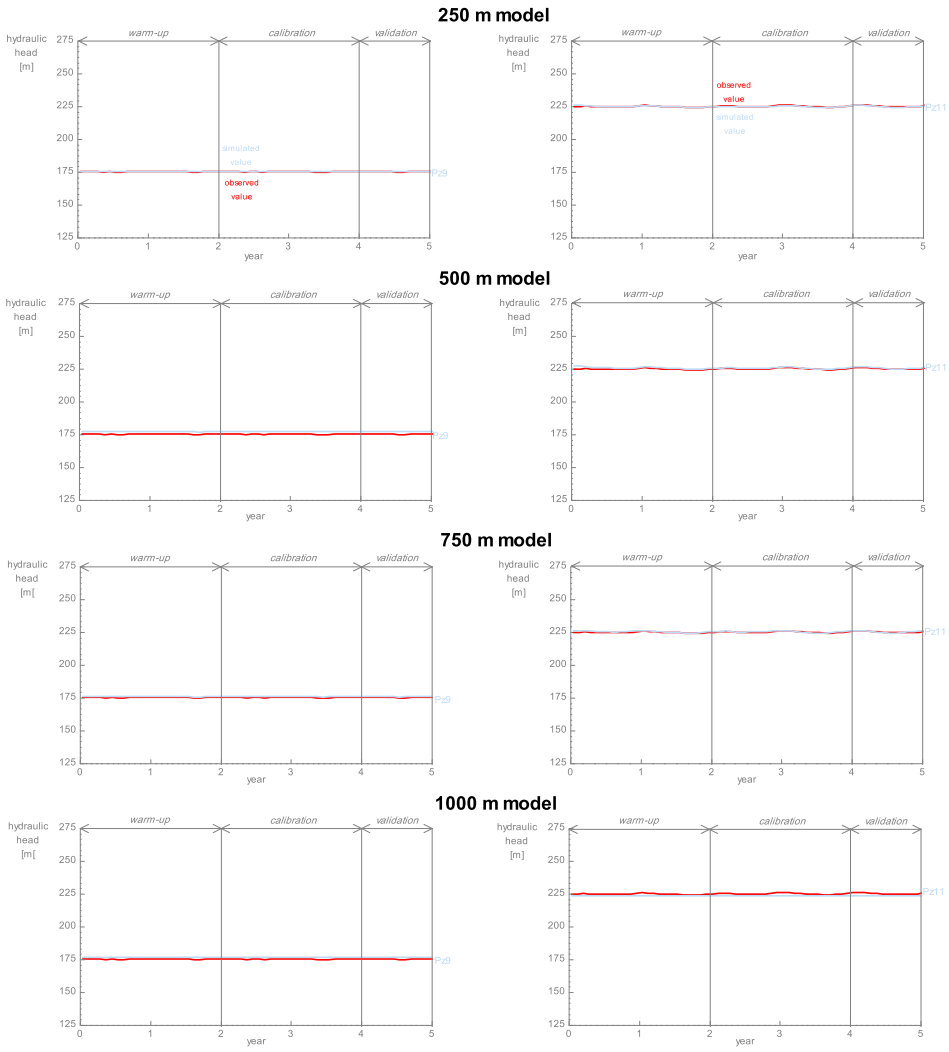

Figure B.34: Comparison of observed and simulated hydraulic heads for Pz9 and Pz11 produced with the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

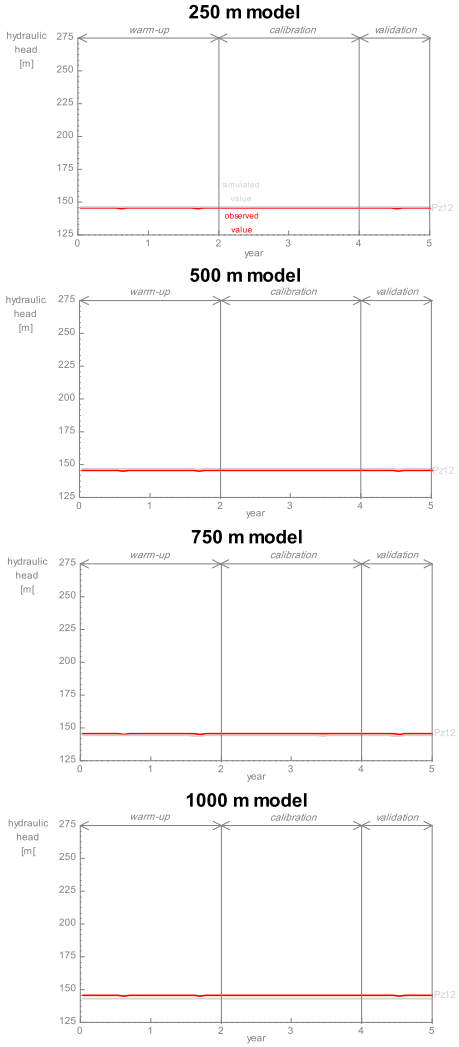

Figure B.35: Comparison of observed and simulated hydraulic heads for Pz12 produced with the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors.

Table B.1: Comparison of reference and optimised parameter values for the *250 m*, *500 m*, *750 m*, and *1000 m* models with average monthly stress factors.

Parameter	Reference value	Optimised value			
		250 m	*500 m*	*750 m*	*1000 m*
K_I	5.00×10^{-7}	1.00×10^{-6}	1.66×10^{-6}	1.92×10^{-7}	7.38×10^{-7}
K_{II}	1.00×10^{-6}	8.48×10^{-7}	2.87×10^{-6}	1.01×10^{-6}	6.34×10^{-7}
K_{III}	1.00×10^{-5}	5.88×10^{-5}	5.49×10^{-5}	5.46×10^{-5}	6.15×10^{-5}
K_{IV}	1.00×10^{-4}	9.91×10^{-5}	1.04×10^{-4}	9.75×10^{-5}	8.05×10^{-5}
K_V	2.50×10^{-4}	4.47×10^{-4}	1.28×10^{-4}	1.66×10^{-4}	4.36×10^{-4}
K_{VI}	5.00×10^{-5}	6.17×10^{-5}	4.67×10^{-5}	4.71×10^{-5}	5.07×10^{-5}
S_{S-I}	1.00×10^{-4}	9.20×10^{-5}	1.08×10^{-4}	1.68×10^{-5}	1.02×10^{-4}
S_{S-II}	1.00×10^{-4}	1.04×10^{-4}	3.52×10^{-4}	5.12×10^{-5}	1.47×10^{-4}
S_{S-III}	1.00×10^{-4}	2.47×10^{-4}	2.43×10^{-4}	5.27×10^{-4}	9.87×10^{-5}
S_{S-IV}	1.00×10^{-4}	9.50×10^{-5}	2.72×10^{-4}	3.45×10^{-5}	1.95×10^{-4}
S_{S-V}	1.00×10^{-4}	1.80×10^{-4}	1.08×10^{-4}	1.81×10^{-5}	3.93×10^{-5}
S_{S-VI}	1.00×10^{-4}	7.42×10^{-5}	1.82×10^{-4}	1.03×10^{-6}	1.02×10^{-4}
θ_{s-I}	4.10×10^{-1}	5.00×10^{-1}	4.50×10^{-1}	4.58×10^{-1}	5.00×10^{-1}
θ_{s-II}	4.10×10^{-1}	2.50×10^{-1}	3.43×10^{-1}	5.00×10^{-1}	5.00×10^{-1}
θ_{s-III}	2.50×10^{-2}	3.61×10^{-2}	1.87×10^{-2}	5.46×10^{-2}	2.29×10^{-2}
θ_{s-IV}	1.00×10^{-1}	7.62×10^{-2}	1.50×10^{-1}	5.66×10^{-2}	1.49×10^{-1}
θ_{s-V}	1.00×10^{-1}	3.70×10^{-2}	1.43×10^{-1}	1.95×10^{-2}	1.50×10^{-1}
θ_{s-VI}	7.50×10^{-2}	5.86×10^{-2}	2.23×10^{-2}	1.34×10^{-1}	4.65×10^{-2}
$S_{wr-I,II}$	9.76×10^{-2}	1.50×10^{-1}	6.59×10^{-2}	2.73×10^{-2}	7.98×10^{-2}
$\alpha_{vG-I,II}$	2.67×10^{0}	2.29×10^{0}	1.83×10^{0}	6.61×10^{-1}	1.10×10^{0}
$\alpha_{vG-IV,V,VI}$	3.65×10^{-2}	4.98×10^{-2}	4.56×10^{-2}	5.04×10^{-2}	3.04×10^{-2}
$\beta_{vG-I,II}$	1.45×10^{0}	1.47×10^{0}	1.32×10^{0}	1.53×10^{0}	1.39×10^{0}
$\beta_{vG-IV,V,VI}$	1.83×10^{0}	2.21×10^{0}	2.42×10^{0}	2.18×10^{0}	1.77×10^{0}
n_{x-1}	1.20×10^{-2}	1.84×10^{-2}	1.40×10^{-2}	3.06×10^{-1}	2.27×10^{-2}
n_{x-2}	2.00×10^{-1}	3.80×10^{-1}	5.27×10^{-1}	1.03×10^{-1}	7.50×10^{-1}
n_{x-3}	6.00×10^{-1}	5.30×10^{-1}	5.11×10^{-1}	7.50×10^{-1}	7.50×10^{-1}
n_{x-4}	2.50×10^{-2}	5.13×10^{-1}	4.05×10^{-2}	5.04×10^{-1}	6.12×10^{-1}
H_{sto-1}	2.00×10^{-3}	1.49×10^{-3}	7.94×10^{-4}	8.19×10^{-4}	9.46×10^{-3}
H_{sto-2}	2.00×10^{-3}	1.16×10^{-3}	2.80×10^{-3}	6.83×10^{-4}	4.53×10^{-3}
H_{sto-3}	2.00×10^{-3}	3.49×10^{-3}	1.67×10^{-3}	1.89×10^{-3}	2.86×10^{-3}
H_{sto-4}	2.00×10^{-3}	9.56×10^{-4}	1.56×10^{-3}	1.00×10^{-2}	3.76×10^{-3}
$L_{c-1,2,3,4}$	1.00×10^{-1}	2.00×10^{-2}	3.18×10^{-1}	2.00×10^{-1}	1.12×10^{-1}

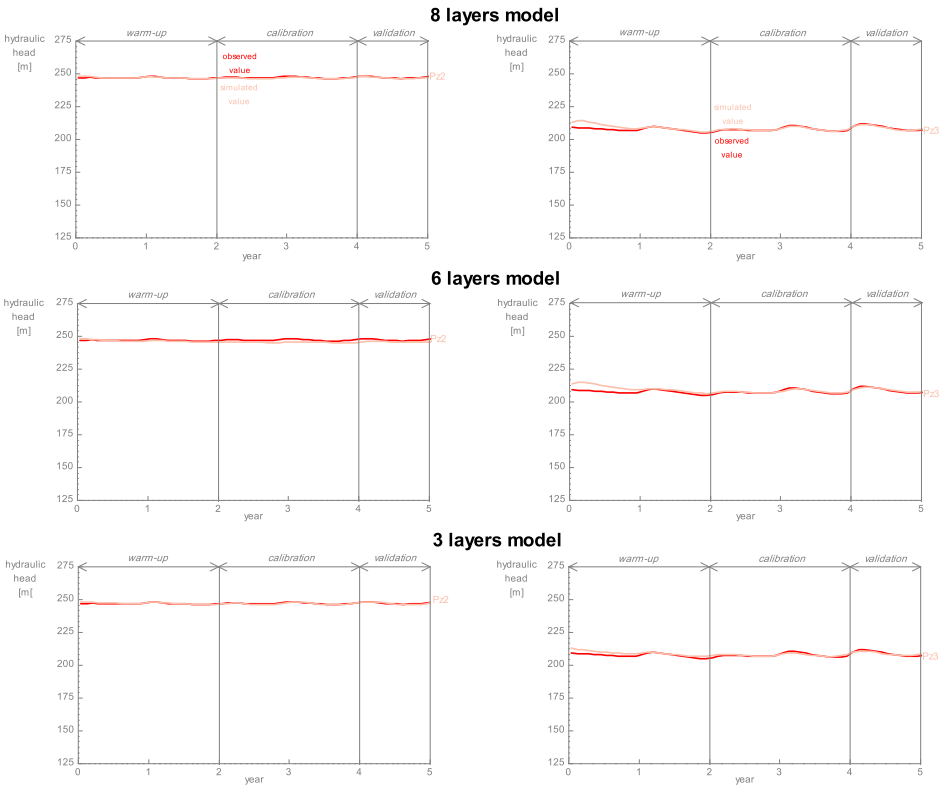

Figure B.36: Comparison of observed and simulated hydraulic heads for Pz2 and Pz3 produced with the optimised *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

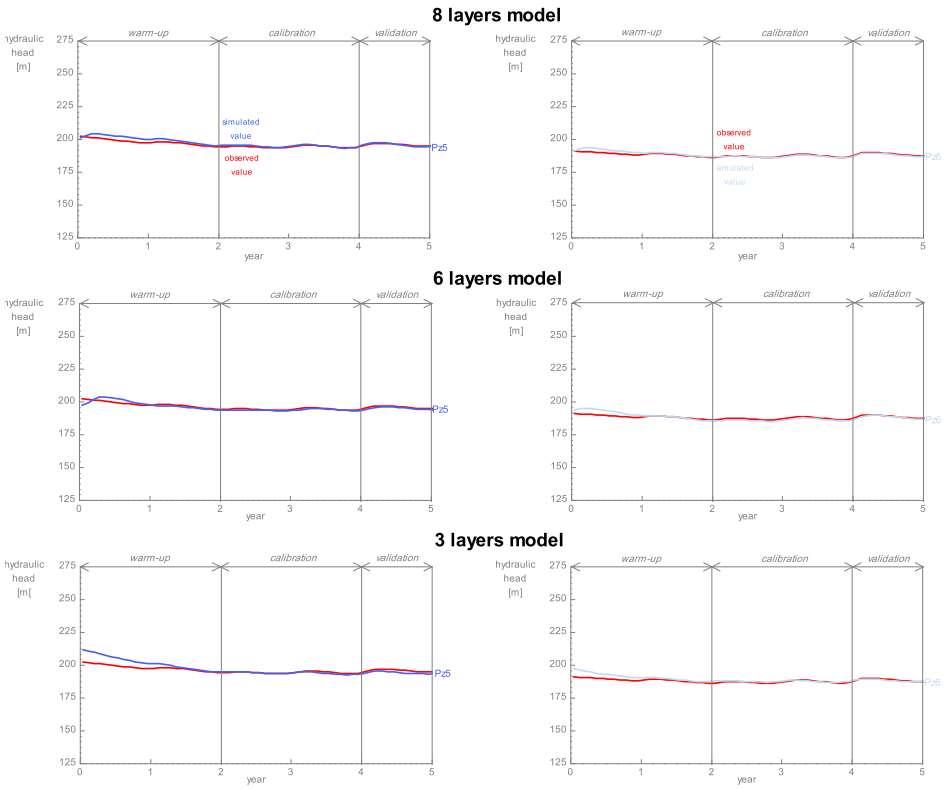

Figure B.37: Comparison of observed and simulated hydraulic heads for Pz5 and Pz6 produced with the optimised *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

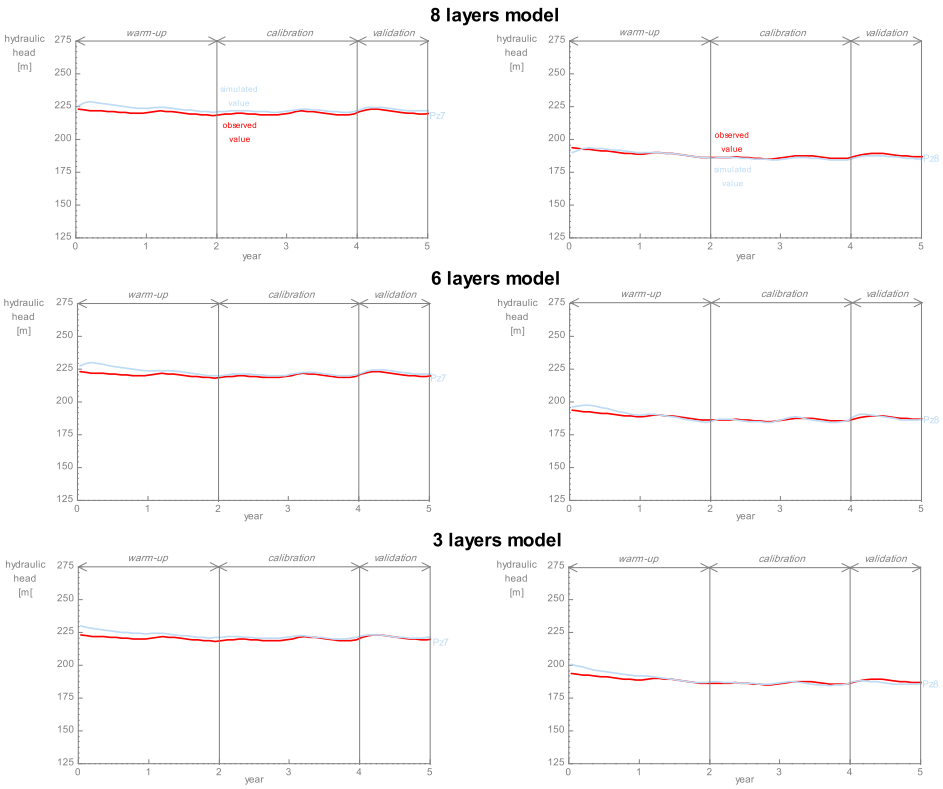

Figure B.38: Comparison of observed and simulated hydraulic heads for Pz7 and Pz8 produced with the optimised *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

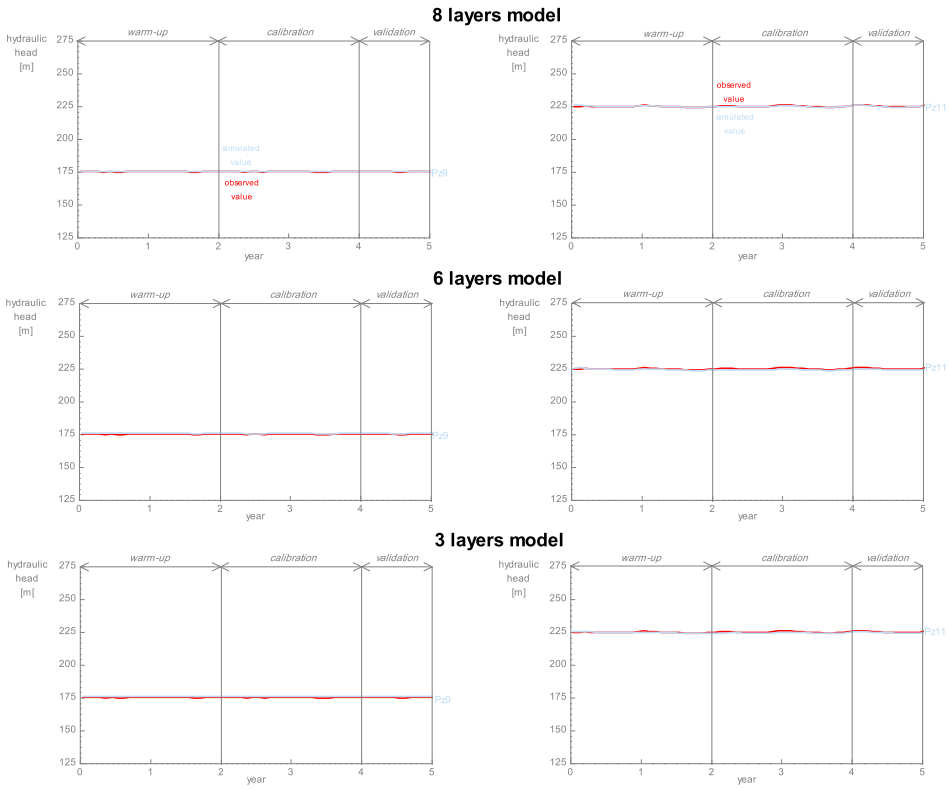

Figure B.39: Comparison of observed and simulated hydraulic heads for Pz9 and Pz11 produced with the optimised *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

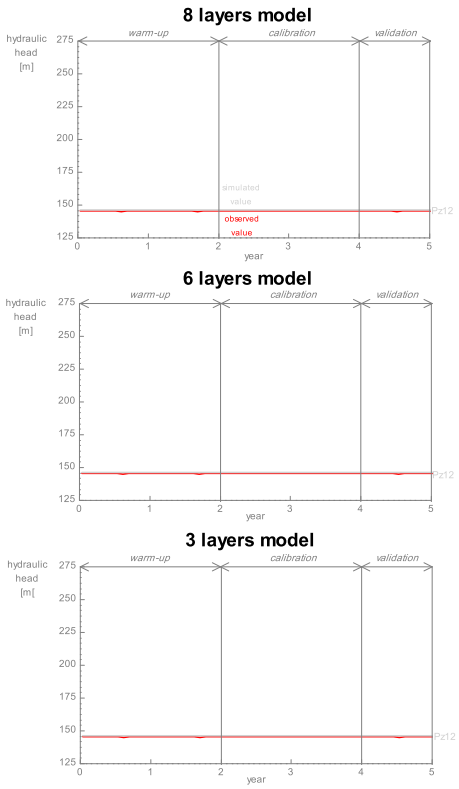

Figure B.40: Comparison of observed and simulated hydraulic heads for Pz12 produced with the optimised *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors.

Table B.2: Comparison of reference and optimised parameter values for the *8 layers*, *6 layers*, and *3 layers* models with average monthly stress factors.

Parameter	Reference value	Optimised value		
		8 layers	*6 layers*	*3 layers*
K_I	5.00×10^{-7}	1.00×10^{-6}	2.53×10^{-7}	2.66×10^{-7}
K_{II}	1.00×10^{-6}	8.48×10^{-7}	4.32×10^{-5}	1.93×10^{-7}
K_{III}	1.00×10^{-5}	5.88×10^{-5}	2.75×10^{-5}	7.41×10^{-6}
K_{IV}	1.00×10^{-4}	9.91×10^{-5}	1.44×10^{-4}	1.52×10^{-4}
K_V	2.50×10^{-4}	4.47×10^{-4}	9.30×10^{-5}	1.37×10^{-4}
K_{VI}	5.00×10^{-5}	6.17×10^{-5}	1.10×10^{-4}	3.80×10^{-5}
S_{S-I}	1.00×10^{-4}	9.20×10^{-5}	1.08×10^{-4}	2.70×10^{-5}
S_{S-II}	1.00×10^{-4}	1.04×10^{-4}	1.89×10^{-4}	1.36×10^{-3}
S_{S-III}	1.00×10^{-4}	2.47×10^{-4}	3.68×10^{-4}	2.19×10^{-4}
S_{S-IV}	1.00×10^{-4}	9.50×10^{-5}	3.57×10^{-4}	3.68×10^{-4}
S_{S-V}	1.00×10^{-4}	1.80×10^{-4}	1.63×10^{-3}	8.62×10^{-6}
S_{S-VI}	1.00×10^{-4}	7.42×10^{-5}	9.30×10^{-5}	5.54×10^{-5}
θ_{s-I}	4.10×10^{-1}	5.00×10^{-1}	5.00×10^{-1}	5.00×10^{-1}
θ_{s-II}	4.10×10^{-1}	2.50×10^{-1}	2.50×10^{-1}	5.00×10^{-1}
θ_{s-III}	2.50×10^{-2}	3.61×10^{-2}	2.01×10^{-2}	4.61×10^{-1}
θ_{s-IV}	1.00×10^{-1}	7.62×10^{-2}	2.25×10^{-2}	7.48×10^{-2}
θ_{s-V}	1.00×10^{-1}	3.70×10^{-2}	2.28×10^{-2}	1.50×10^{-1}
θ_{s-VI}	7.50×10^{-2}	5.86×10^{-2}	1.22×10^{-2}	7.14×10^{-2}
$S_{wr-I,II}$	9.76×10^{-2}	1.50×10^{-1}	2.09×10^{-1}	2.44×10^{-1}
$\alpha_{vG-I,II}$	2.67×10^{0}	2.29×10^{0}	1.39×10^{0}	1.08×10^{1}
$\alpha_{vG-IV,V,VI}$	3.65×10^{-2}	4.98×10^{-2}	8.08×10^{-2}	4.57×10^{-2}
$\beta_{vG-I,II}$	1.45×10^{0}	1.47×10^{0}	1.85×10^{0}	1.54×10^{0}
$\beta_{vG-IV,V,VI}$	1.83×10^{0}	2.21×10^{0}	2.52×10^{0}	2.49×10^{0}
n_{x-1}	1.20×10^{-2}	1.84×10^{-2}	5.34×10^{-2}	1.36×10^{-2}
n_{x-2}	2.00×10^{-1}	3.80×10^{-1}	5.02×10^{-1}	7.50×10^{-1}
n_{x-3}	6.00×10^{-1}	5.30×10^{-1}	1.53×10^{-1}	7.50×10^{-1}
n_{x-4}	2.50×10^{-2}	5.13×10^{-1}	3.94×10^{-1}	7.50×10^{-1}
H_{sto-1}	2.00×10^{-3}	1.49×10^{-3}	2.45×10^{-3}	9.70×10^{-3}
H_{sto-2}	2.00×10^{-3}	1.16×10^{-3}	3.92×10^{-3}	3.68×10^{-3}
H_{sto-3}	2.00×10^{-3}	3.49×10^{-3}	1.00×10^{-2}	2.52×10^{-3}
H_{sto-4}	2.00×10^{-3}	9.56×10^{-4}	1.05×10^{-3}	3.84×10^{-3}
$L_{c-1,2,3,4}$	1.00×10^{-1}	2.00×10^{-2}	1.70×10^{-1}	5.25×10^{-2}

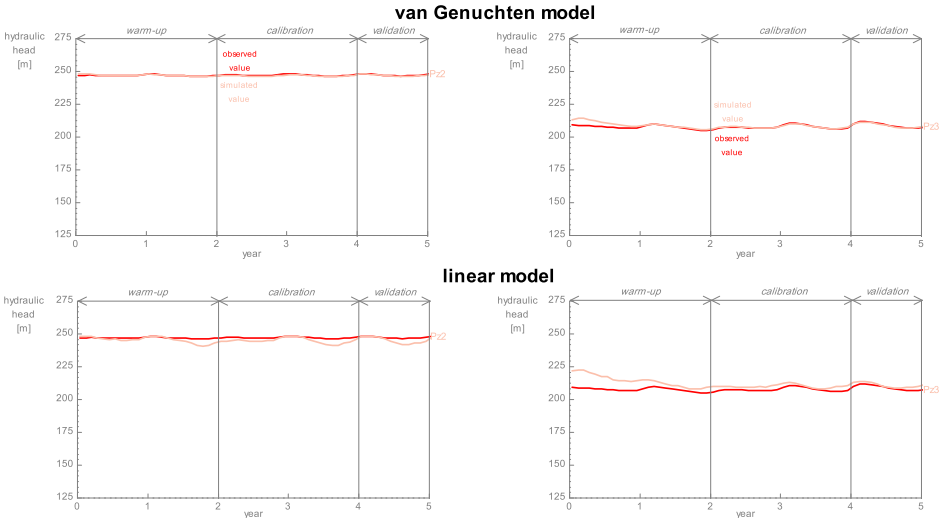

Figure B.41: Comparison of observed and simulated hydraulic heads for Pz2 and Pz3 produced with the optimised *van Genuchten* and *linear* models with monthly stress factors.

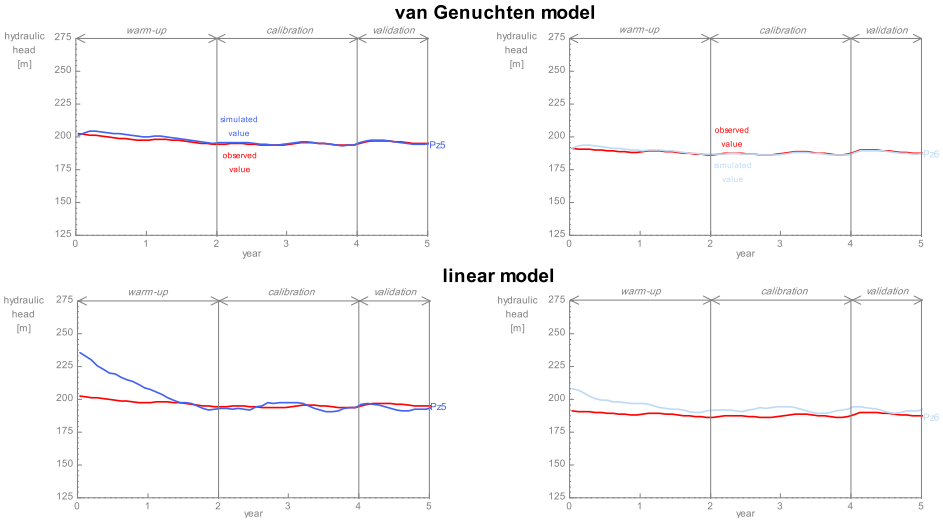

Figure B.42: Comparison of observed and simulated hydraulic heads for Pz5 and Pz6 produced with the optimised *van Genuchten* and *linear* models with monthly stress factors.

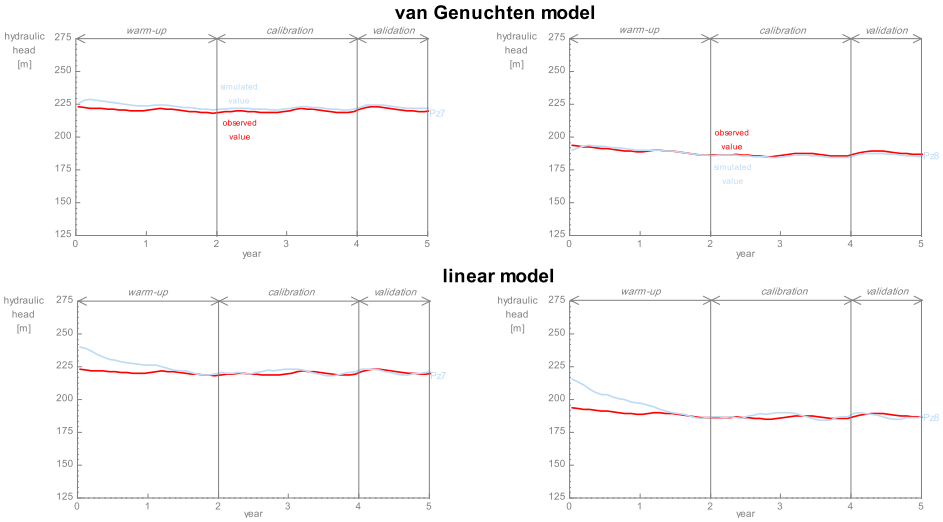

Figure B.43: Comparison of observed and simulated hydraulic heads for Pz7 and Pz8 produced with the optimised *van Genuchten* and *linear* models with monthly stress factors.

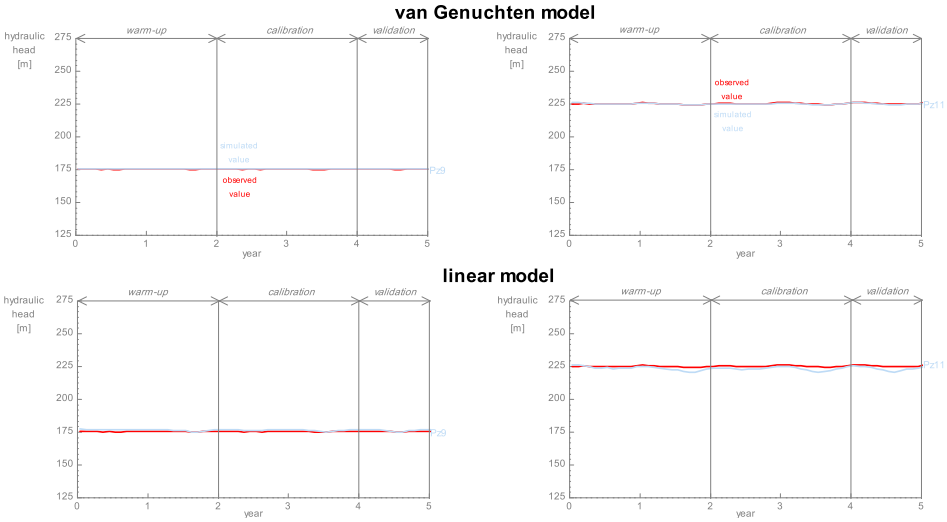

Figure B.44: Comparison of observed and simulated hydraulic heads for Pz9 and Pz11 produced with the optimised *van Genuchten* and *linear* models with monthly stress factors.

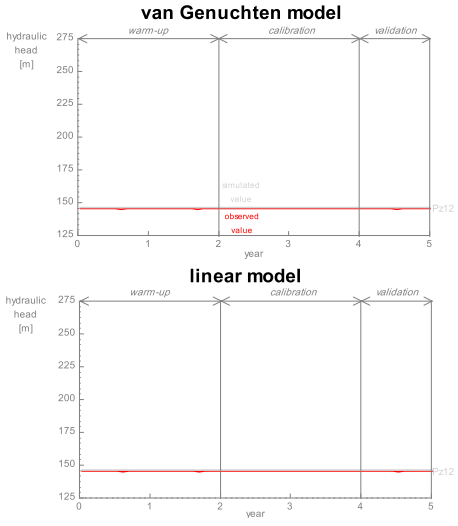

Figure B.45: Comparison of observed and simulated hydraulic heads for Pz12 produced with the optimised *van Genuchten* and *linear* models with monthly stress factors.

Table B.3: Comparison of reference and optimised parameter values for the *van Genuchten* and *linear* models with average monthly stress factors.

Parameter	Reference value	Optimised value	
		van Genuchten	*linear*
K_I	5.00×10^{-7}	1.00×10^{-6}	7.44×10^{-8}
K_{II}	1.00×10^{-6}	8.48×10^{-7}	6.38×10^{-7}
K_{III}	1.00×10^{-5}	5.88×10^{-5}	2.43×10^{-6}
K_{IV}	1.00×10^{-4}	9.91×10^{-5}	1.87×10^{-4}
K_V	2.50×10^{-4}	4.47×10^{-4}	5.34×10^{-5}
K_{VI}	5.00×10^{-5}	6.17×10^{-5}	1.40×10^{-5}
S_{S-I}	1.00×10^{-4}	9.20×10^{-5}	1.92×10^{-4}
S_{S-II}	1.00×10^{-4}	1.04×10^{-4}	1.29×10^{-4}
S_{S-III}	1.00×10^{-4}	2.47×10^{-4}	2.02×10^{-3}
S_{S-IV}	1.00×10^{-4}	9.50×10^{-5}	1.63×10^{-4}
S_{S-V}	1.00×10^{-4}	1.80×10^{-4}	9.59×10^{-5}
S_{S-VI}	1.00×10^{-4}	7.42×10^{-5}	1.56×10^{-4}
θ_{s-I}	4.10×10^{-1}	5.00×10^{-1}	5.00×10^{-1}
θ_{s-II}	4.10×10^{-1}	2.50×10^{-1}	3.68×10^{-1}
θ_{s-III}	2.50×10^{-2}	3.61×10^{-2}	4.72×10^{-2}
θ_{s-IV}	1.00×10^{-1}	7.62×10^{-2}	1.50×10^{-1}
θ_{s-V}	1.00×10^{-1}	3.70×10^{-2}	1.50×10^{-1}
θ_{s-VI}	7.50×10^{-2}	5.86×10^{-2}	1.50×10^{-1}
$S_{wr-I,II}$	9.76×10^{-2}	1.50×10^{-1}	2.44×10^{-1}
$\alpha_{vG}/h_{a-I,II}$	2.67×10^{0} / -1.20×10^{1}	2.29×10^{0}	-3.18×10^{1}
$\alpha_{vG}/h_{a-IV,V,VI}$	3.65×10^{-2} / -2.20×10^{2}	4.98×10^{-2}	-3.01×10^{2}
$\beta_{vG}/h_{b-I,II}$	1.45×10^{0} / -2.07×10^{-1}	1.47×10^{0}	-5.07×10^{-3}
$\beta_{vG}/h_{b-IV,V,VI}$	1.83×10^{0} / -1.30×10^{1}	2.21×10^{0}	-1.88×10^{1}
n_{x-1}	1.20×10^{-2}	1.84×10^{-2}	3.10×10^{-2}
n_{x-2}	2.00×10^{-1}	3.80×10^{-1}	1.16×10^{-1}
n_{x-3}	6.00×10^{-1}	5.30×10^{-1}	3.24×10^{-1}
n_{x-4}	2.50×10^{-2}	5.13×10^{-1}	4.25×10^{-3}
H_{sto-1}	2.00×10^{-3}	1.49×10^{-3}	7.59×10^{-4}
H_{sto-2}	2.00×10^{-3}	1.16×10^{-3}	1.00×10^{-2}
H_{sto-3}	2.00×10^{-3}	3.49×10^{-3}	1.94×10^{-4}
H_{sto-4}	2.00×10^{-3}	9.56×10^{-4}	2.74×10^{-4}
$L_{c-1,2,3,4}$	1.00×10^{-1}	2.00×10^{-2}	7.78×10^{-2}

Appendix C

Complement to Chapter 7

Table C.1: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with daily stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	250 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.46	0.09	0.33	0.19	0.79	0.40	1.58	0.76	0.43	0.99	0.23	0.73
MBE_h [%]	0.23	-0.01	-0.15	-0.03	0.40	-0.20	0.72	-0.41	0.24	0.41	-0.10	0.50
$HHVE_h^1$ [%]	3.36	-5.93	-6.34	19.69	5.19	8.70	-9.46	1.99	-13.73	-15.41	4.84	-45.26
$HHVE_h^2$ [%]	-0.39	-1.15	5.64	7.27	1.36	2.47	0.33	1.23	-23.64	4.04	-1.70	-43.33
	500 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.58	0.40	0.17	1.57	1.59	2.94	1.44	2.46	2.06	0.18	0.84	1.27
MBE_h [%]	-0.24	0.15	0.00	-0.67	-0.82	1.57	0.65	-1.32	1.18	0.01	0.37	0.87
$HHVE_h^1$ [%]	28.57	-4.44	-2.42	21.26	-0.74	26.96	-8.11	6.62	-21.57	-7.89	6.45	-47.37
$HHVE_h^2$ [%]	29.34	2.30	-1.04	9.55	-0.45	8.64	0.66	4.92	-38.18	10.43	-5.11	-48.33
	750 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	3.59	0.94	1.11	3.33	0.78	0.67	1.09	1.68	0.84	0.54	0.49	1.69
MBE_h [%]	-1.78	-0.36	0.53	-1.43	0.40	0.35	0.50	0.90	0.48	0.22	0.20	-1.16
$HHVE_h^1$ [%]	0.84	-24.44	5.44	15.75	-9.63	4.35	-8.78	1.32	-29.41	-12.90	-17.74	-3.16
$HHVE_h^2$ [%]	-12.74	-32.18	8.56	5.91	-0.45	3.70	-0.33	6.97	-36.36	5.74	-27.27	16.67
	1000 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.87	0.52	0.83	0.41	1.13	0.85	1.32	4.20	1.66	0.52	1.38	2.66
MBE_h [%]	0.43	0.21	0.33	0.17	-0.58	-0.45	-0.59	-2.26	0.94	-0.20	-0.59	-1.82
$HHVE_h^1$ [%]	6.72	-6.67	-20.24	12.60	0.00	6.96	12.84	6.62	-39.22	-3.23	-58.06	-28.42
$HHVE_h^2$ [%]	9.65	3.45	-22.96	8.64	9.50	-0.82	-0.66	0.82	-45.45	2.34	-63.07	-1.67

Table C.2: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *8 layers*, *6 layers*, and *3 layers* models with daily stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	8 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.46	0.09	0.33	0.19	0.79	0.40	1.58	0.76	0.43	0.99	0.23	0.73
MBE_h [%]	0.23	-0.01	-0.15	-0.03	0.40	-0.20	0.72	-0.41	0.24	0.41	-0.10	0.50
$HHVE_h^1$ [%]	3.36	-5.93	-6.34	19.69	5.19	8.70	-9.46	1.99	-13.73	-15.41	4.84	-45.26
$HHVE_h^2$ [%]	-0.39	-1.15	5.64	7.27	1.36	2.47	0.33	1.23	-23.64	4.04	-1.70	-43.33
	6 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	1.94	0.29	0.72	1.18	0.88	1.24	2.82	0.34	0.41	1.92	0.42	0.70
MBE_h [%]	-0.96	-0.11	0.29	-0.49	-0.44	0.66	1.28	0.13	0.23	0.77	-0.18	0.48
$HHVE_h^1$ [%]	10.92	7.41	-38.67	43.31	16.30	20.87	-5.41	25.17	-7.84	-46.24	2.42	-18.95
$HHVE_h^2$ [%]	-8.11	10.92	-18.79	48.18	14.03	12.35	-6.25	31.15	-10.91	-33.19	5.68	-18.33
	3 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	7.96	0.27	3.32	9.82	12.67	8.18	7.92	10.90	0.46	4.15	0.29	0.74
MBE_h [%]	3.91	-0.06	1.55	4.21	6.49	4.35	3.59	5.83	0.26	1.69	-0.12	0.50
$HHVE_h^1$ [%]	139.50	17.04	-35.65	114.17	106.67	64.35	16.89	38.41	-68.63	-38.35	-15.32	-87.37
$HHVE_h^2$ [%]	51.35	20.11	-24.22	79.09	82.35	23.87	-2.30	40.16	-78.18	-35.96	-1.70	-85.00

Table C.3: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *van Genuchten* and *linear* models with daily stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	van Genuchten											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.46	0.09	0.33	0.19	0.79	0.40	1.58	0.76	0.43	0.99	0.23	0.73
MBE_h [%]	0.23	-0.01	-0.15	-0.03	0.40	-0.20	0.72	-0.41	0.24	0.41	-0.10	0.50
$HHVE_h^1$ [%]	3.36	-5.93	-6.34	19.69	5.19	8.70	-9.46	1.99	-13.73	-15.41	4.84	-45.26
$HHVE_h^2$ [%]	-0.39	-1.15	5.64	7.27	1.36	2.47	0.33	1.23	-23.64	4.04	-1.70	-43.33
	linear											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	4.81	1.61	5.49	8.39	11.08	10.19	10.94	12.80	0.41	6.95	0.99	0.70
MBE_h [%]	1.54	-0.58	2.55	3.23	5.38	5.30	4.89	6.65	0.22	2.85	-0.28	0.48
$HHVE_h^1$ [%]	635.29	132.59	-14.50	579.53	537.04	372.17	214.19	390.07	162.75	-29.75	112.10	44.21
$HHVE_h^2$ [%]	348.65	155.17	-7.31	415.91	417.65	197.94	112.50	298.36	141.82	-43.40	119.32	111.67

Table C.4: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	250 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.45	0.24	1.01	1.26	0.36	1.12	0.85	1.93	0.45	0.28	0.30	0.74
MBE_h [%]	-0.21	-0.08	-0.47	-0.54	-0.17	-0.59	0.38	-1.03	0.26	0.01	-0.13	0.50
$HHVE_h^1$ [%]	19.82	-14.29	4.12	24.39	9.02	14.55	13.45	6.16	-80.95	5.18	-22.55	-91.67
$HHVE_h^2$ [%]	-8.91	3.66	-3.82	-2.35	-7.41	-5.65	-12.24	-6.20	-76.92	-7.63	2.10	-86.96
	500 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	1.36	0.26	0.55	2.75	2.58	2.05	0.79	3.65	2.08	0.89	0.80	1.28
MBE_h [%]	-0.66	0.09	-0.25	-1.18	-1.33	1.09	0.35	-1.96	1.18	-0.36	0.35	0.88
$HHVE_h^1$ [%]	51.35	-5.88	0.52	26.83	5.26	30.91	9.24	13.70	-85.71	13.47	-18.63	-91.67
$HHVE_h^2$ [%]	24.70	7.93	-5.52	-3.76	-9.26	5.65	-7.48	0.41	-76.92	-0.22	0.00	-86.96
	750 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	4.38	0.99	0.56	4.37	0.16	0.13	0.46	0.58	0.86	0.42	0.46	1.64
MBE_h [%]	-2.18	-0.38	0.26	-1.89	-0.07	0.04	0.20	0.31	0.49	-0.15	0.19	-1.12
$HHVE_h^1$ [%]	12.61	-18.49	0.00	19.51	-1.50	6.36	1.68	6.16	-80.95	3.63	-6.86	-91.67
$HHVE_h^2$ [%]	-19.84	-28.05	2.76	-3.29	-8.80	-1.30	-8.16	-2.07	-73.08	-3.92	-13.99	-78.26
	1000 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.12	0.37	0.54	0.73	2.05	1.52	1.85	5.32	1.68	1.18	1.33	2.61
MBE_h [%]	0.04	0.14	0.13	-0.31	-1.05	-0.81	-0.84	-2.86	0.96	-0.49	-0.57	-1.79
$HHVE_h^1$ [%]	21.62	-12.61	-29.90	16.26	6.02	13.64	-4.20	6.16	-85.71	3.63	-98.04	-94.44
$HHVE_h^2$ [%]	0.00	10.98	-29.30	-6.57	-0.93	-8.26	-4.42	-6.20	-80.77	-1.53	-94.41	-78.26

Table C.5: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	8 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.45	0.24	1.01	1.26	0.36	1.12	0.85	1.93	0.45	0.28	0.30	0.74
MBE_h [%]	-0.21	-0.08	-0.47	-0.54	-0.17	-0.59	0.38	-1.03	0.26	0.01	-0.13	0.50
$HHVE_h^1$ [%]	19.82	-14.29	4.12	24.39	9.02	14.55	13.45	6.16	-80.95	5.18	-22.55	-91.67
$HHVE_h^2$ [%]	-8.91	3.66	-3.82	-2.35	-7.41	-5.65	-12.24	-6.20	-76.92	-7.63	2.10	-86.96
	6 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	2.66	0.51	0.53	2.15	1.85	0.52	2.19	0.88	0.44	1.38	0.49	0.72
MBE_h [%]	-1.32	-0.20	0.02	-0.92	-0.95	0.26	0.99	-0.46	0.25	0.52	-0.22	0.49
$HHVE_h^1$ [%]	15.32	-19.33	-13.40	41.46	9.77	24.55	14.29	28.08	-76.19	-34.20	-15.69	-88.89
$HHVE_h^2$ [%]	-10.53	10.98	-25.27	37.09	5.09	7.83	-10.54	21.90	-69.23	-39.43	9.09	-82.61
	3 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	8.24	0.29	3.16	9.84	12.82	8.20	7.78	10.87	0.46	3.96	0.31	0.74
MBE_h [%]	4.06	-0.08	1.47	4.23	6.57	4.36	3.53	5.82	0.26	1.61	-0.13	0.50
$HHVE_h^1$ [%]	66.67	8.40	-22.68	43.09	33.83	7.27	-2.52	-13.70	-90.48	-37.31	-26.47	-97.22
$HHVE_h^2$ [%]	25.51	25.00	-27.39	49.30	47.69	12.17	-8.16	17.36	-92.31	-38.56	0.70	-91.30

Table C.6: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the *van Genuchten* and *linear* models with monthly stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	van Genuchten											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.45	0.24	1.01	1.26	0.36	1.12	0.85	1.93	0.45	0.28	0.30	0.74
MBE_h [%]	-0.21	-0.08	-0.47	-0.54	-0.17	-0.59	0.38	-1.03	0.26	0.01	-0.13	0.50
$HHVE_h^1$ [%]	19.82	-14.29	4.12	24.39	9.02	14.55	13.45	6.16	-80.95	5.18	-22.55	-91.67
$HHVE_h^2$ [%]	-8.91	3.66	-3.82	-2.35	-7.41	-5.65	-12.24	-6.20	-76.92	-7.63	2.10	-86.96
	linear											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	6.58	1.57	6.30	10.35	13.27	11.47	12.11	14.56	0.46	7.18	0.92	0.74
MBE_h [%]	2.77	-0.57	2.98	4.18	6.57	6.02	5.45	7.65	0.26	2.95	-0.23	0.51
$HHVE_h^1$ [%]	600.90	99.16	22.16	593.50	517.29	311.82	205.88	332.19	-85.71	-14.51	97.06	-94.44
$HHVE_h^2$ [%]	302.02	158.54	-33.97	350.23	359.26	157.83	71.77	224.38	-73.08	-45.10	144.06	-91.30

Table C.7: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	250 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.26	0.42	0.33	0.53	0.68	0.36	1.99	1.00	0.42	0.41	0.22	0.74
MBE_h [%]	-0.05	-0.16	0.01	-0.20	0.26	-0.15	0.90	-0.48	0.24	0.00	-0.09	0.51
$HHVE_h^1$ [%]	3.60	-40.34	-19.59	20.33	49.62	4.55	-4.20	10.96	-95.24	-9.33	-23.53	-77.78
$HHVE_h^2$ [%]	8.91	-9.15	-15.29	17.37	38.43	-9.57	-19.73	-11.16	-88.46	-10.89	0.00	8.70
	500 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	1.21	0.20	0.49	0.91	0.55	2.83	1.78	2.20	2.00	0.42	0.57	1.26
MBE_h [%]	0.38	-0.02	-0.23	0.29	-0.09	1.47	0.77	-1.09	1.14	0.00	0.25	0.86
$HHVE_h^1$ [%]	98.20	5.04	-4.64	16.26	5.26	-4.55	-18.49	0.68	-85.71	-19.69	-2.94	-83.33
$HHVE_h^2$ [%]	93.93	40.24	6.79	-11.27	-4.63	-49.13	-37.41	-42.98	-80.77	-23.75	24.48	39.13
	750 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	1.41	1.47	0.90	1.49	1.58	0.27	1.07	0.84	0.97	0.73	0.40	1.41
MBE_h [%]	-0.68	-0.58	0.37	-0.62	0.79	0.09	0.48	-0.32	0.55	0.25	0.12	-0.96
$HHVE_h^1$ [%]	-5.41	7.56	-21.13	49.59	8.27	7.27	-7.56	52.74	9.52	-10.88	7.84	-36.11
$HHVE_h^2$ [%]	-23.08	28.66	-28.45	35.21	-10.19	11.30	-15.65	40.08	96.15	-14.38	16.08	91.30
	1000 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.47	0.46	0.68	0.94	0.83	1.43	0.36	1.95	1.63	1.01	1.32	2.51
MBE_h [%]	0.05	-0.13	0.09	0.39	-0.34	0.74	0.01	-0.98	0.93	-0.38	-0.56	-1.72
$HHVE_h^1$ [%]	61.26	-29.41	-30.93	14.63	27.82	10.00	-26.89	15.75	-90.48	-24.35	-98.04	-86.11
$HHVE_h^2$ [%]	63.16	46.95	-35.03	-6.10	4.63	-7.83	-15.65	-21.90	-88.46	-19.83	-96.50	-26.09

Table C.8: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the optimised *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	8 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.26	0.42	0.33	0.53	0.68	0.36	1.99	1.00	0.42	0.41	0.22	0.74
MBE_h [%]	-0.05	-0.16	0.01	-0.20	0.26	-0.15	0.90	-0.48	0.24	0.00	-0.09	0.51
$HHVE_h^1$ [%]	3.60	-40.34	-19.59	20.33	49.62	4.55	-4.20	10.96	-95.24	-9.33	-23.53	-77.78
$HHVE_h^2$ [%]	8.91	-9.15	-15.29	17.37	38.43	-9.57	-19.73	-11.16	-88.46	-10.89	0.00	8.70
	6 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.59	1.71	0.53	0.51	0.67	1.07	1.25	0.81	0.59	1.11	0.80	0.82
MBE_h [%]	0.28	-0.67	0.06	-0.02	-0.28	-0.56	0.56	-0.18	0.34	0.06	-0.35	0.56
$HHVE_h^1$ [%]	29.73	-57.14	-28.35	45.53	-52.63	0.91	18.49	51.37	-52.38	-52.85	-50.98	-80.56
$HHVE_h^2$ [%]	12.15	-51.83	-30.57	64.32	-28.70	-7.39	-5.10	79.34	-34.62	-71.02	-23.08	-65.22
	3 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.60	0.32	0.77	0.91	0.68	0.69	1.62	0.90	0.63	1.10	0.61	0.76
MBE_h [%]	0.15	-0.10	0.01	-0.06	-0.14	0.21	0.67	-0.05	0.36	-0.25	-0.25	0.52
$HHVE_h^1$ [%]	-36.04	4.20	-30.93	43.09	9.77	-6.36	-3.36	18.49	-90.48	-48.19	-65.69	-94.44
$HHVE_h^2$ [%]	-23.89	23.17	-39.70	61.03	1.85	-10.00	-23.13	13.64	-80.77	-54.03	-42.66	-91.30

Table C.9: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the optimised *van Genuchten* and *linear* models with monthly stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the calibration period. $HHVE_h^2$ corresponds to $HHVE_h$ for the 2nd year of the calibration period.

	van Genuchten											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.26	0.42	0.33	0.53	0.68	0.36	1.99	1.00	0.42	0.41	0.22	0.74
MBE_h [%]	-0.05	-0.16	0.01	-0.20	0.26	-0.15	0.90	-0.48	0.24	0.00	-0.09	0.51
$HHVE_h^1$ [%]	3.60	-40.34	-19.59	20.33	49.62	4.55	-4.20	10.96	-95.24	-9.33	-23.53	-77.78
$HHVE_h^2$ [%]	8.91	-9.15	-15.29	17.37	38.43	-9.57	-19.73	-11.16	-88.46	-10.89	0.00	8.70
	linear											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	4.14	2.75	2.73	2.59	2.47	5.04	2.02	2.34	1.14	4.12	2.06	0.75
MBE_h [%]	1.66	-0.93	1.20	-0.35	-0.16	2.55	0.56	0.57	0.61	-1.61	-0.84	0.51
$HHVE_h^1$ [%]	420.72	221.01	26.29	339.02	304.51	188.18	193.28	194.52	157.14	38.86	118.63	-75.00
$HHVE_h^2$ [%]	190.69	317.07	0.00	214.08	212.04	110.87	71.09	141.32	546.15	6.32	197.20	-60.87

Table C.10: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the optimised *250 m*, *500 m*, *750 m*, and *1000 m* models with monthly stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the validation period.

	250 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.66	0.28	0.40	0.90	0.47	0.54	1.70	1.71	0.45	0.37	0.24	0.75
MBE_h [%]	-0.28	-0.11	-0.13	-0.35	0.03	-0.28	0.76	-0.90	0.24	-0.14	-0.10	0.51
$HHVE_h^1$ [%]	7.34	1.88	-11.65	18.52	34.04	-5.84	-15.95	-14.93	-96.08	-4.17	1.27	-52.94
	500 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	1.57	0.27	0.47	0.97	1.87	1.63	0.83	4.35	2.01	0.38	0.56	1.25
MBE_h [%]	0.00	-0.01	-0.21	-0.40	-0.93	0.79	0.29	-2.29	1.15	-0.09	0.24	0.85
$HHVE_h^1$ [%]	102.10	36.25	4.14	-20.16	-7.23	-57.30	-40.74	-53.73	-92.16	-15.67	24.68	-35.29
	750 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	2.18	1.55	0.83	1.97	0.68	0.35	0.82	1.25	0.97	0.26	0.32	1.41
MBE_h [%]	-1.07	-0.62	0.17	-0.83	0.34	-0.16	0.36	-0.63	0.55	0.03	0.05	-0.97
$HHVE_h^1$ [%]	-19.23	29.38	-26.88	39.51	-5.96	17.88	-9.97	40.30	0.00	-10.52	6.33	-17.65
	1000 m											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.83	0.42	1.01	0.37	1.78	0.68	0.70	3.22	1.65	1.47	1.48	2.49
MBE_h [%]	-0.11	-0.04	-0.32	0.16	-0.90	0.34	-0.29	-1.70	0.94	-0.59	-0.62	-1.70
$HHVE_h^1$ [%]	66.78	48.75	-32.52	-2.88	0.85	-18.98	-20.80	-23.88	-96.08	-18.85	-96.20	-80.88

Table C.11: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the optimised *8 layers*, *6 layers*, and *3 layers* models with monthly stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the validation period.

	8 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.66	0.28	0.40	0.90	0.47	0.54	1.70	1.71	0.45	0.37	0.24	0.75
MBE_h [%]	-0.28	-0.11	-0.13	-0.35	0.03	-0.28	0.76	-0.90	0.24	-0.14	-0.10	0.51
$HHVE_h^1$ [%]	7.34	1.88	-11.65	18.52	34.04	-5.84	-15.95	-14.93	-96.08	-4.17	1.27	-52.94
	6 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.42	1.50	0.67	1.37	0.89	0.59	1.41	1.20	0.62	1.69	0.75	0.86
MBE_h [%]	-0.03	-0.58	0.02	0.39	-0.42	-0.26	0.62	-0.01	0.35	-0.39	-0.33	0.58
$HHVE_h^1$ [%]	19.93	-51.25	-28.01	122.63	9.79	9.49	-1.99	65.67	-68.63	-68.25	-26.58	-85.29
	3 layers											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.81	0.31	0.89	1.62	1.68	0.67	0.76	1.86	0.66	1.86	0.51	0.81
MBE_h [%]	-0.36	-0.08	-0.16	-0.55	-0.84	-0.23	0.22	-0.86	0.37	-0.65	-0.19	0.54
$HHVE_h^1$ [%]	-30.42	18.12	-39.66	40.74	-12.34	-24.45	-35.04	-4.10	-92.16	-51.79	-48.73	-97.06

Table C.12: Values of MBE_h, $HHVE_h^1$, and $HHVE_h^2$ calculated for the optimised *van Genuchten* and *linear* models with monthly stress factors (calibration period). $HHVE_h^1$ corresponds to $HHVE_h$ for the 1st year of the validation period.

	van Genuchten											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	0.66	0.28	0.40	0.90	0.47	0.54	1.70	1.71	0.45	0.37	0.24	0.75
MBE_h [%]	-0.28	-0.11	-0.13	-0.35	0.03	-0.28	0.76	-0.90	0.24	-0.14	-0.10	0.51
$HHVE_h^1$ [%]	7.34	1.88	-11.65	18.52	34.04	-5.84	-15.95	-14.93	-96.08	-4.17	1.27	-52.94
	linear											
	Pz1	Pz2	Pz3	Pz4	Pz5	Pz6	Pz7	Pz8	Pz9	Pz10	Pz11	Pz12
RMS_h [m]	1.85	2.97	1.92	3.60	2.81	3.26	1.29	2.11	1.10	5.75	2.34	0.79
MBE_h [%]	0.50	-0.94	0.77	-1.33	-1.08	1.55	-0.18	-0.54	0.56	-2.30	-0.95	0.53
$HHVE_h^1$ [%]	92.66	301.88	-4.14	126.75	138.72	70.07	31.34	94.03	254.90	1.79	189.87	-86.76

Bibliography

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. Petrov (Ed.), *Second International Symposium on Information Theory*, pp. 267–281.

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control 19*(6), 716–723.

Anderman, E., M. Hill, and E. Poeter (1996). Two-dimensional advective transport in ground-water flow parameter estimation. *Ground Water 34*(6), 1001–1009.

Andersen, J., G. Dybkjaer, K. Jensen, J. Refsgaard, and K. Rasmussen (2002). Use of remotely sensed precipitation and leaf area index in a distributed hydrological model. *Journal of Hydrology 264*(1-4), 34–50.

Anderson, D. (2003). Multi-model inference based on Kullback-Leibler information. In E. Poeter, C. Zheng, and M. Hill (Eds.), *Proceedings of MODFLOW and More 2003: Understanding Through Modeling*.

Anderson, M. (2005). Heat as a ground water tracer. *Ground Water 43*(6), 951–968.

Aricò, C., C. Nasello, and T. Tucciarelli (2009). Using steady-state water level

data to estimate channel roughness and discharge hydrograph. *Advances in Water Resources 32*(8), 1223–1240.

Asner, G., J. Scurlock, and J. Hicke (2003). Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global Ecology and Biogeography 12*(3), 191–205.

Barthel, R., J. Jagelke, J. Götzinger, T. Gaiser, and A. Printz (2008). Aspects of choosing appropriate concepts for modelling groundwater resources in regional integrated water resources management - Examples from the Neckar (Germany) and Ouémé catchment (Benin). *Physics and Chemistry of the Earth, Parts A/B/C 33*(1-2), 92–114.

Batlle-Aguilar, J. (2008). *Groundwater flow and contaminant transport in an alluvial aquifer: in-situ investigation and modelling of a brownfield with strong groundwater - surface water interactions*. Ph. D. thesis, Université de Liège.

Bauer, S., C. Beyer, and O. Kolditz (2006). Assessing measurement uncertainty of first-order degradation rates in heterogeneous aquifers. *Water Resources Research 42*(W01420).

Beldring, S. (2002). Multi-criteria validation of a precipitation-runoff model. *Journal of Hydrology 257*(1-4), 189–211.

Beven, K. (1989). Changing ideas in hydrology - The case of physically-based models. *Journal of Hydrology 105*(1-2), 157–172.

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources 16*(1), 41–51.

Beven, K. (1996a). *Distributed Hydrological Modeling*, Chapter A discussion of distributed hydrological modeling.

Beven, K. (1996b). *The Scientific Nature of Geomorphology*, Chapter Equifinality and uncertainty in geomorphological modeling.

Beven, K. (2000). On the future of distributed modelling in hydrology. *Hydrological Processes 14*(16-17), 3183–3184.

Beven, K. (2001a). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences 5*(1), 1–12.

Beven, K. (2001b). On explanatory depth and predictive power. *Hydrological Processes 15*(15), 3069–3072.

Beven, K. (2002a). Towards a coherent philosophy for modeling the environment. *Proceedings of the Royal Society of London, Part A 458*, 2465–2484.

Beven, K. (2002b). Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes 16*(2), 189–206.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology 320*(1-2), 18–36.

Beven, K. and J. Freer (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes 6*(3), 279–298.

Beven, K. and J. Freer (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology 249*(1-4), 11–29.

Beyer, C., S. Bauer, and O. Kolditz (2006). Uncertainty assessment of contaminant plume length estimates in heterogeneous aquifers. *Journal of Contaminant Hydrology 87*(1-2), 73–95.

Bravo, H., F. Jiang, and R. Hunt (2002). Using groundwater temperature data to constrain parameter estimation in a groundwater flow model of a wetland system. *Water Resources Research 38*(8).

Brouyère, S., G. Carabin, and A. Dassargues (2004). Climate change impacts on groundwater resources: modelled deficits in a chalky aquifer, Geer basin, Belgium. *Hydrogeology Journal 12*(2), 123–134.

Brouyère, S., J. Gesels, P. Goderniaux, T. Robert, L. Thomas, A. Dassargues, J. Bastien, F. Van Wittenberge, A. Rorive, F. Dossin, J.-L. Lacour, D. Le Madec, P. Nogarède, and V. Hallet (2009). Caractérisation hydrogéologique et support à la mise en oeuvre de la Directive Européenne 2000/60 sur les masses d'eau souterraine en Région Wallonne (Projet Synclin'EAU). Délivrable D.2.22 - partie MESO RWM021, Convention RW et SPGE-Aquapôle.

Brutsaert, W. (2005). *Hydrology: an introduction*. Cambridge, United Kingdom: Cambridge University Press.

Burnham, K. and D. Anderson (2002). *Model selection and multi-model inference: a practical information-theoric approach*. New York, NY, USA: Springer-Verlag.

Burnham, K. and D. Anderson (2004). Multi-model inference: understanding AIC and BIC model selection. *Sociological Methods and Research 33*(2), 261–304.

Burow, K., J. Constantz, and R. Fujii (2005). Heat as a tracer to estimate dissolved organic carbon flux from a restored wetland. *Ground Water 43*(4), 545–556.

Canadell, J., R. Jackson, J. Ehleringer, H. Mooney, O. Sala, and E.-D. Schulze

(1996). Maximum rooting depth of vegetation types at the global scale. *Oecologia 108*(4), 583–595.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. Slooten (2005). Inverse problem in hydrogeology. *Hydrogeology Journal 13*(1), 206–222.

Carrera, J. and S. Neuman (1986a). Estimation of aquifer parameters under transient and steady sate conditions: 1. Maximum likelihood method incorporating prior information. *Water Resources Research 22*(2), 199–210.

Carrera, J. and S. Neuman (1986b). Estimation of aquifer parameters under transient and steady sate conditions: 2. Uniqueness, stability, and solution algorithms. *Water Resources Research 22*(2), 211–227.

Certes, C. and G. de Marsily (1991). Application of the pilot point method to the identification of aquifer transmissivities. *Advances in Water Resources 14*(5), 284–300.

Chiew, F. and T. McMahon (1994). Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments. *Journal of Hydrology 153*(1-4), 383–416.

Christensen, S. and R. Cooley (2005). User guide to the UNC process and three utility programs for computation of nonlinear confidence and prediction intervals using MODFLOW-2000. Techniques and Methods Report 2004-1349, U.S. Geological Survey.

Das, T., A. Bárdossy, E. Zehe, and Y. He (2008). Comparison of conceptual model performance using different representations of spatial variability. *Journal of Hydrology 356*(1-2), 106–118.

Dassargues, A., J.-P. Radu, and R. Charlier (1988). Finite elements modelling

of a large water table aquifer in transient conditions. *Advances in Water Resources 11*(2), 58–66.

de Marsily, G., G. Lavedan, M. Boucher, and G. Fasanino (1984). Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In G. Verly, M. David, A. Journel, and A. Marechal (Eds.), *Geostatistics for Natural Resources Characterization*, pp. 831–849.

Dickinson, R., A. Henderson-Sellers, C. Rosenzweig, and P. Sellers (1991). Evapotranspiration models with canopy resistance for use in climate models, a review. *Agricultural and Forest Meteorology 54*(2-4), 373–388.

Doherty, J. (2003). Ground water model calibration using pilot points and regularization. *Ground Water 41*(2), 170–177.

Doherty, J. (2005). *PEST - Model-independent parameter estimation - User manual - 5th edition*. Watermark Numerical Computing.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B 57*(1), 45–97.

Ebel, B. and K. Loague (2006). Physics-based hydrologic-response simulation: seeing through the fog of equifinality. *Hydrological Processes 20*(13), 2887–2900.

Fazal, M., M. Imaizumi, S. Ishida, T. Kawachi, and T. Tsuchihara (2005). Estimating groundwater recharge using SMAR conceptual model calibrated by genetic algorithm. *Journal of Hydrology 303*(1-4), 56–78.

Fetter, C. (2001). *Applied hydrogeology*. Upper Saddle River, NJ, USA: Prentice Hall.

Foglia, L., M. Hill, W. Mehl, and P. Burlando (2009). Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function. *Water Resources Research 45*(W06427).

Freeze, R. and J. Cherry (1979). *Groundwater*. Upper Saddle River, NJ, USA: Prentice Hall.

Gómez-Hernández, J. (2006). Complexity. *Ground Water 44*(6), 782–785.

Goderniaux, P. (2010). *Impact of climate change on groundwater reserves*. Ph. D. thesis, Université de Liège.

Goderniaux, P., S. Brouyère, H. Fowler, S. Blenkinsop, R. Therrien, P. Orban, and A. Dassargues (2009). Large scale surface-subsurface hydrological model to assess climate change impacts on groundwater reserves. *Journal of Hydrology 373*(1-2), 122–138.

Graham, D. and L. Kilde (2002). MIKE-SHE users guide. Manual, Danish Hydraulic Institute.

Grandjean, M., Y. Hanin, and V. Rousseaux (2006). Occupation du territoire en Région wallonne - TERRIT1 - Urbanisation et occupation du sol. Technical report.

Grayson, R., I. Moore, and T. McMahon (1992). Physically based hydrologic modeling 2. Is the concept realistic? *Water Resources Research 28*(10), 2659–2666.

Gupta, H., H. Kling, K. Yilmaz, and G. Martinez (2009). Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology 377*(1-2), 80–91.

Gupta, H., S. Sorooshian, and P. Yapo (1998). Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resources Research 34*(4), 751–763.

Gupta, H., S. Sorooshian, and P. Yapo (1999). Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. *Journal of Hydrologic Engineering 4*(2), 135–143.

Heng, H. and N. Nikolaidis (1998). Modeling of nonpoint source pollution of nitrogen at the watershed scale. *Journal of the American Water Resources Association 34*(2), 359–374.

Henriksen, H., L. Troldborg, P. Nyegaard, T. Sonnenborg, J. Refsgaard, and M. B. (2003). Methodology for construction, calibration and validation of a national hydrological model for denmark. *Journal of Hydrology 280*(1-4), 52–71.

Hernández, A., S. Neuman, G. A., and J. Carrera (2003). Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion. *Stochastic Environmental Research and Risk Assessment 17*(5), 319–328.

Hill, M. (1992). A computer program (MODFLOWP) for estimating parameters of a transient, three-dimensional, ground-water flow model using nonlinear regression. Open-File Report 91-484, U.S. Geological Survey.

Hill, M. (2006). The practical use of simplicity in developing ground water models. *Ground Water 44*(6), 775–781.

Hill, M., R. Cooley, and D. Pollock (1998). A controlled experiment in groundwater flow model calibration using nonlinear regression. *Ground Water 36*(3), 520–535.

Hill, M. and C. Tiedeman (2007). *Effective groundwater model calibration with analysis of sensitivities, predictions, and uncertainty*. Hoboken, NJ, USA: John Wiley & Sons.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1999). Bayesian model averaging: a tutorial. *Statistical Science 14*(4), 382–417.

Hornberger, M., J. Raffensperger, P. Wilberg, and K. Eshleman (1998). *Elements of physical hydrology*. Baltimore, MD, USA: Johns Hopkins University Press.

Hunt, R., J. Doherty, and M. Tonkin (2007). Are models too simple? Arguments for increased parameterization. *Ground Water 45*(3), 254–262.

Hurvich, C. and C.-L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika 76*(2), 297–307.

Hurvich, C. and C.-L. Tsai (1994). Autoregressive model selection in small samples using a bias-corrected version of AIC. In H. Bozdogan (Ed.), *Engineering and Scientific Applications*.

Islam, M. (2004). *Automated calibration of a physically-based hydrologic model to simulate water balance variables for water and crop management*. Ph. D. thesis, University of California at Davis.

Jones, J. (2005). *Simulating hydrologic systems using a physically-based surface-subsurface model: issues concerning flow, transport and parametrization*. Ph. D. thesis, University of Waterloo.

Kashyap, R. (1982). Optimal choice of AR and MA parts in autoregressive moving average models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 4*(2), 99–104.

Kass, R. and A. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 190*(430), 773–798.

Keating, E., J. Doherty, J. Vrugt, and Q. Kang (2010). Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality. *Water Resources Research 46*(W10517).

Kowalsky, M., S. Finsterle, and Y. Rubin (2004). Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone. *Advances in Water Resources 27*(6), 583–599.

Kristensen, K. and S. Jensen (1975). A model for estimating actual evapotranspiration from potential evapotranspiration. *Nordic Hydrology 6*(3), 170–188.

Kullback, S. and R. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics 22*(1), 79–86.

LaVenue, A., B. RamaRao, G. de Marsily, and M. Marietta (1995). Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields 2. Application. *Water Resources Research 31*(3), 495–516.

Legates, D. and G. McCabe (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimate model validation. *Water Resources Research 35*(1), 233–241.

Lehmann, P., S. Assouline, and D. Or (2008). Characteristic lengths affecting evaporative drying of porous media. *Physical Review E 77*(056309).

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quaterly of Applied Mathematics 2*, 164–168.

Li, Q., A. Unger, E. Sudicky, D. Kassenaar, E. Wexler, and S. Shikaze (2008). Simulating the multi-seasonal response of a large-scale watershed with a 3D physically-based hydrologic model. *Journal of Hydrology 357*(3-4), 317–336.

Liu, C. and W. Ball (1999). Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware. *Water Resources Research 32*(5), 1131–1161.

Loague, K. and R. Green (1991). Statistical and graphical methods for evaluating solute transport modesl: overview and applications. *Journal of Contaminant Hydrology 7*(1-2), 51–73.

Loague, K., C. Heppner, B. Mirus, B. Ebel, Q. Ran, A. Carr, S. BeVille, and J. VanderKwaak (2006). Physics-based hydrologic-response simulation: foundation for hydroecology and hydrogeomorphology. *Hydrological Processes 20*(5), 1231–1237.

Loague, K. and J. VanderKwaak (2004). Physics-based hydrologic response simulation: platinum bridge, 1958 Edsel, or useful tool. *Hydrological Processes 18*(15), 2949–2956.

Manning, C. and S. Ingebritsen (1999). Permeability of the continental crust: the implications of geothermal data and metamorphic systems. *Reviews of Geophysics 37*(1-2), 127–150.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinears parameters. *Journal of the Society of Industrial and Applied Mathematics 11*(2), 431–441.

McCuen, R. (1989). *Hydrologic analysis and design*. Englewood Cliffs, NJ, USA: Prentice Hall.

Mishra, S., N. Deeds, and G. Ruskauff (2009). Global sensitivity analysis techniques for probabilistic ground water modeling. *Ground Water 47*(5), 727–744.

Moore, C. and J. Doherty (2006). The cost of uniqueness in groundwater model calibration. *Advances in Water Resources 29*(4), 605–623.

Moore, C., T. Wöhling, and J. Doherty (2010). Efficient regularization and uncertainty analysis using a global optimization methodology. *Water Resources Research 46*(W08527).

Moriasi, D., J. Arnold, M. Van Liew, R. Bingner, R. Harmel, and T. Veith (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE 50*(3), 885–900.

Murphy, A. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review 116*, 2417–2424.

Nash, J. and J. Sutcliffe (1970). River flow forecasting through conceptual models I — A discussion of principles. *Journal of Hydrology 10*(3), 282–290.

Neuman, S. (2003). Maximum Likelihood Bayesian Averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment 17*(5), 291–305.

Neuman, S. and P. Wierenga (2003). A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites. NUREG/CR-6940 (PNNL-16396), U.S. Nuclear Regulatory Commission.

Niswonger, R., D. Prudic, G. Pohll, and J. Constantz (2005). Incorporating seepage losses into the unsteady streamflow equations for simulat-

ing intermittent flow along mountain front streams. *Water Resources Research 41*(W06006).

Orban, P., S. Brouyère, J. Batlle-Aguilar, J. Couturier, P. Goderniaux, M. Leroy, P. Maloszewski, and A. Dassargues (2010). Regional transport modelling for nitrate trend assessment and forecasting in a chalk aquifer. *Journal of Contaminant Hydrology 118*(1-2), 79–93.

Oreskes, N. (2000). *The Earth Around Us*, pp. 70–82. New York, NY, USA: Freedman, W.H.

Panday, S. and P. Huyakorn (2004). A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. *Advances in Water Resources 27*(4), 361–382.

Poeter, E. and D. Anderson (2005). Multimodel ranking and inference in ground water modeling. *Ground Water 46*(4), 597–605.

Poeter, E. and M. Hill (1997). Inverse models: a necessary next step in ground-water modeling. *Ground Water 35*(2), 250–260.

Poeter, E. and M. Hill (2007). MMA: a computer code for multi-model analysis. Techniques and Methods 6-E3, U.S. Geological Survey.

Poeter, E., M. Hill, E. Banta, and S. Mehl (2005). UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation. Techniques and Methods 6-A11, U.S. Geological Survey.

Poeter, E. and S. McKenna (1995). Reducing uncertainty associated with ground-water flow and transport predictions. *Ground Water 33*(6), 899–904.

Radcliffe, A. (2000). *Physical hydrogeology and the impact of urbanization at the Waterloo West Side: a groundwater modeling approach*. Ph. D. thesis, University of Waterloo.

RamaRao, B., A. LaVenue, G. de Marsily, and M. Marietta (1995). Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields 1. theory and computational experiments. *Water Resources Research 31*(3), 475–493.

Ramos da Silva, M., C. Schroeder, and J.-C. Verbrugge (2008). Unsaturated rock mechanics applied to a low-porosity shale. *Engineering Geology 97*(1-2), 42–52.

Refsgaard, J. (1997). Parameterisation, calibrationnext term and validation of distributed hydrological models. *Journal of Hydrology 198*(1-4), 69–97.

Refsgaard, J. and H. Henriksen (2004). Modelling guidelines—terminology and guiding principles. *Advances in Water Resources 27*(1), 71–82.

Refsgaard, J., B. Storm, and M. Abbott (1996). *Distributed Hydrological Modeling*, pp. 279–287. The Netherlands: Kluwer.

Refsgaard, J., J. van der Sluijs, J. Brown, and P. van der Keur (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources 29*(11), 1586–1597.

Refsgaard, J., J. van der Sluijs, A. Højberg, and P. Vanrolleghem (2007). Uncertainty in the environmental modelling process - A framework and guidance. *Environmental Modelling & Software 22*, 1543–1556.

Remy, N., A. Boucher, and J. Wu (2009). *Applied geostatistics with SGeMS: a user's guide*. Cambridge, United Kingdom: Cambridge University Press.

Rojas, R., O. Batelaan, L. Feyen, and A. Dassargues (2009). Assessment of model conceptual uncertainty for the regional aquifer pampa del tamarugal - north chile. *Hydrology and Earth System Sciences Discussions 6*(5), 5881–5935.

Rojas, R., L. Feyen, and A. Dassargues (2008). Conceptual model uncertainty in groundwater modelling: combining generalized likelihood uncertainty estimations and bayesian model averaging. *Water Resources Research 44*(W124118).

Roulier, S., N. Baran, C. Mouvet, F. Stenemo, X. Morvan, H.-J. Albrechtsen, L. Clausen, and N. Jarvis (2006). Controls on atrazine leaching through a soil-unsaturated fractured limestone sequence at Brévilles, france. *Journal of Contaminant Hydrology 84*(1-2), 81–105.

Saiers, J., D. Genereux, and C. Bolster (2004). Influence of calibration methodology on ground water flow predictions. *Ground Water 42*(1), 32–44.

Saltelli, A., K. Chan, and M. Scott (2000). *Sensitivity analysis*. Chichester, United Kingdom: John Wiley.

Sanford, W., L. Plummer, D. McAda, L. Bexfield, and S. Anderholm (2004a). Hydrochemical tracers in the middle Rio Grande Basin, USA: 2. Calibration of a groundwater-flow model. *Hydrogeology Journal 12*(4), 389–407.

Sanford, W., L. Plummer, D. McAda, L. Bexfield, and S. Anderholm (2004b). Use of environmental tracers to estimate parameters for a predevelopment ground-water flow model of the middle Rio Grande Basin, New Mexico. Water-Resources Investigation Report 03-4286, U.S. Geological Survey.

Schäfer, D., B. Schlenz, and A. Dahmke (2004). Evaluation of exploration and

monitoring methods for verification of natural attenuation using the virtual aquifer approach. *Biodegradation 15*(6), 453–465.

Schroeder, P., N. Aziz, C. Lloyd, and P. Zappi (1994). The Hydrologic Evaluation of Landfill Performance (HELP) model: user's guide for version 3. EPA/600/R-94/168a, US EPA.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*(2), 461–464.

Singh, A., S. Mishra, and G. Ruskauff (2010). Model averaging techniques for quantifying conceptual model uncertainty. *Ground Water 48*(5), 701–715.

Skaggs, T. and Z. Kabala (1994). Recovering the release history of groundwater contaminant. *Water Resources Research 30*(1), 71–79.

Skahill, B. and J. Doherty (2006). Efficient accommodation of local minima in watershed model calibration. *Journal of Hydrology 329*(1-2), 122–139.

Smith, P., K. Beven, and J. Tawn (2008). Informal likelihood measures in model assessment: theoretic development and investigation. *Advances in Water Resources 31*(8), 1087–1100.

Smith, R., D. Goodrich, D. Woolhiser, and J. Simanton (1994). Comment on "Physically based hydrologic modeling, 2. Is the concept realistic?". *Water Resources Research 30*(3), 851–854.

Solomatine, D., Y. Dibike, and N. Kukuric (1999). Automatic calibration of groundwater models using global optimization techniques. *Hydrological Sciences Journal 44*(6), 879–894.

Spanoudaki, K., A. Stamou, and A. Nanou-Giannarou (2009). Development

and verification of a 3-D integrated surface water-groundwater model. *Journal of Hydrology 375*(3-4), 410–427.

Sugira, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods 7*(1), 13–26.

Therrien, R., E. Sudicky, R. McLaren, and S. Panday (2005). HydroGeoSphere - A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Manual, Groundwater Simulations Group.

Tikhonov, A. and V. Arsenin (1977). *Solutions of ill-posed problems*. Hoboken, NJ, USA: John Wiley & Sons.

Tonkin, M. and J. Doherty (2005). A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research 41*(W10414).

Tsai, F.-C., N.-Z. Sun, and W.-G. Yeh (2003). Global-local optimization for parameter structure identification in three-dimensional groundwater modeling. *Water Resources Research 39*(1043).

van den Doel, K. and U. Ascher (2006). On level set of regularization for highly ill-posed distributed parameter estimation problems. *Journal of Computational Physics 216*(2), 707–723.

van Genuchten, M. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal 44*(5), 892–898.

Vecchia, A. and R. Cooley (1987). Simultaneous confidence and prediction

intervals for nonlinear regression models with application to a groundwater flow model. *Water Resources Research 22*(2), 95–108.

Vesselinov, V., S. Neuman, and W. Illman (2001). Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff 2. equivalent parameters, high-resolution stochastic imaging and scale effects. *Water Resources Research 37*(12), 3019–3041.

Vrugt, J., H. Gupta, L. Bastidas, W. Bouten, and S. Sorooshian (2003). Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research 39*(1214).

Vázquez, R. and J. Feyen (2003). Effect of potential evapotranspiration estimates on effective parameters and performance of the MIKE SHE-code applied to a medium-size catchment. *Journal of Hydrology 270*(3-4), 309–327.

Vázquez, R., L. Feyen, J. Feyen, and J. Refsgaard (2002). Effect of grid size on effective parameters and model performance of the MIKE-SHE code. *Hydrologic Processes 16*(2), 355–372.

Weglarczyk, S. (1998). The interdependence and applicability of some statistical quality measures for hydrological models. *Journal of Hydrology 206*(1-2), 98–103.

Wood, T., C. Helm-Clark, H. Huang, S. Magnuson, T. McLing, B. Orr, M. Roddy, M. Rohe, M. Plummer, and R. Podgorney (2005). Operable unit 10-08. Summary report of the subregional-scale two-dimensional aquifer model. ICP/EXT-05-00979, U.S. Departement of Energy.

Woodbury, A. and L. Smith (1988). Simultaneous inversion of hydrogeologic and thermal data: 2. Incorporation of thermal data. *Water Resources Research 24*(3), 356–372.

Wylie, A. and J. Doherty (2005). Eastern Snake Plain Aquifer model calibration report. DDM-008, University of Idaho.

Yapo, P., H. Gupta, and S. Sorooshian (1996). Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology 181*(1-4), 23–48.

Ye, M., P. Meyer, and S. Neuman (2008). On model selection criteria in multimodel analysis. *Water Resources Research 44*(W03428).

Ye, M., S. Neuman, and P. Meyer (2004). Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff. *Water Resources Research 40*(W05113).

Ye, M., K. Pohlmann, J. Chapman, G. Pohll, and D. Reeves (2010). A model-averaging method for assessing groundwater conceptual model uncertainty. *Ground Water 48*(5), 716–728.

Yeh, W.-G. (1986). Review of parameter identification procedures in groundwater hydrology: the inverse problem. *Water Resources Research 22*(2), 95–108.

Yeh, W.-G. and Y. Yoon (1981). Aquifer parameter identification with optimum dimension in parameterization. *Water Resources Research 17*(3), 664–672.

Zheng, C. and P. Wang (1996). Parameter structure identification using tabu search and simulated annealing. *Advances in Water Resources 19*(4), 215–224.