A highly parameterised groundwater modelling study: Insights into recharge estimability in the face of non-uniqueness and the impacts of climate versus pumping

Matthew James Knowling

BSc EnvSc (Hons)

Thesis submitted for the degree of

Doctor of Philosophy in the School of Environment,

Faculty of Science and Engineering,

Flinders University,

South Australia

February, 2016

Table of contents

Summary

Groundwater models are used routinely for water resources management and environmental decision-making support. The recent advances in computing power and technology have led to the development of state-of-the-art modelling methodologies that are capable of supporting the application of complex and highly parameterised models to real-world field-scale settings. Despite this, the use of highly parameterised models to address practical hydrology questions is largely lacking. The objectives of this thesis are to: (1) demonstrate the application of a highly parameterised modelling approach for disentangling climate and human impacts for a regional setting, (2) evaluate the estimability of recharge and its spatial variability through calibration of field-scale steady-state groundwater models, and (3) assess the extent to which time-varying recharge can be informed through field-scale transient model calibration.

First, this thesis presents the application of a highly parameterised modelling strategy to quantify climate and pumping contributions to aquifer depletion for a regional setting (Uley South Basin, USB; southern Australia). The strategy involves calibration-constrained model predictions of natural and pumped conditions. Results show that, while both climate and pumping impacts are highly variable in both space and time, the impact of pumping is, in general terms, greater than that of climate (by 2.7 times on the basis of time-averaged impacts). The results serve as a response to a recent Parliamentary Enquiry into the cause of USB groundwater-level decline.

Second, this thesis investigates the extent to which recharge and its spatial variability can be informed via the calibration of field-scale steady-state groundwater models. Recharge estimation by these means is known to be hampered by the non-uniqueness between recharge and aquifer parameter (e.g., hydraulic conductivity; K) values. Here, a systematic analysis of calibration-based recharge estimates is undertaken subject to varying degrees of hydraulic parameter information. Results show that, for a synthetic reality based on a highly parameterised model of USB, a surprisingly large amount of K information (>100 K preferred values) is required to obtain reasonable recharge estimates (<10% average error). The use of pumping data reduces error in both average and spatially variable recharge estimates, whereas submarine groundwater discharge (as a calibration target) reduces average recharge error only. This study suggests that the estimation of recharge through calibration may be impractical for real-world settings.

Third, this thesis evaluates time-varying recharge estimability via calibration of transient groundwater models. Transient model calibration requires the additional consideration of aquifer storage parameters (e.g., specific yield; S_y). The analysis undertaken here is similar to that of the second study, i.e., recharge estimates, subject to varying degrees of aquifer parameter information and water-level data, are investigated. Results show that reasonable estimates of monthly recharge (<30% recharge root-mean-squared error) require a large amount of transient water-level data, and that the spatial distribution of K is known (i.e., through joint recharge-and-S_y estimation). Joint estimation of recharge, S_y and K, however, does not yield reasonable recharge values. This study indicates that the estimation of

recharge through calibration for real-world settings may require an impractical amount of water-level and hydraulic parameter data.

Dedication

This thesis is dedicated to my mother, who remains a constant source of inspiration.

Declaration

I certify that this thesis does not incorporate, without acknowledgment, any material previously submitted for a degree or diploma in any other university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

.....

Matthew J. Knowling

Co-authorship

Matthew Knowling is the primary author on all chapters in this thesis. The co-authors listed on various chapters provided intellectual supervision and editorial support.

Acknowledgements

Firstly, I wish to thank my supervisor, Adrian Werner, for his guidance, encouragement and enthusiasm over the last four years. I admire your commitment to addressing important and meaningful research questions. I also wish to thank my co-author Daan Herckenrath for sharing some of his inverse modelling wisdom. I would like to thank Ty Watson for many helpful discussions regarding optimisation and uncertainty analysis, and for programming assistance. Thanks also go to Etienne Bresciani for fruitful discussions on the topic of recharge estimation through inverse modelling, and Dan Partington for coding assistance.

I am very grateful for the financial support of an Australian Postgraduate Award, the Goyder Institute for Water Research, and the National Centre for Groundwater Research and Training (NCGRT). I especially wish to acknowledge Craig Simmons and the NCGRT for providing me with opportunities to present at a number of international conferences, and the Goyder Institute for facilitating interaction with industry stakeholders.

I would also like to thank all of the great people at Flinders University and the NCGRT. A special thank-you goes to Megan Sebben, my PhD compadré.

Lastly, I wish to sincerely thank my family and friends, and my partner, Maddi, for their unwavering love and support.

List of Figures

List of Tables

Chapter 1

1. Introduction and objectives

Numerical groundwater models are commonly used for the purposes of devising effective water management strategies, and for improving conceptual understanding of hydrological systems. The former usually involves the assessment of model predictions subject to a range of potential future climate and management scenarios (e.g., Scibek and Allen, 2006; Candela et al., 2009; Ferguson and Maxwell, 2012). The latter is often achieved by means of inverse modelling, whereby model parameters (and/or stresses, boundary conditions, etc.) are estimated through the process of calibration or history matching (i.e., the minimisation of the discrepancy between field measurements and model-generated outputs) (e.g., Sanford et al., 2004; Dausman et al., 2010a; Maneta and Wallender, 2013).

Following the recent advances in computing power and technology, there has been a surge in the development of modelling methodologies and software. Computationally efficient inverse modelling approaches constitute many of these methodologies, such as model calibration and uncertainty analysis techniques that can be used in conjunction with complex field-scale groundwater models containing hundreds or even thousands of parameters (herein referred to as "highly parameterised models") (e.g., Vrugt et al., 2003; Tonkin and Doherty, 2009; Burrows and Doherty, 2014). Highly parameterised models provide a basis for: (1) enhanced extraction of information from observation data, and (2) comprehensive evaluation of parameter

and predictive uncertainty (Hunt et al., 2007). Prior to the development of these methodologies, the computational burden associated with the calibration of highly parameterised models, and the quantification of the uncertainty associated with the predictions based on these models (both of which require a very large number of model runs), precluded the application of highly parameterised models for real-world regional-scale aquifer settings.

Despite that the application of highly parameterised models for regional settings is supported by recently developed modelling methodologies, their use in addressing practical hydrology problems are largely lacking at present. This thesis investigates two such research problems:

First, the quantification of climate and human impacts on hydrological systems is investigated. This has been the subject of many studies over the last two decades (e.g., Ye et al., 2003; von Asmuth et al., 2008; Joodaki et al., 2014). These studies are critical to the development of effective water management strategies. Previous modelling investigations of climate and human impacts adopt simple (i.e., lumped-parameter, one-dimensional) models (e.g., Heuvelmans et al., 2011; van Loon and van Lanen, 2013). While simple models constitute rapid assessment tools for informing environmental decision-making, they preclude the representation of complexities that are inherent in field-scale groundwater systems such as aquifer heterogeneity, and spatially and temporally variable boundary conditions and stresses, all of which can be accounted for in spatially distributed, physically based models. There is therefore need to extend previous modelling studies that disentangle

climate and human impacts on groundwater systems by applying a highly parameterised and spatially distributed modelling strategy.

Second, the estimation of aquifer recharge, which has been another subject of much hydrology research over the last two decades (e.g., Wood and Sanford, 1995; Crosbie et al., 2005; Wang et al., 2008), is investigated. Despite that reliable estimates of recharge and its temporal and spatial variability are often prerequisites for effective groundwater modelling and management, recharge is widely considered to be one of the most difficult water balance components to quantify given that it cannot be measured directly, and therefore must be inferred from indirect approaches (Scanlon et al., 2002). The estimation of recharge through groundwater model calibration is one such increasingly popular approach (e.g., Essaid et al., 2003; Liu et al., 2008; Hashemi et al., 2013). Recharge estimation by these means is hampered by the non-uniqueness between estimated values of recharge and aquifer hydraulic properties such as hydraulic conductivity (K) and specific yield (S_y) (Sanford, 2002). For example, a unique solution for recharge and K does not exist; only their ratio can be estimated uniquely. Previous studies that jointly (i.e., simultaneously) estimate recharge and hydraulic parameters have not evaluated the ramification of non-uniqueness in terms of recharge estimability for real-world, regional-scale aquifer settings. Therefore, the extent to which recharge and its variability in time and space can be informed through groundwater model calibration for practical situations is presently unknown.

The broad objective of this thesis is to provide insights into these practical and important hydrology research problems through the use of highly parameterised modelling strategies. More specifically, the primary objectives of the thesis are to:

1. Demonstrate the application of a highly parameterised modelling strategy to quantify the relative contributions of climate variability and groundwater pumping to aquifer depletion within a regional-scale setting (addressed in Chapter 2). The Uley South Basin (USB; southern Australia), where there is conjecture around the causal factors of aquifer depletion (NRC, 2013), is used as a case study.
2. Evaluate the estimability of recharge and its spatial variability through calibration of steady-state field-scale groundwater models (addressed in Chapter 3). To address this objective requires the investigation of the effect of non-uniqueness between estimates of spatially distributed recharge and K.
3. Evaluate the estimability of time-varying recharge through transient field-scale groundwater model calibration (addressed in Chapter 4). To address this objective requires the examination of the non-uniqueness between estimates of transient recharge and the spatially distribution of both K and S_y.

This thesis contains three distinct bodies of work, presented as manuscripts (Chapters 2, 3 and 4), the first two of which are published in *Journal of Hydrology*, and the third of which is currently under review at *Journal of Hydrology*. References to these manuscripts are given at the beginning of each chapter. The conclusions of the thesis are summarised in Chapter 5.

Chapter 2

2. Quantifying climate and pumping contributions to aquifer depletion using a highly parameterised groundwater model: Uley South Basin (South Australia)

This chapter is based on the following paper:

Knowling, M.J., Werner, A.D., Herckenrath, D., 2015. Quantifying climate and pumping contributions to aquifer depletion using a highly parameterised groundwater model: Uley South Basin (South Australia). Journal of Hydrology 523, 515-530, doi: 10.1016/j.jhydrol.2015.01.081.

2.1. Abstract

The relative contributions of climate and human stresses to aquifer depletion in real-world settings are rarely quantified, particularly where complex patterns of depletion arise from the spatial and temporal variability in aquifer stresses. These impacts can be assessed using calibration-constrained model predictions of disturbed (i.e., subject to human activity) and undisturbed (i.e., natural) conditions. Prior investigations that adopt this approach employ lumped-parameter or one-dimensional models. Here, we extend previous studies by using a highly parameterised, spatially distributed

groundwater model to investigate the relative impacts of climate variability and pumping on aquifer depletion. The Uley South Basin (USB), South Australia, where there is conjecture surrounding the cause of declining groundwater levels, serves as a case study. The relative contributions of climate variability and pumping to USB depletion are shown to be highly variable in time and space. Temporal trends reflect variability in rainfall and pumping, as expected. Spatial trends are primarily dependent on the proximity to both the coastal boundary and pumping wells, and to the distribution of recharge and hydraulic properties. Results show that pumping impacts exceed those of climate between 1978 and 2012, and over the majority of the spatial extent of USB. The contribution of pumping to aquifer depletion is shown to be 2.9 and 1.4 times that of climate in terms of the time-averaged and maximum-in-time basin-scale water budget, respectively. Confidence in model predictions is enhanced by the outcomes of a linear predictive uncertainty analysis, which indicates that predictive uncertainty is lower than climatic and pumping impacts. This study demonstrates the application of a relatively simple analysis that can be used in combination with highly parameterised, spatially distributed groundwater models to differentiate causal factors of aquifer depletion.

2.2. Introduction

Understanding the relative contributions of climatic and human impacts on hydrological systems is essential for developing effective water resources management strategies. The majority of previous attempts to distinguish between climatic and human impacts on hydrological systems have focused on surface water processes, and in particular, stream flow responses (e.g., Ma et al., 2008; Lorenzo-

Lacruz et al., 2010; Wang and Hejazi, 2011), as reviewed recently by Wang (2014). However, the complex nature of groundwater systems often precludes accurate assessment of the contributions of different causal factors to aquifer depletion, which is becoming an increasingly widespread issue (Wada et al., 2010; Werner et al., 2013). Such complexities occur in the form of spatial variability in hydraulic properties, intermittency in natural and human stresses, poorly constrained recharge, pumping and boundary condition controls, and difficulties in accurately determining surface water-groundwater interactions (e.g., Custodio, 2002; Skøien et al., 2003; Panda et al., 2007). As a result, quantification of the extent to which groundwater storage behaviour is impacted by various climate- and human-based controlling factors remains a challenge (van Loon and van Lanen, 2013).

The influence of climate variability and human activity on groundwater systems has previously been investigated using a range of approaches, including statistical- and concept-based time-series analysis (e.g., Shamsudduha et al., 2009; von Asmuth et al., 2008), artificial neural networks (e.g., Ghose et al., 2010), satellite-based gravity observations (e.g., Joodaki et al., 2014), and numerical model applications (e.g., Heuvelmans et al., 2011). Numerical modelling approaches are increasingly applied in this context because they allow for: (1) the representation of system nonlinearities (e.g., watertable response to climate signals), and (2) generation of time series of state variables under natural conditions (e.g., groundwater levels not impacted by pumping) (van Loon and van Lanen, 2013). Model predictions of natural (i.e., in the absence of human activity) and human-influenced (or "disturbed") conditions are used to investigate climate and human effects in a relative manner. Differences between simulated natural and disturbed conditions are assumed to reflect the

consequences of human activity, whereas climatic impacts are inferred directly from the variability caused by natural stresses (e.g., Ruud et al., 2004). This method requires the availability of: (1) well-constrained climate and anthropogenic stress data, and (2) historical field observations from both undisturbed and disturbed periods to provide confidence in the ability of the model, enhanced through calibration, to replicate hydrological system responses (e.g., Cong et al., 2009; Yan et al., 2013; van Loon and van Lanen, 2013).

Prior model-based investigations of the causal factors of groundwater decline employ either lumped-parameter or one-dimensional models. For example, van Loon and van Lanen (2013) applied a lumped-parameter rainfall-runoff model (HBV; Seibert, 2005) to the Upper-Guadiana catchment (Spain), and used an anomaly analysis to demonstrate that the influence of pumping on groundwater levels was, on average, four times higher than that of climate variability. Heuvelmans et al. (2011) used a one-dimensional unsaturated zone model (SWAP; Kroes et al., 2000) to investigate climatic and anthropogenic effects on phreatic groundwater level trends from 245 observation wells in a catchment in northern Belgium. They emphasised the importance of watertable depth and the time series duration in determining the controls on groundwater level trends.

Simple groundwater models (e.g., lumped-parameter models) provide rapid insight for management decision making by avoiding the large computational times and data requirements associated with more complex models (Refsgaard, 1997). However, field-scale groundwater problems require consideration of the spatial and temporal variability in groundwater levels, fluxes, aquifer properties and boundary conditions

that invariably occur. Spatially distributed, physically based groundwater models provide a basis for representing many commonly encountered field-scale complexities (Cuthbert, 2014). Given recent advances in computing power, these models, and highly parameterised versions thereof, have been applied in regional groundwater contexts (e.g., Fienen et al., 2010; Dausman et al., 2010a). Highly parameterised models allow for: (1) enhanced extraction of information from observation data, and (2) comprehensive evaluation of predictive uncertainty (Hunt et al., 2007). However, these have not been adopted in previous efforts to distinguish between climate- and human-induced groundwater impacts.

The primary objective of this study is to extend previous model-based strategies for distinguishing between climatic and human impacts by using a highly parameterised, calibration-constrained groundwater model within a regional setting. A critical and systematic evaluation of the model is undertaken to offer new insights into the performance of such models for assessing causal factors of hydrological impacts. The analysis is applied to the investigation of the Uley South Basin (USB), southern Eyre Peninsula (South Australia). The sustainability of this resource is of concern given significant water level decline between 1970 and 2005 (Werner et al., 2011). Groundwater hydrographs have since stabilised at low levels relative to historical conditions. The causes of groundwater trends are the subject of ongoing debate, as indicated by a recent parliamentary inquiry (NRC, 2013) into Eyre Peninsula water management, which reports that "the cause of the decline of water quantity and quality in the limestone basins cannot be clearly attributed to either natural causes or over-extraction". Moreover, the USB provides an ideal setting for the application of a highly parameterised, spatially distributed groundwater model given the

availability of relatively widespread and long-term water-level monitoring data, comprehensive model- and field-based recharge estimates obtained through a concurrent study (Ordens, 2014), relatively well-constrained groundwater extraction information, and a lack of persistent surface water systems and catchment runoff more generally (Werner et al., 2011).

2.3. Study area

USB is a topographically enclosed surface drainage basin of 129 km^2 bounded by coastal cliffs of up to 140 m AHD (Australian Height Datum, approximately mean sea level) and inland reliefs of between 30 and 180 m AHD (Figure 2.1). The region has a temperate climate characterised by winter-dominant rainfall (May-October), and hot, dry summer months (November-April) (Harrington et al., 2006). Average annual rainfall and pan evaporation rates are 560 and 1547 mm/y, respectively (Bureau of Meteorology, 2010). The land surface is composed predominantly of exposed calcrete or skeletal soils of sandy and clayey loam (Evans, 1997). Solution features (e.g., sinkholes) are widespread across USB, and serve as a mechanism for rapid groundwater recharge (Ordens et al., 2012). USB's vegetation consists primarily of Mallee scrub, drooping she-oak and significant areas of sparse grassland (Ordens et al., 2012).

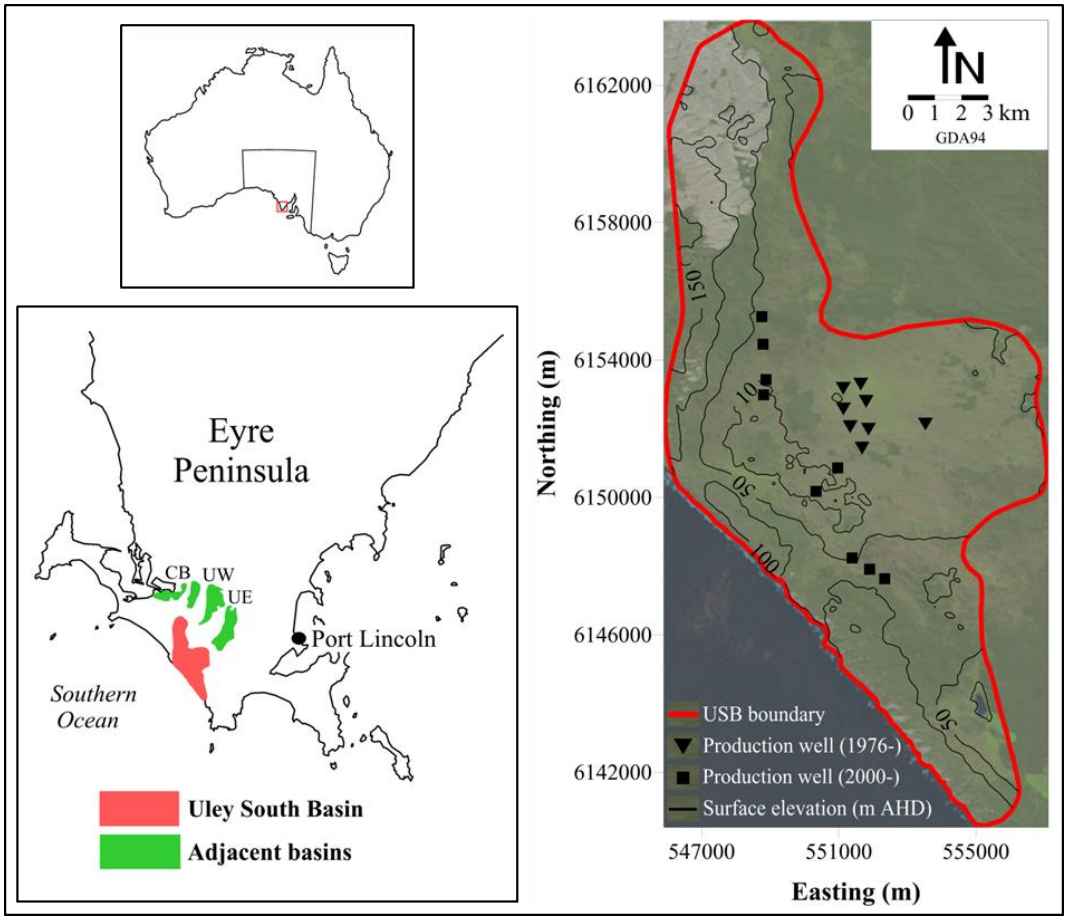

Figure 2.1. Locality map of Uley South Basin (Eyre Peninsula, South Australia). The location of adjacent groundwater basins Uley East, Uley Wanilla and Coffin Bay is given by "UE", "UW" and "CB", respectively.

Groundwater in USB occurs predominantly within an unconfined Quaternary sand and limestone aquifer (QL) (Evans, 1997). The QL is underlain by a discontinuous Tertiary clay aquitard (TC) and a semi-confined Tertiary aquifer comprising silty and clayey sand (TS) (Harrington et al., 2006). These sediments overlay an Archaean metamorphic basement (Harrington et al., 2006).

The USB constitutes the primary freshwater supply for urban, agricultural and industrial activity in the Eyre Peninsula (Harrington et al., 2006). Groundwater pumping began in 1976 (Barnett, 1978), and occurs solely from the QL aquifer. The

production well field expanded from 8 to 17 wells (Figure 2.1) in 2000 (Clarke et al., 2003). Traditionally, USB groundwater management has been undertaken in accordance with a flux-based approach (Werner et al., 2011), whereby recharge estimates have been used to allocate extraction volumes (EPNRM, 2006). Specifically, 60% of the estimated recharge volume is reserved for groundwater-dependent ecosystems; the remainder is deemed available for extraction (DFW, 2012).

The hydrostratigraphic model for USB, illustrated in Figure 2.2, is based on drill logs (DEWNR, 2013), cliff-face observations (Bestland, 2010) and downhole and airborne geophysical surveys (Fitzpatrick et al., 2009). Of particular significance is the airborne electromagnetic data (Fitzpatrick et al., 2009), which suggest the presence of a north-south trending basement ridge, which divides the USB from aquifers to the west.

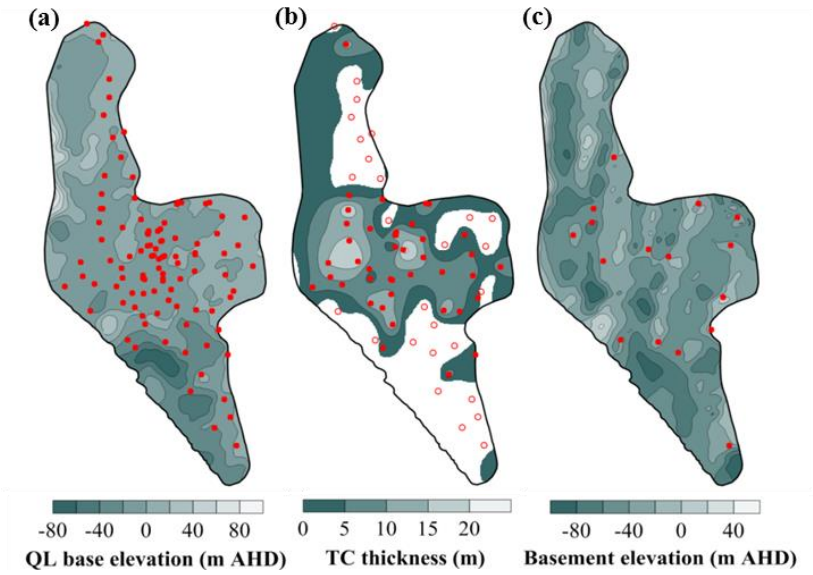

Figure 2.2. Hydrostratigraphy of USB, as shown by (a) QL aquifer base elevation, (b) TC aquitard thickness and (c) basement elevation (which corresponds to the TS aquifer base). Closed markers depict borehole locations that intersect a layer's base, as seen in (a) and (c), or span its entire thickness, as seen in (b). Open markers in (b) show boreholes where TC is not encountered.

The USB boundary is defined following previous approaches as the area of saturated QL material (e.g., Harrington et al., 2006; EPNRM, 2009; DFW, 2012). The extent of the TS aquifer is considerably greater than the region of saturated QL material, and hence we represent the continuation of the TS aquifer using boundary conditions.

Groundwater level data (DEWNR, 2013) from 101 observation wells provide the basis for interpreting the hydrology of the system. The majority (approximately two-thirds) of groundwater level observations pertain to the QL aquifer, although a considerable portion of observation wells are assumed to monitor both aquifers, particularly prior to the improvement in construction of many wells in the 1990s. The differences between groundwater levels in the two aquifers appear to be small, suggesting significant hydraulic connection, albeit there are few nested-well sites for obtaining vertical head gradients. Groundwater level data are of varying quality, and

some observations may not be representative of regional groundwater conditions. For example, there is uncertainty regarding which aquifer many of the wells are monitoring due to inadequate well-screen information, and in north-west USB, there appears to be a local perched watertable, although there is insufficient information to confirm this.

Temporal groundwater behaviour is assessed on the basis of 24,507 groundwater level observations. Four characteristic USB hydrographs and spatially interpolated groundwater level distributions are given in Figure 2.3. Several distinct periods of groundwater level behaviour are apparent, namely: (1) a pre-development period (1960–1976) prior to major groundwater extraction, (2) a period of general groundwater level decline (1976–2005), and (3) a period of groundwater level stabilisation (2005–2009; referred to hereafter as "contemporary conditions"). Since 2009, groundwater levels have either remained stable or trended upwards.

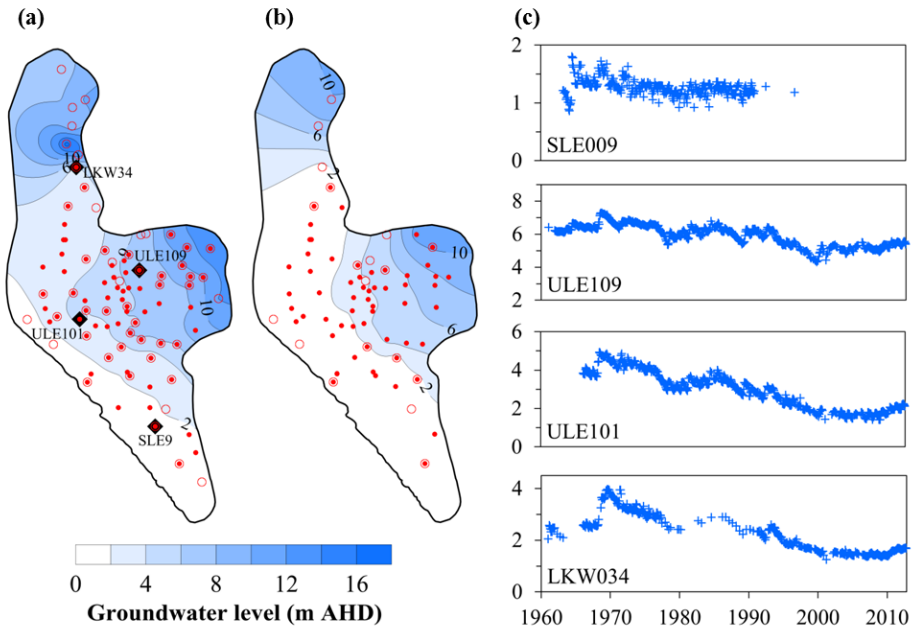

Figure 2.3. Spatial and temporal groundwater level distributions. (Interpolated) spatial groundwater levels (within the QL aquifer) are shown for (a) pre-development and (b)

contemporary steady-state conditions. QL and TS groundwater level observations are distinguished using closed and open red markers, respectively. (c) Groundwater level hydrographs (in m AHD) are shown for observation wells SLE009, ULE109, ULE101 and LKW034 (see diamond markers on (a) for well locations).

Regional groundwater flow in USB is oriented predominantly towards the ocean (Figure 2.3). In the north-west portion of USB, hydraulic gradients suggest complex flow patterns, involving north-westerly flow towards Coffin Bay (Figure 2.1), albeit there are few groundwater level observations in this region. Hydraulic heads at the coastal boundary are determined using the corrections suggested by Morgan et al. (2012). At the shoreline, QL density-corrected heads range from 0.017-0.668 m AHD, and TS density-corrected heads range from 0.020-1.441 m AHD. Tidal effects are not considered.

USB interacts with three adjacent groundwater basins: Uley East, Uley Wanilla and Coffin Bay (Figure 2.1). Inflow to USB occurs from Uley East and Uley Wanilla via the TS aquifer (Evans, 1997; Harrington et al., 2006). USB is expected to discharge towards the western portion of Coffin Bay through QL sediments located north of the basement ridge (Figure 2.2).

For the QL aquifer, eleven horizontal hydraulic conductivity (K_h) estimates have been obtained from pump tests in the central portion of USB that vary between 150 and 1320 m/d (Table 2.1; Sibenaler, 1976). Shephard (1980) and Zulfic et al. (2007) presumed that the K_h of the QL aquifer falls within a wider range of values across the entire basin, namely 30-2000 and 40-1400 m/d, respectively, in developing previous groundwater models. A single K_h estimate of 22 m/d is available for the TS aquifer (Sibenalar, 1976). Morton and Steel (1968) estimated the vertical hydraulic

conductivity (K_v) of the TC to be 6.8×10^{-4} m/d. Pumping test-based estimates of QL specific yield (S_y) range between 0.017 and 0.35 (Table 2.1; Sibenaler, 1976). For the TS aquifer, a single storage property estimate of 0.007 was obtained; however, it is unknown whether this value represents the S_y or storativity of TS because of the uncertainty surrounding the degree of aquifer confinement at this location (Morton and Steel, 1968).

Table 2.1. Pumping test-based hydraulic property estimates.

Observation well	Easting (m)	Northing (m)	Best estimate	Range	Best estimate	Range
			QL K_h (m/d)		QL S_y (-)	
Pt2	551132	6153175	690	410-690	0.07	0.05-0.18
Pt3	552176	6149817	150	-	0.10	-
Ptu4	553530	6152134	1320	-	0.017	-
Ptu5	549814	6151424	800	-	0.29	-
Ptu6	548850	6154438	-	>570	-	-
HDO6	551134	6152575	930	860-2040	0.14	0.03-0.15
HDO7	551336	6152061	1050	390-1490	0.12	0.07-0.13
HDO8	551681	6151436	370	280-470	0.026	0.02-0.10
HDO11	551866	6151986	1170	870-1260	0.079	0.07-0.10
HDO12	551798	6152800	1130	1130-1890	0.35	0.13-0.41
HDO13a	551642	6153301	420	180-690	-	-
HDO22	554251	6154742	171	-	-	-
			TS K_h (m/d)		TS storativity (-)	
Pt1	552582	6154685	22	-	0.007	-

Ordens et al. (2012, 2014) used field-based and one-dimensional modelling approaches to investigate USB recharge. In a subsequent study, Ordens (2014) extend field-based estimates to the basin scale through modelling within a GIS framework. Their work is used to provide spatially and temporally variable recharge for the purpose of this investigation (Figure 2.4). They obtain a spatially and temporally averaged value of 84 mm/y. This estimate is within ranges obtained by Bresciani et al. (2014), who modified the chloride mass balance-derived estimates of Ordens et al. (2012) to account for chloride deposition enhancement by vegetation. Although Ordens (2014) accounted for evapotranspiration (ET) at the surface and within the unsaturated zone, ET of shallow groundwater was not simulated explicitly. Groundwater ET is expected to occur only in localised regions and period of shallow watertables (Harrington et al., 2006; Figure 2.4).

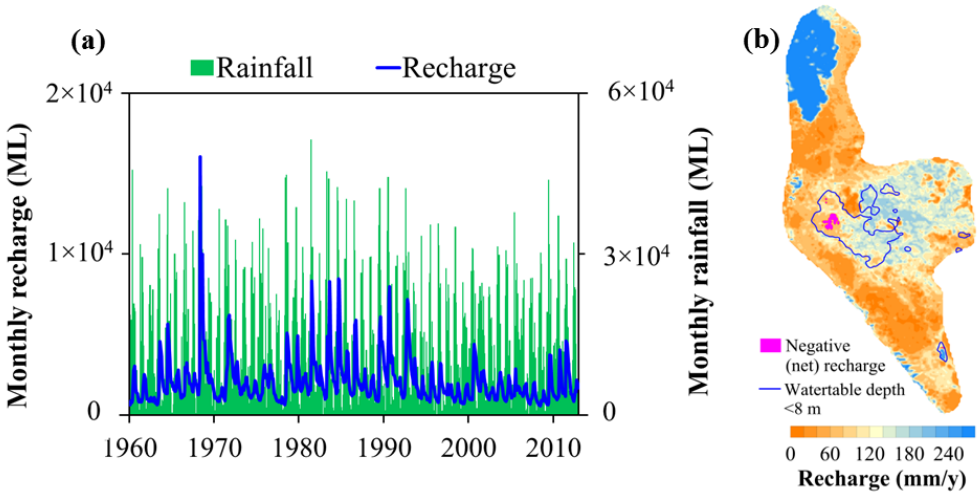

Figure 2.4. (a) Temporal (i.e., spatially averaged) and (b) spatial (i.e., temporally averaged over the period 1960–2012) recharge distribution estimates. Areas of negative net recharge (i.e., where ET rate exceeds that of recharge) are shown. The blue line indicates the extent of shallow watertable conditions, which is taken as the area in which pre-development groundwater levels are within 8 m of the land surface.

Groundwater is extracted from the QL aquifer by South Australian Water Corporation. Well extraction rates are monitored using automated meters (Werner et al., 2011). Monthly extraction rates for individual wells are approximated by combining the metered data reported by Werner et al. (2011) and extrapolation (both in time and to individual wells) based on averages from historical pumping records, in the absence of more recent pumping figures for individual wells. Figure 2.5 illustrates the resulting spatial distribution and time series of pumping.

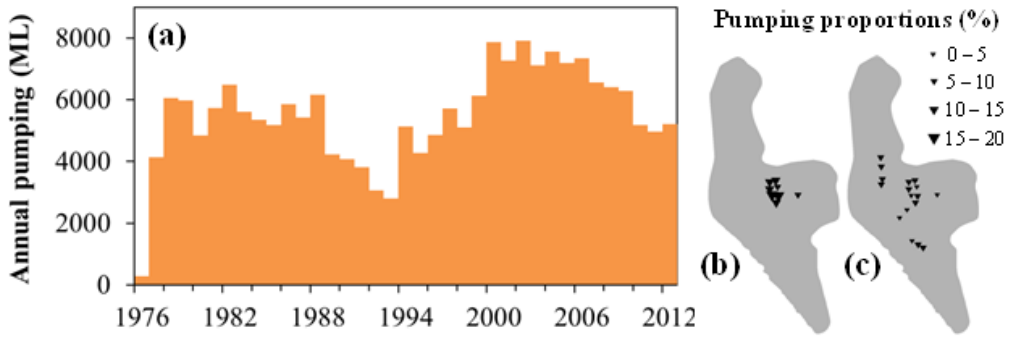

Figure 2.5. (a) Total pumping volume time series. Spatially distributed pumping proportions (as a % of total USB pumping) are given for periods (b) 1976–2000 and (c) 2000–2012.

2.4. Methodology

In order to assess the relative contribution of climatic and human impacts on groundwater decline in USB, we adopt a modelling strategy that involves comparison of simulated natural and disturbed conditions. This approach first requires the development of a calibration-constrained groundwater model using observation data from both pre-development and disturbed periods. Using the natural and disturbed condition predictions, in combination with observed conditions, the relative impact of climate variability and pumping is assessed on a temporally and spatially variable basis. Herein, it is assumed that climatic and human impacts are the

result only of recharge variability and groundwater pumping (given that USB is uninhabited and relatively minor land-use change has occurred during the simulation period; Evans, 1997), respectively. The performance of the model is evaluated using predictive uncertainty analysis based on linear theory.

2.4.1. Groundwater model

MODFLOW (Harbaugh et al., 2000) is used to simulate groundwater flow under both steady-state and transient conditions. It is presumed that steady-state conditions are a reasonable approximation of the groundwater level behaviour occurring during pre-development (1960–1976) and contemporary periods (2005–2009). A Newtonian solver (PCGN; Naff and Banta, 2008) is used to maximise the model's robustness in resolving water levels in parts of the aquifer that become dry.

The finite-difference grid is comprised of 2 layers (corresponding to the QL and TS aquifers), 245 rows and 132 columns (Figure 2.6). A uniform horizontal model discretisation of 100 m by 100 m is employed. A "quasi-three-dimensional" approach (Chiang and Kinzelbach, 1998) is adopted whereby groundwater flow within the TC unit is represented implicitly (i.e., vertical flow only and no TC storage effects) given that preliminary USB groundwater modelling efforts revealed that horizontal groundwater flow and storage within the TC were negligible at the basin scale.

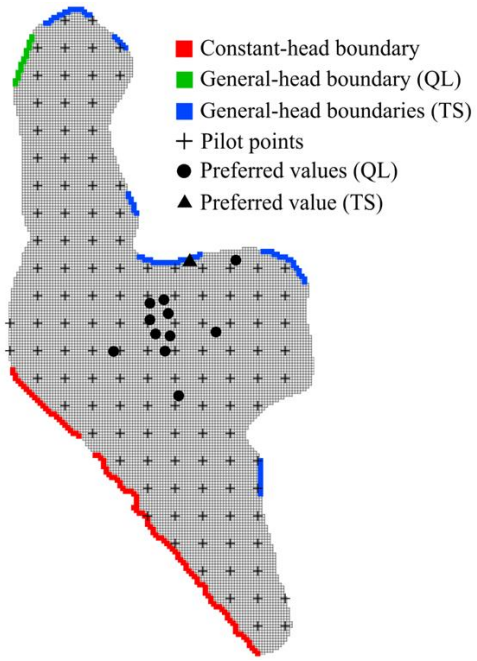

Figure 2.6. USB model grid, boundary conditions and parameterisation.

The transient simulation spans the period January 1960 to December 2012. Transient model stresses are applied using monthly stress periods. Each stress period contains 18 time steps, which increase in size starting from daily time steps. Boundary conditions are illustrated in Figure 2.6. Constant-head conditions are used to represent density-corrected mean sea-level at the shoreline. Head-dependent flux boundary conditions ("general-head" boundary (GHB) package; Harbaugh et al., 2000) are used to account for inter-basin flow. GHB forcing heads are specified according to time-averaged groundwater levels within adjacent basins. For transient simulations, initial conditions are specified using the results from the pre-development steady-state model.

Spatial and temporal recharge distributions from Ordens (2014) (Figure 2.4) are applied to the uppermost active layer of the model. Shallow groundwater ET is represented using the EVT package of MODFLOW. The ET rate for groundwater

levels at the land surface is equal to the average annual pan evaporation rate multiplied by a conversion factor of 0.85 (Taylor, 2003). An extinction depth of 2 m is used, above which the ET rate decreases with depth linearly.

2.4.2. Model calibration

Model calibration, which is used to estimate model parameters on the basis of steady-state and transient groundwater conditions, is performed using the automatic parameter estimation software PEST (Doherty, 2013). PEST adopts a gradient-search algorithm (i.e., Gauss-Marquardt-Levenberg method) to minimise an objective function on the basis of the sum of squared weighted residuals (i.e., the difference between model outputs and corresponding field observations). The parameter estimation process is parallelised using BeoPEST (Schreüder, 2009; Hunt et al., 2010).

The steady-state calibration is undertaken on the basis of both the pre-development and contemporary steady-state models. Average groundwater levels within each of these periods are used as the steady-state calibration dataset. The transient calibration dataset includes groundwater level measurements as deviations-from-the-mean, rather than absolute values, to better infer aquifer storage parameters (e.g., Ackerman et al., 2010; Peeters et al., 2011).

Weighting of groundwater level data is used as a means of ranking observations within the model calibration process in accordance to their credibility, whereby observations of higher credibility are assigned greater weights. For the steady-state

calibration, weights are applied to groundwater level observations based on an approximation of the inverse of the measurement error standard deviation (σ_m) (e.g., Doherty and Hunt, 2010), which is assumed to be 0.15 m for reliable observations. However, low-quality observations (e.g., see Section 2.3) are assigned a higher σ_m (i.e., ranging from 0.3 to 6 m). A minimum objective function value (e.g., Fienen et al., 2009) is specified to avoid fitting observations beyond a root-mean-squared weighted error (RMSE) of 0.15 m for the steady-state calibration. For transient model calibration, observation weights are assigned on the basis of the product of σ_m (equal to weights used in the steady-state calibration) and temporal groundwater level measurement density (i.e., number of measurements per year).

Spatial variability in both hydraulic conductivity (K_h and K_v) and storage parameters are specified using pilot point parameterisation (de Marsily et al., 1984). A total of 125 pilot points are distributed in a regular grid configuration (Figure 2.6). Additional pilot points are located where hydraulic parameters have been inferred from pumping tests, as recommended by Doherty et al. (2010). Cell-by-cell hydraulic property variability is achieved using interpolation (kriging) of the pilot point parameter values. A site-specific variogram could not be defined given insufficient data to support such an approach. Kriging is therefore based on an isotropic variogram of exponential type with a range of 1200 m (which corresponds to the distance between pilot point locations; Doherty et al., 2010) and a nugget of zero. A spatially and temporally constant recharge multiplier parameter is also employed in the calibration process. The GHB boundary is subdivided into segments (Figure 2.6) of piecewise constancy in conductance. Hydraulic conductivities, GHB conductances

and the recharge multiplier are estimated using the steady-state model. Storage parameters are estimated using the transient model.

Calibration of hydraulic parameters is constrained using "expert knowledge" in the form of likely ranges and best estimates of aquifer hydraulic properties, based on field evidence (Table 2.1; Section 2.3). In the absence of field estimates and geological information, textbook values (e.g., Fetter, 2001) that are considered representative of the setting are used to constrain estimation of TS K_h, S_y and S_s, and TC K_v. Parameter constraints are given in Table 2.2. Upper and lower parameter bounds are based on field estimates (Table 2.1) and typical literature values where reliable field estimates are lacking.

Table 2.2. Parameters estimated through model calibration, and their constraints. [a]Weights assigned to preferred values are relative to those assigned to preferred homogeneity constraints between neighbouring pilot point parameters. [b]Preferred S_y value weights are not comparable to preferred K_h, K_v, recharge multiplier and GHB conductance value weights.

Parameter(s)	Parameterisation method	Bounds	Preferred values	Relative weight[a]
Horizontal hydraulic conductivity (K_h) (m/d)				
QL	Pilot points	$1\text{-}10^4$	(Table 2.1)	2.0
TS	Pilot points	0.1-1320	(Table 2.1)	2.0
Vertical hydraulic conductivity (K_v) (TC) (m/d)	Pilot points	10^{-4}-1	-	-
Specific yield (S_y) (-)				
QL	Pilot points	0.007-0.41	(Table 2.1)	0.2[b]
TS	Single value	0.005-0.3	0.15	0.2[b]
Specific storage (S_s) (TS) (m^{-1})	Pilot points	10^{-7}-10^{-4}	-	-
Recharge multiplier (-)	Single value	0.5-1.5	1.0	2.0
GHB conductances (m^2/d)				
Uley East (TS)	Single value	0.36-3600	36	0.02
Uley Wanilla 1 (TS)	Single value	0.054-540	5.4	0.02
Uley Wanilla 2 (TS)	Single value	0.059-590	5.9	0.02
Uley Wanilla 3 (TS)	Single value	0.23-2300	23	0.02
Coffin Bay 1 (TS)	Single value	0.29-2900	29	0.02
Coffin Bay 2 (TS)	Single value	0.26-2600	26	0.02
Coffin Bay 3 (outflow; QL)	Single value	0.18-1800	18	0.02
South-east USB (TS)	Single value	0.07-770	7.7	0.02

To constrain and stabilise the calibration of the USB model, Tikhonov (Tikhonov and Arsenin, 1977) and subspace (e.g., Aster et al., 2005) regularisation schemes are employed. Tikhonov regularisation involves imposition of "soft" information, typically in the form of preferred parameter values and relationships (e.g., Alcolea et al., 2006), to supplement the calibration dataset such that parameters can be estimated uniquely (e.g., Menke, 1989). In this way, deviation from preferred values and relationships will arise only to the extent that is necessary to fit the data (Moore and Doherty, 2006). Preferred values are imposed on pilot point parameters that coincide with pumping test locations (where preferred values equal the pumping test-based hydraulic property estimates; Table 2.1 and 2.2). Estimation of hydraulic properties at remaining pilot points are constrained using preferred homogeneity. Preferred values are also assigned to boundary conductance parameters (equal to conductance estimates based on information pertaining to K_h and aquifer thickness, and the distance between the boundary and the GHB specified head) and the recharge multiplier (equal to unity). The relative weights used for preferred values (Table 2.2) reflect the degree of uncertainty in field-based estimates.

Subspace regularisation, in the form of singular value decomposition (SVD), constrains the inverse problem through a data-informed simplification process that transforms the large number of estimable parameters into a highly reduced number of estimable data-sensitive parameter combinations referred to as "super parameters" (e.g., Aster et al., 2005). The subspace approach used here is described by Tonkin and Doherty (2005), and is implemented within the PEST suite (Doherty, 2013). The computational savings afforded through SVD are invoked only for the transient calibration, which involves significantly longer run times. Here, 265 transient model

parameters are estimated in the form of 160 super parameters. The number of super parameters employed is based on the minimisation of predictive error variance, as suggested by Moore and Doherty (2005).

2.4.3. Distinguishing between climatic and pumping impacts

To investigate the relative contributions of climatic and pumping impacts to USB aquifer depletion, the calibration-constrained model is used to simulate groundwater levels under both natural and disturbed conditions. Natural conditions are simulated by running the model under historical climate stresses in the absence of pumping. Time series of groundwater level deviations attributable to climate and pumping effects are calculated, respectively, as:

$$\Delta h_c = h_{nat} - h_{nat}^{\text{ave}}, \tag{2.1a}$$

$$\Delta h_p = h_{nat} - h_{dist}, \tag{2.1b}$$

where Δh_c (L) represents the time series of climate-induced groundwater level deviations, h_{nat} (L) is the time-varying natural groundwater level, h_{nat}^{ave} (L) is the time-averaged h_{nat}, Δh_p (L) represents the time series of pumping-induced groundwater level declines (herein referred to as drawdown), and h_{dist} (L) is the time-varying disturbed groundwater level. Note that the contribution of climate to groundwater level decline is identified by negative values of Δh_c, and hence, we consider only negative Δh_c values herein. Time series of groundwater level deviations are obtained at every model cell for the purposes of analysing the variability in climate and pumping effects. Climate and pumping impacts are also

evaluated in terms of their contribution to basin-scale aquifer storage changes, as given respectively by:

$$\Delta S_c = S_{nat} - S_{nat}^{\text{ave}}, \tag{2.2a}$$

$$\Delta S_p = S_{nat} - S_{dist}, \tag{2.2b}$$

where ΔS_c (L^3) represents the time series of climate-induced groundwater storage deviations, S_{nat} (L^3) is the time-varying natural groundwater storage, S_{nat}^{ave} (L^3) is the time-averaged S_{nat}, ΔS_p (L^3) represents the time series of pumping-induced groundwater storage declines, and S_{dist} (L^3) is the time-varying disturbed groundwater storage. Only negative ΔS_c values are considered. The QL aquifer storage volume, S (L^3), is calculated, for each stress period, using:

$$S = \sum_{i=1}^{n} (S_y b)_i \Delta x \Delta y, \tag{2.3}$$

where n is the number of model-grid cells, b (L) is the aquifer thickness, and $\Delta x \Delta y$ (L^2) represents the area of the model-grid cell (each equal to 10^4 m^2). Here, it is assumed that S_y provides a reasonable surrogate for effective porosity in the absence of such information (e.g., Bear, 1972). Although basin-scale volumetric storage changes are analysed on the basis of the QL aquifer alone, this is not expected to influence the relative contributions of climate and pumping to groundwater depletion within USB.

2.4.4. Predictive uncertainty

Predictive uncertainty analysis is undertaken following the approach of Hill and Tiedeman (2007), Christensen and Doherty (2008), and Dausman et al. (2010b). That is, linearity between model parameters and predictions is presumed, and we neglect model structural defects (e.g., Refsgaard, 2006). The post-calibration uncertainty of a prediction, based on linear theory, is given by (Christensen and Doherty, 2008):

$$\sigma_s^2 = \mathbf{y}^t\mathbf{C}(\mathbf{p})\mathbf{y} - \mathbf{y}^t\mathbf{C}(\mathbf{p})\mathbf{X}^t[\mathbf{X}\mathbf{C}(\mathbf{p})\mathbf{X}^t + \mathbf{C}(\boldsymbol{\varepsilon})]^{-1}\mathbf{X}\mathbf{C}(\mathbf{p})\mathbf{y}\,. \tag{2.4}$$

In equation (2.4), σ_s^2 represents the post-calibration uncertainty variance of a prediction *s*. The sensitivity of σ_s^2 to the calibrated model parameters, which are encapsulated by the vector **p** (assumed to be normally distributed), is represented by the vector **y**. The innate variability of model parameters in **p** is expressed by the covariance matrix **C**(**p**). By assuming model linearity, the pre-calibration prediction variance can be estimated using the parameter covariance matrix **C**(**p**), as represented by the first term on the right-hand side of equation (2.4). The second term on the right-hand side of (4) represents the amount by which this prediction variance is reduced by constraining the model on the basis of the calibration dataset. Note that a post-calibration Jacobian matrix, **X**, which contains sensitivities of model outputs (that correspond to field observations) to calibrated model parameters, is required to compute the second term in (4). **C**(**ε**) is the covariance matrix of measurement error, where **ε** is also assumed to be normally distributed.

While efficient nonlinear methods have been presented and applied to highly parameterised groundwater models (e.g., Tonkin et al., 2007; Tonkin and Doherty, 2009; Herckenrath et al., 2011), linear uncertainty analysis has been shown to provide robust estimates of uncertainty, even if applied to nonlinear models (e.g.,

James et al., 2009; Brunner et al., 2012). Given that the objective here is to evaluate the relative magnitude of predictive uncertainty with respect to climate and pumping impacts, linear methods are considered appropriate.

The 95% confidence interval of the prediction, obtained on the basis of the standard deviation calculated via equation (2.4), is used to represent predictive uncertainty for the USB model. The Jacobian matrix used for this analysis is computed on the basis of all 662 parameters (at their calibrated values) involved in the steady-state and transient calibrations.

2.5. Results

2.5.1. Steady-state calibration

Figure 2.7 shows simulated-versus-observed steady-state groundwater levels in the QL aquifer for both pre-development and contemporary conditions. The RMSE is reduced from 2.44 to 0.154 m through calibration of the steady-state model. The post-calibration RMSE is approximately equal to σ_m, therefore indicating that the model is capable of reproducing groundwater levels during pre-development and contemporary periods to a level considered appropriate (e.g., Fienen et al., 2009). However, two outliers are evident in the pre-development calibration scatterplot (Figure 2.7a), which correspond to observation wells LKW10 and LKW11 (Figure 2.7c). These observations were assigned a low weighting prior to calibration because: (1) few groundwater level measurements are available (nine for LKW10 and seven for LKW11), and (2) groundwater level observations at LKW10 and LKW11 are

significantly higher than surrounding measurements, perhaps due to local watertable perching. We nonetheless include these observations in the calibration dataset for completeness.

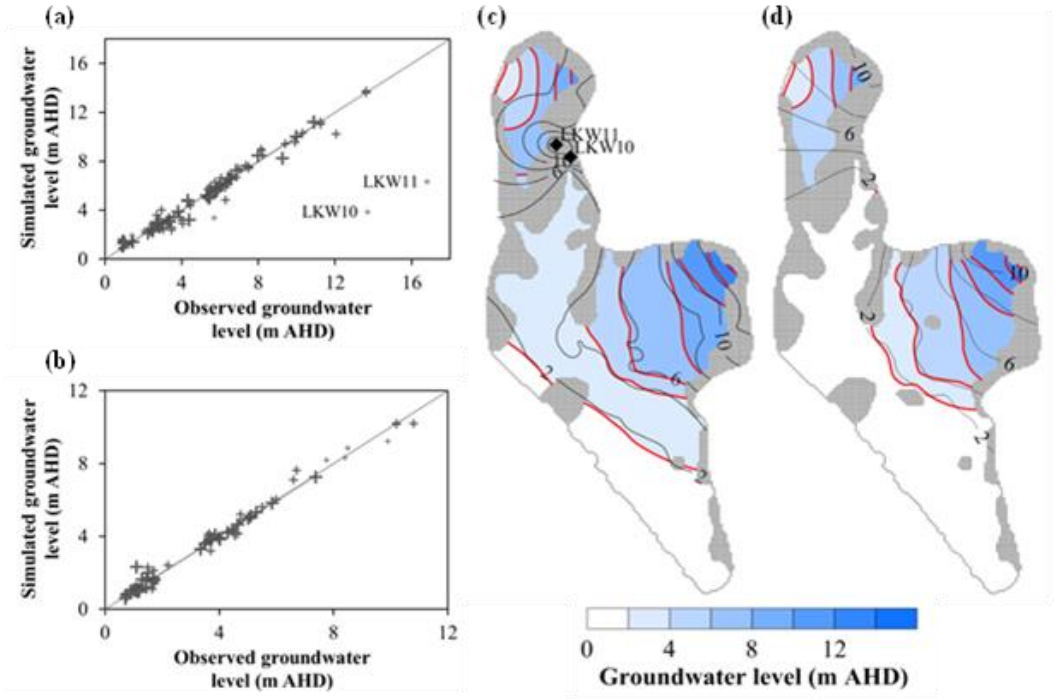

Figure 2.7. Steady-state calibration results for the QL aquifer, including simulated-versus-observed scatterplots of (a) pre-development and (b) contemporary conditions (where marker size reflects the relative weighting applied to each groundwater level observation and the line represents a model-to-measurement misfit of zero), and simulated (red)-versus-observed (black) groundwater level contours for (c) pre-development and (d) contemporary conditions. Observed contours are based on groundwater level observations shown in Figure 2.3a and b. Regions showing the model grid illustrate where the QL aquifer is dry.

Simulated groundwater level contours suggest that a distinct flow-divide exists in the QL aquifer between the north-west and central portions of USB, caused by a region where the QL is unsaturated (Figure 2.7c and d). These dry zones occur as a result of the undulating nature of the QL aquifer base. A number of observation wells in the north-west portion of USB, and a recent geophysical study (Davis et al., 2013), support the notion that the QL aquifer is dry in places. Natural discharge from USB's

QL aquifer is therefore expected to occur via two distinct pathways: (1) to the southern ocean, and (2) towards Coffin Bay. The lower groundwater levels during contemporary conditions (Figure 2.7d) cause a small reduction (1.2%) in the extent of saturated QL material.

Calibrated K_h, K_v and boundary conductance distributions are shown in Figure 2.8. The greatest degree of heterogeneity introduced through calibration occurs within the QL K_h field (Figure 2.8a). Most notable is the high K_h feature in the western portion of USB where low hydraulic gradients are observed within the QL (i.e., 5×10^{-5} m/m; Figure 2.3). Here, four pilot point parameters exceed 7×10^{3} m/d. A minimum QL K_h value of 2 m/d in central USB is obtained, where the hydraulic gradients are relatively steep (i.e., 2.5×10^{-3} m/m; Figure 2.3). The contrast between minimum and maximum QL K_h values is considered reasonable given the highly heterogeneous nature of the USB sediments (e.g., Harrington et al., 2006). Relative to the QL, a reduced adjustment of the TS K_h field (Figure 2.8c) arises via calibration. Minimum and maximum K_h values of 1 and 100 m/d are obtained for the TS, in the north-west and south-central parts of USB, respectively. The spatial variability in TS K_h reflects differences in hydraulic gradients, in a similar manner to the calibration results for the QL K_h field. The calibrated TC K_v field (Figure 2.8b) largely reflects expert knowledge regarding the discontinuous nature of the TC (see Figure 2.2). In areas where the TC is known to be present, the calibrated TC K_v ranges between 2.9×10^{-4} and 9.1×10^{-4} m/d. A high degree of variability in boundary conductance values across USB is obtained through calibration (Figure 2.8d). The largest modification to GHB conductance through calibration is 18 to 984 m^2/d, which occurs within the westernmost Coffin Bay GHB reach (red line; Figure 2.6). This level of modification

is deemed reasonable given the uncertainty in the boundary conductance. Finally, recharge rates were essentially unmodified by the steady-state calibration (i.e., 3×10^{-6}% adjustment).

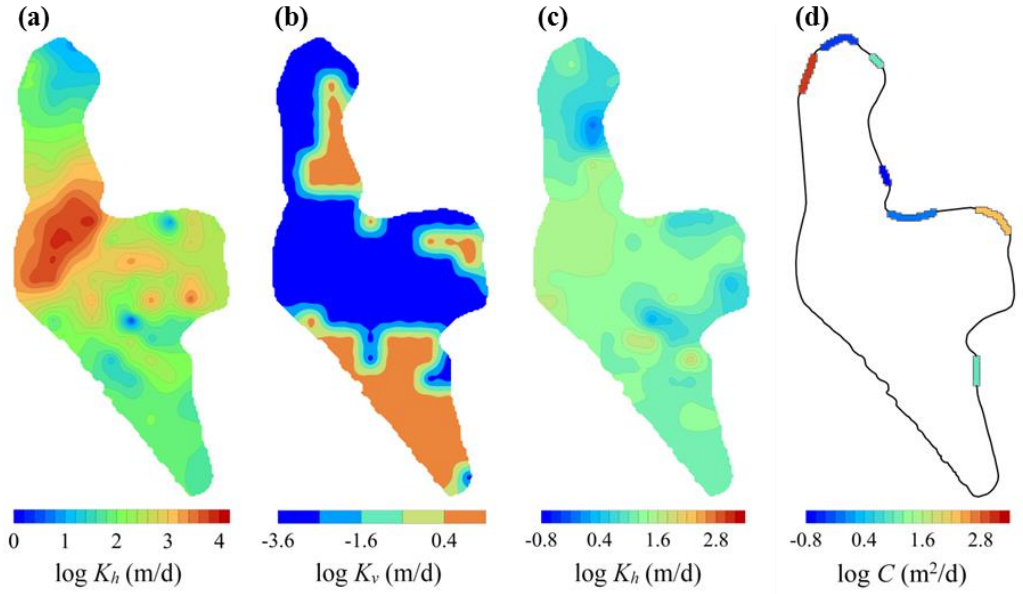

Figure 2.8. Calibrated parameter distributions: (a) K_h of layer 1 (QL aquifer), (b) K_h of layer 2 (TS aquifer), (c) K_v of the implicit TC "layer" and (d) boundary conductance for each GHB reach.

The maximum deviation of a pilot point parameter value from its preferred value, through calibration, is 0.6 m/d, which occurs within the QL K_h field. The magnitude of this deviation is considered relatively insignificant given the degree of variability within the calibrated K_h field (Figure 2.8a). Here, preferred parameter values are hardly modified by calibration because of the high relative weighting used to impose field-based K_h estimates (Table 2.2). A high relative weighting is also the reason for the spatially uniform recharge multiplier remaining at its preferred value (of unity). The average relative deviation of boundary conductances from preferred values is 734%. Such deviation is comparably large, and is due to the low relative weighting used to reflect the uncertainty of preferred conductance values (Table 2.2). A

discussion on the trade-off between fitting observation data and the imposition of preferred parameter values is provided in Section 2.6.

Volumetric water balances for pre-development and contemporary steady-state conditions are given in Figure 2.9a and 2.9b, respectively. The reduction in groundwater discharge to the ocean from 45 ML/d (pre-development) to 18 ML/d (contemporary) is of particular significance given its association with increased vulnerability to seawater intrusion (Werner et al., 2012). Inflow through the USB coastal boundary is due to the application of density-corrected head conditions to the QL and TS aquifers. The resulting groundwater re-circulation in the coastal fringe is typical of the flow patterns in coastal aquifers (e.g., Post et al., 2013). The (net) inflow from adjacent basins is approximately 24% greater under the lower groundwater levels of the contemporary period than under pre-development conditions. ET of shallow groundwater is a relatively minor component of the water balance (0.012 ML/d under pre-development conditions and 0 ML/d under contemporary conditions), as expected, and is therefore not shown in Figure 2.9.

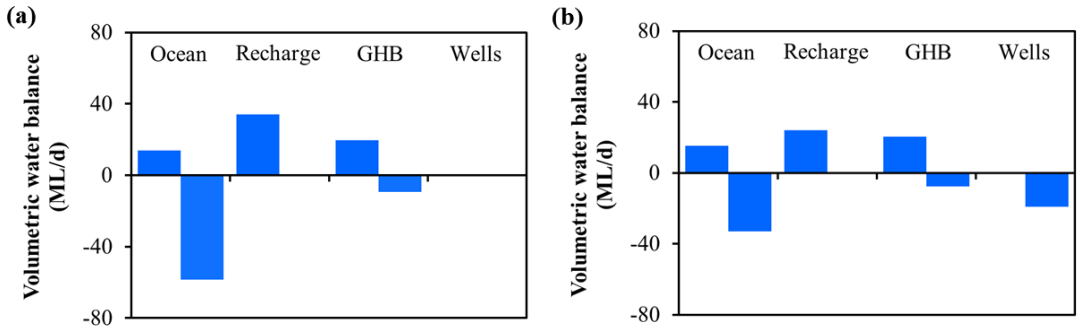

Figure 2.9. Volumetric water balance for (a) pre-development and (b) contemporary steady-state conditions. Positive fluxes indicate inflow to USB; negative fluxes indicate outflow from USB.

2.5.2. Transient calibration

Figure 2.10 shows the scatterplot of simulated-versus-observed deviation-from-the-mean groundwater levels obtained through calibration of the transient USB model. The RMSE is reduced from 0.25 to 0.24 m through calibration. Here, the RMSE value prior to transient calibration is low because it is based on parameters obtained from the steady-state calibration. For individual simulated-versus-observed (absolute) hydrographs, the bias (i.e., average weighted error) varies from -1.22 to 0.50 m (average of 0.01 m) and the average absolute weighted error varies from 0.01 to 1.22 m (average of 0.14 m), which indicates a reasonable level of model-to-measurement misfit. The positive bias values indicate an overestimation of groundwater levels by the model. As expected, the largest model errors correspond to observations that are of low credibility (Section 2.4.2). The simulated and observed hydrographs from four representative observation wells (SLE006, ULE136, ULE112 and LKW034) are shown in Figure 2.10. These are selected because they: (1) span a large spatial extent of USB, (2) have relatively long-term water level records, (3) represent groundwater behaviour from both aquifers, and (4) highlight variations in the calibration model-to-measurement misfit. Both temporal groundwater level fluctuations and time-average water levels are well-captured by the model at SLE006, ULE136 and ULE112, where average groundwater level misfits equal 0.01, -0.17 and -0.04 m, respectively (negative misfit indicates model underestimation). For LKW034, the groundwater level is overestimated, on average, by 0.40 m, which is commensurate with the fit obtained from the steady-state calibration.

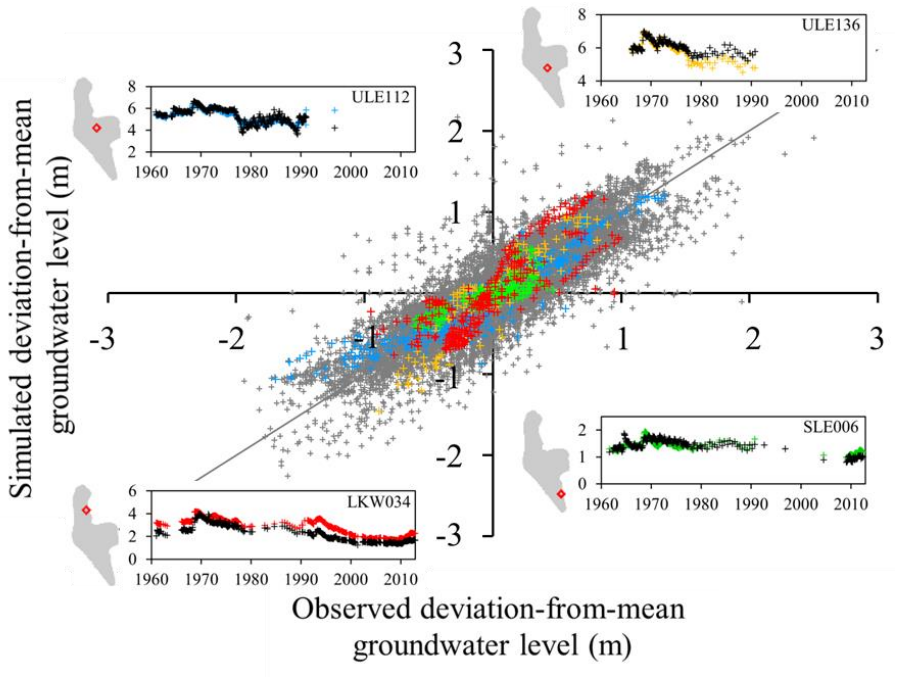

Figure 2.10. Transient calibration results, as shown by the scatterplot of simulated-versus-observed deviation-from-the-mean groundwater levels (m), along with simulated-versus-observed (absolute) groundwater level hydrographs (m AHD) for four representative observation wells (SLE006, ULE136, ULE112 and LKW034). Well locations are indicated to the left of each hydrograph.

Calibrated storage parameter fields (S_y of QL and S_s of TS) are presented in Figure 2.11. A greater degree of adjustment occurs within the QL S_y field (Figure 2.11a). By visual inspection of the calibrated S_y field, a general relationship between calibrated S_y values and the magnitude of groundwater level fluctuation (within the QL aquifer) is evident whereby high values of S_y correspond to regions where relatively low fluctuation in groundwater levels are observed, and vice versa. For example, four pilot point parameters located in north-central USB reach their upper bound (0.41), which reflects the relatively low groundwater level fluctuation observed on the basis of three observation wells (ULE120, ULE121 and ULE122) located within this region. The lowest S_y values correspond to preferred values (Figure 2.6) imposed on the basis of pumping test-based S_y estimates (Table 2.1). Calibrated S_y values do not

deviate significantly from preferred values (i.e., the maximum deviation is 1.5×10^{-3}), as discussed further in Section 2.6. Calibrated S_s values range from 2.43×10^{-6} to 3.24×10^{-6} m^{-1} (Figure 2.11b), which generally reflects the variability in magnitude of TS groundwater level fluctuation in a similar manner to the relationship between calibrated S_y values and groundwater level fluctuations within the QL.

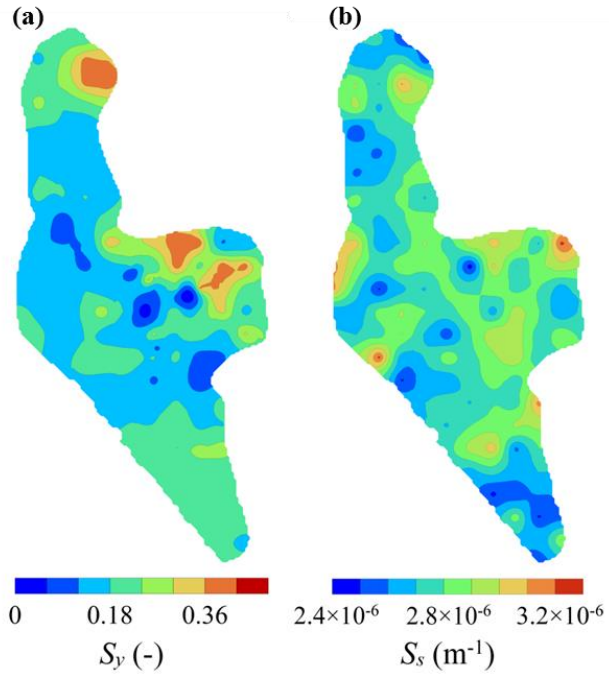

Figure 2.11. Calibrated storage parameter fields: (a) S_y of layer 1 (QL) and (b) S_s of layer 2 (TS).

Figure 2.12 shows the transient volumetric water balance. Discharge to the ocean varies significantly over time (mean of 37 ML/d, with a standard deviation of 11 ML/d or 29%). Historical discharge displays a similar trend to groundwater levels (e.g., decline during 1994–2004), as expected, and represents an attenuated response to recharge and pumping. Net inter-basin flow is relatively stable throughout the investigation period (mean of 10 ML/d, with a standard deviation of 1.5 ML/d or 15%). The largest positive storage changes (i.e., in a single year) occur prior to 1994,

after which recharge events and resulting increases in storage are more subdued. Groundwater ET is not shown here due to its minor contribution to the water balance.

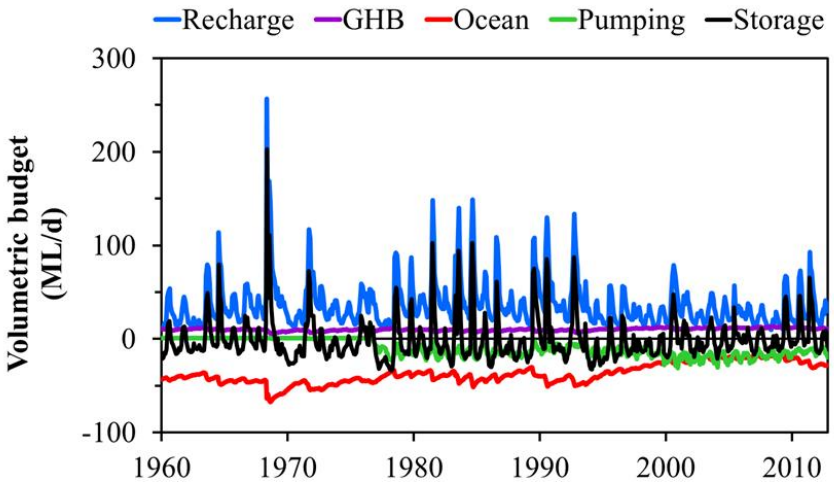

Figure 2.12. Transient volumetric water balance. The time series of groundwater discharge to the ocean (labelled "ocean") represents the net oceanic discharge (i.e., the difference between inflow and outflow through the coastal boundary; Section 2.4.1). Similarly, the time series of flow between adjacent basins (labelled "GHB") represents the net inter-basin flow (i.e., the difference between inflow from Uley East, Uley Wanilla and Coffin Bay and outflow to Coffin Bay; Section 2.2).

2.5.3. Climatic and pumping impacts

Figure 2.13 shows the time series of h_{nat} and h_{dist} for wells SLE006, ULE136, ULE112 and LKW034. Observed groundwater levels and the uncertainty of simulated groundwater levels (as demonstrated using the 95% confidence interval) are also shown.

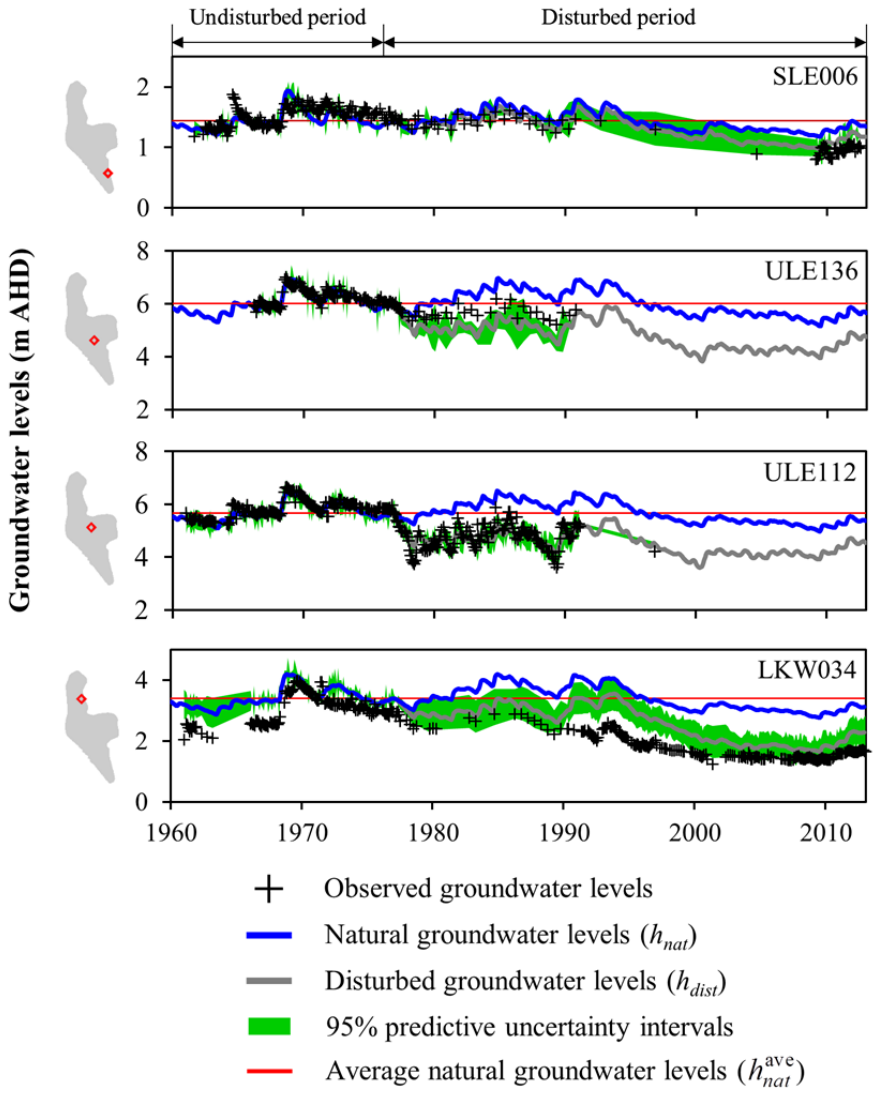

Figure 2.13. Time series of h_{nat} and h_{dist} at observation wells SLE006, ULE136, ULE112 and LKW034 (well locations shown to the left of hydrographs).

The time series of h_{nat} display distinct seasonal and inter-annual variability (Figure 2.13). The time-varying Δh_c at these observation wells, along with the annual cumulative deviation-from-the-mean rainfall (CDMR; e.g., Buishand, 1982), is shown in Figure 2.14. Long-term h_{nat} trends are characterised by (Figures 2.13 and 2.14): (1) a low phase from 1960–1967 (where the mean rainfall rate during this period is 535 mm/y), (2) rapid increase during 1968–1969 (mean rainfall of 702 mm/y), (3) somewhat steady decline between 1970 and 1977 (mean rainfall of 548 mm/y), (4) gradual increase from 1978–1986 (mean rainfall of 612 mm/y), (5)

decline during 1987–1989 (mean rainfall of 534 mm/y), (6) mostly incline between 1990 and 1992 (mean rainfall of 652 mm/y), (7) somewhat gradual decline during 1993–2009 (mean rainfall of 520 mm/y), and (8) incline between 2010 and 2012 (mean rainfall of 483 mm/y), due primarily to the 2010 annual rainfall of 656 mm (i.e., the largest annual rainfall since 1992). These trends generally resemble patterns in annual rainfall variability. Despite these fluctuations, h_{nat} is somewhat stable over the 52-year simulation period, as evidenced by a regression line slope of -10^{-4} m/y. This is consistent with a mostly stable rainfall trend for the period of investigation (i.e., slope of -1.2 mm/y^2 for 1960–2012). The largest-in-time Δh_c at SLE006, ULE136, ULE112 and LKW034 is 0.26, 0.86, 0.70 and 0.65 m, respectively (Figure 2.14). These maxima occur during 2009 following the long-term below-average rainfall period between 1993 and 2009. Time-averaged (1960–2012) Δh_c for these wells is 0.10, 0.32, 0.26 and 0.27 m, respectively.

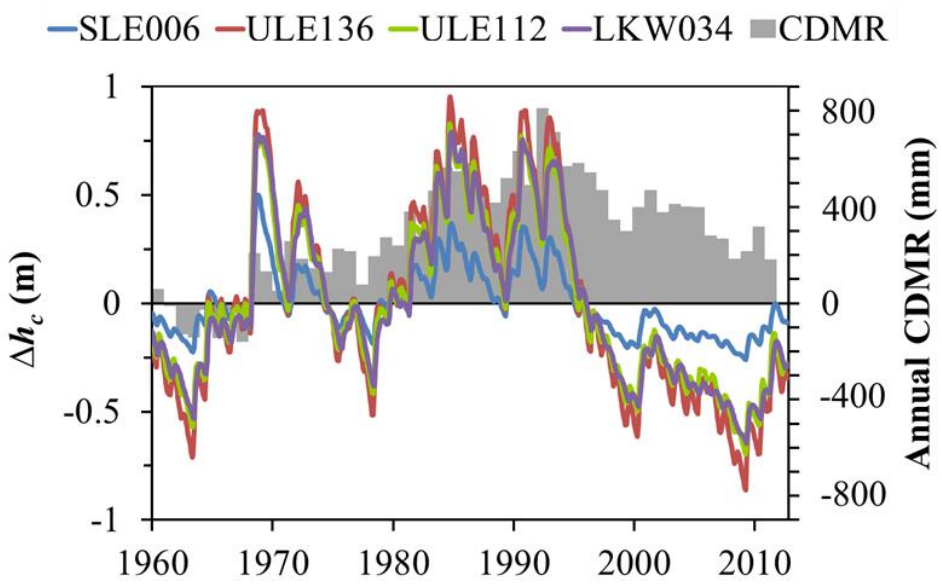

Figure 2.14. Time series of Δh_c at observation wells SLE006, ULE136, ULE112 and LKW034, and the annual cumulative deviation-from-the-mean rainfall (CDMR) (grey bars).

The time series of h_{dist} display considerable variability in Δh_p following the commencement of pumping in 1976 (Figure 2.13). The time-varying Δh_p in these wells, along with annual pumping volumes, is shown in Figure 2.15. The following Δh_p trends are mostly consistent for ULE136, ULE112 and LKW034 (Figures 2.15): (1) increase between 1976 and 1988 (where the mean pumping rate during this period is 5161 ML/y), (2) gradual reduction from 1989–1993 (mean pumping of 3593 ML/y), (3) steady increase from 1994–2000 (mean pumping of 5584 ML/y), (4) somewhat stable period from 2001–2007 (mean pumping of 7280 ML/y), and (5) reduction between 2008 and 2012 (mean pumping of 5608 ML/y). These trends are strongly correlated to the variability in annual pumping volumes. A different response is evident in SLE006, whereby Δh_p is mostly stable prior to a distinct increase in 2001. This increase in Δh_p, also clear in LKW034 (during 2000), occurs as a result of the pumping well-field expansion in 2000 (Figure 2.1). At this time, a decrease in Δh_p at ULE112 is apparent due to a significant reduction in pumping in central USB. In general terms, pumping and rainfall are somewhat inversely related,

as expected given the strong seasonality of the USB climate. This inhibits the interpretation of pumping and climate impacts from simple hydrograph analysis (i.e., in the absence of the numerical model). The maximum-in-time Δh_p at SLE006, ULE136, ULE112 and LKW034 is 0.24, 1.57, 1.61 and 1.28 m, and occurs during 2007, 2000, 2000 and 2004, respectively. Time-averaged (1976–2012; i.e., the period of pumping) Δh_p for these wells is 0.13, 1.19, 1.16 and 0.81 m, respectively.

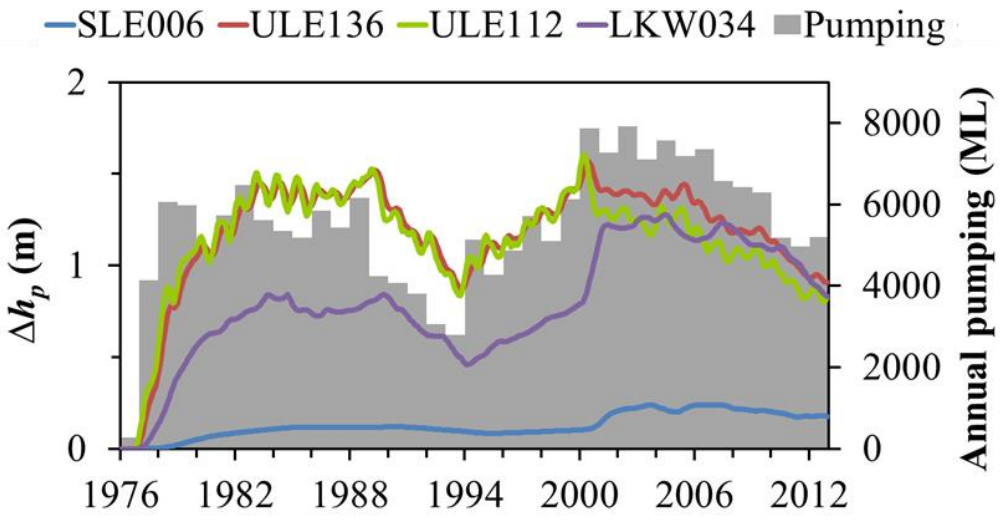

Figure 2.15. Time series of Δh_p at observation wells SLE006, ULE136, ULE112 and LKW034, and the annual pumping volumes (grey bars). Time series are truncated prior to pumping (1976).

The uncertainty of groundwater level predictions at SLE006, ULE136, ULE112 and LKW034, on average, is 0.08, 0.22, 0.30 and 0.35 m, respectively (Figure 2.13). Because the h_{nat} lies mostly outside of the 95% uncertainty interval of h_{dist} (Figure 2.13), the relative magnitude of predictive uncertainty is generally lower than the relative impact of pumping. The 95% predictive uncertainty intervals are also, in general terms, lower than the influence of climate on groundwater levels.

Figure 2.16 shows time series of S_{nat} and S_{dist}. Time-varying ΔS_c and ΔS_p display good agreement with observation well-averaged Δh_c and Δh_p trends (Figure 2.13), as evidenced by the combined linear regression between ΔS_c and Δh_c and ΔS_p and Δh_p, with a coefficient of determination (r^2) value of 0.99. ΔS_c exceeds ΔS_p only within the first two years following the initiation of pumping (i.e., 1976–1978). The largest-in-time ΔS_c of 1.1×10^4 ML occurs during June 2009, after a period of below-average annual rainfall between 1993 and 2009 (Figure 2.14). The maximum-in-time ΔS_p of 1.6×10^4 ML (143% that of ΔS_c) occurs during January 2001, following the approximately equal largest annual pumping rate in 2000 (7873 ML/y; Figure 2.15). The time-averaged (1960–2012) ΔS_c is 4.3×10^3 ML, compared to the time-averaged (1976–2012) ΔS_p of 1.2×10^4 ML, i.e., time-averaged ΔS_p is 2.9 times that of ΔS_c.

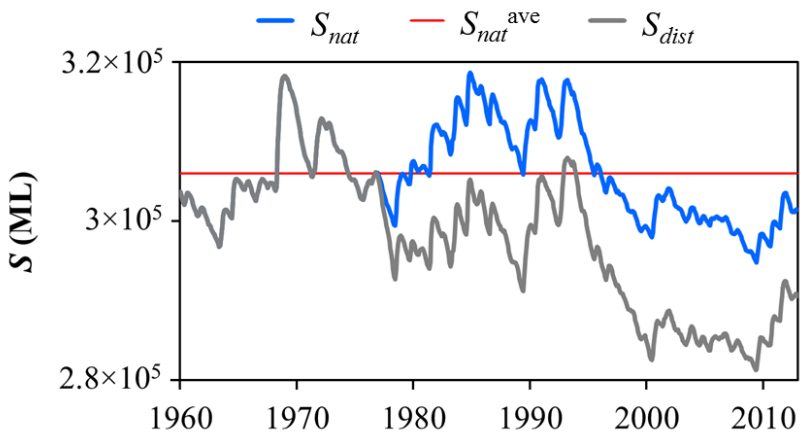

Figure 2.16. Time series of S_{nat} and S_{dist}.

Figure 2.17a and b show the spatial distribution of Δh_c and Δh_p during June 2009 and January 2001 (i.e., the times at which the largest ΔS_c and ΔS_p occur), respectively. The spatially distributed difference between the maximum-in-time Δh_c (Δh_c^{max}) and Δh_p (Δh_p^{max}), on a cell-by-cell basis, is shown in Figure 2.17c.

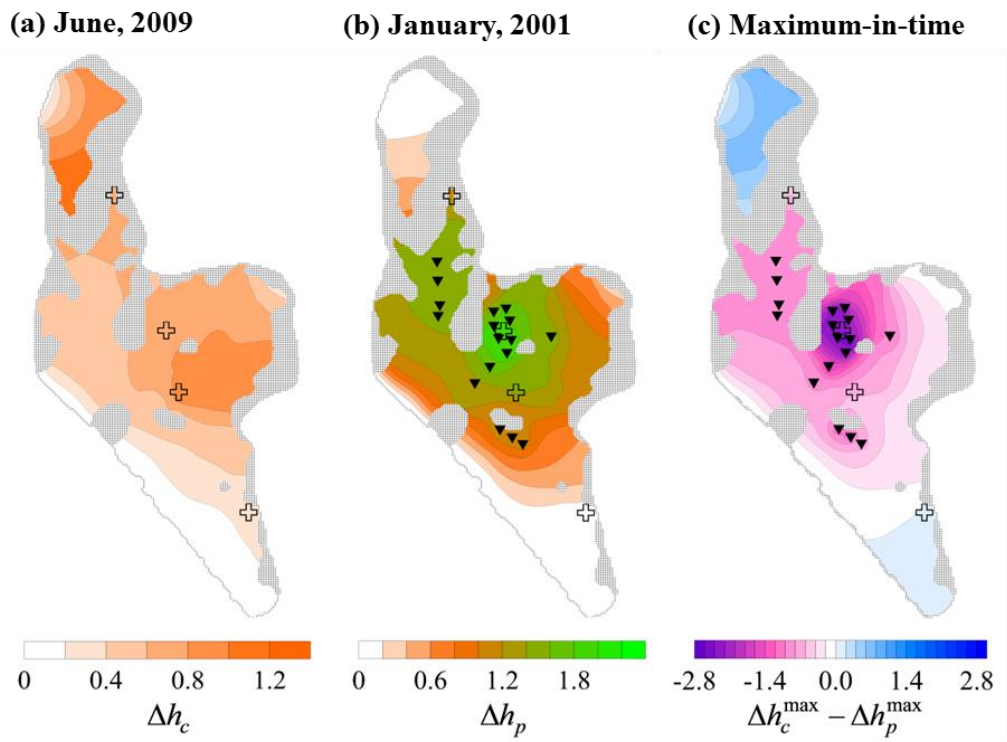

Figure 2.17. The spatial distribution of (a) Δh_c during June, 2009, (b) Δh_p during January, 2001, which correspond to the times at which maximum ΔS_c and ΔS_p occur, respectively. (c) The spatially distribution (on a cell-by-cell basis) of $\Delta h_c^{max} - \Delta h_p^{max}$. Negative $\Delta h_c^{max} - \Delta h_p^{max}$ (shown in purple) values correspond to regions where the maximum Δh_p exceeds that of Δh_c, whereas positive values (shown in blue) correspond to areas where the largest Δh_c exceeds that of Δh_p.

The spatial distribution of Δh_c (within the QL aquifer) exhibits a variability of 1.1 m across USB during June 2009 (Figure 2.17a). Areas of relatively low Δh_c values (<0.5 m) are found in close proximity to constant- and general-head boundary conditions (e.g., near Coffin Bay and Uley East GHBs, and coastal USB), where groundwater levels exhibit smaller fluctuation. The dampened groundwater level fluctuation near GHBs occurs as a result of the time-averaged nature of the inter-basin flows. Otherwise, the spatial distribution of Δh_c is related to aquifer properties and recharge in a complicated manner, as demonstrated by the relatively weak correlation between Δh_c and hydraulic property values on a cell-by-cell basis (e.g.,

the regression between cell-by-cell Δh_c and S_y values produce an r^2 value of 0.17). Similar Δh_c trends are also found within the TS aquifer.

The spatial distribution of Δh_p (within the QL aquifer) varies by 2.1 m across USB during January 2001 (Figure 2.17b). During this time, Δh_p exceeding 1 m spans an area of about 55 km^2 around the USB pumping well field. Smaller Δh_p values (<0.5 m) are found in regions that are: (1) restricted from the effects of pumping by unsaturated sediment, and (2) near head-dependent boundary conditions. In a similar manner to Δh_c trends, the spatial distribution of Δh_p exhibits complicated relationships with those of aquifer properties and recharge, and displays similar trends within the TS aquifer.

Figure 2.17c shows that the spatial distribution of $\Delta h_c^{max} - \Delta h_p^{max}$ values vary between -2.8 and 0.9 m. The variability in space of Δh_p exceeds that of Δh_c because of the localised point-based nature of pumping stresses compared to the diffuse nature of recharge. Negative values of $\Delta h_c^{max} - \Delta h_p^{max}$ are found over the majority (80%) of the spatial area of USB.

2.6. Discussion

The performance of the USB model, in terms of replicating observed groundwater levels, is considered to be of an appropriate level given the complex nature of regional USB groundwater behaviour. Calibration of the steady-state models led to an RMSE approximately equal to the value of σ_m assumed prior to calibration (0.15 m; Figure 2.7). This means that the pre-defined minimum objective function is

achieved, thus avoiding "overfitting" (Fienen et al., 2009). The transient calibration produces an RMSE of 0.24 m (Figure 2.10). A systematic overview of all simulated-versus-observed hydrographs indicates that a reasonable level of model-to-measurement misfit is achieved. The "goodness of fit" is comparable to error statistics reported for the application of spatially distributed models to various regional-scale hydrology problems (e.g., RMSE values of 1.79-2.09 m, 0.12-2.92 m, 1.03 m, 0.16-0.44 m and 3 m are reported by Gurwin and Lubczynski (2005), Reynolds and Marimuthu (2007), El Yaouti et al. (2008), Demissie et al. (2009) and Yihdego and Becht (2013), respectively). Significantly greater reduction in model-to-measurement misfit is achieved through calibration of the steady-state models (1500% RMSE reduction) as opposed to that of the transient model (6% RMSE reduction). This arises partly because parameters obtained from the steady-state calibration are adopted in the transient calibration.

Parameter values (and fields) obtained through calibration are considered to be geologically realistic, at least within the limits of available field evidence, because expert knowledge in the form of field-based hydraulic properties (Table 2.2) are preserved within the models' parameter distributions. That is, calibrated K_h and S_y values do not deviate significantly from their preferred values (a maximum deviation of approximately 1%, occurring within the QL K_h field, is obtained). Moreover, the high degree of heterogeneity within the calibrated QL K_h field (Figure 2.8a) is deemed reasonable for a setting that exhibits karstic features (e.g., Harrington et al., 2006), where K_h values may vary over five orders of magnitude (e.g., Kiraly, 1998; Worthington, 1999). These outcomes highlight the advantages of the high degree of flexibility within the parameterisation of the USB model. That is, the use of many

pilot point parameters allows for an appropriate level of heterogeneity to be accommodated, as informed on the basis of observed groundwater level trends, without the need for parameters to deviate significantly from their preferred values (e.g., Doherty, 2003).

The relative magnitude of the measurement and regularisation components of the minimised objective function provides some insight into the trade-off between the minimisation of model-to-measurement misfit and the deviation from preferred parameter values. For the steady-state calibration, both components are of similar magnitude (differing by 1.3%), thus demonstrating that an appropriate balance between fitting field observations and the imposition of expert knowledge regarding parameter values is obtained. Conversely, the measurement component of the transient calibration objective function is considerably larger (288%) than its regularisation counterpart. This arises primarily from the lack of expert knowledge of storage parameters. Although sophisticated techniques for optimal weight determination are available (e.g., Pareto analysis; Vrugt et al., 2003), their application to regional groundwater models is often avoided given the computational times required, and as such, is beyond the scope of the current study. The different scales that pilot point parameters and pumping test-based estimates represent (e.g., Maneta and Wallender, 2013) may also be an important factor in understanding the trade-off between fitting observation data and imposing expert knowledge.

Calibration of the USB models provides other important insights into hydraulic parameter distributions. This is evidenced by (e.g., Figure 2.11a): (1) pilot point-S_y parameters reaching their upper bound, and (2) calibrated S_y values generally

exceeding pumping test-based estimates. These outcomes suggest, collectively, that calibrated S_y parameters may be assuming surrogate roles, e.g., to offset errors within model structure (given the nonuniqueness of the groundwater model inverse problem; e.g., Carrera and Neuman, 1986). The extent to which calibrated S_y (and K_h) values compensate for errors within spatial and temporal recharge distributions and boundary representations is the subject of future work. The exceedance of field-based S_y estimates by calibrated values suggests that there may be significant, yet unquantified, uncertainty associated with field-based S_y estimates (Table 2.1). Also, by assessing the parameter combinations constructed through SVD regularisation, it is evident that insensitive parameter combinations are composed largely of pilot point S_y parameters located in data-poor areas (e.g., north-west USB). This provides an indication of which parameters are relatively poorly constrained (or "unidentifiable"; Doherty and Hunt, 2009), and where future data collection should potentially be focused.

The model predictions on which the current investigation is based are considered to be of acceptable accuracy. This is highlighted by the outcomes of the linear uncertainty analysis, which demonstrate that predictive uncertainty of the model, in general terms, is significantly smaller compared to the relative contributions of climate and pumping impacts (Figure 2.13).

Both the observation well and basin-wide storage analyses produce similar estimates of relative climate and pumping contributions to aquifer decline (Figure 2.13 to 2.16), as evidenced by the fact that pumping impacts are (by considering both time-averaged and maximum-in-time statistics) approximately 2.7 and 2.2 times that of

climate based on these analyses, respectively. The consistency between results from these analyses demonstrates that the four wells selected provide an adequate collective representation of basin-scale USB behaviour. The lower relative contribution of climate to groundwater decline from time-averaged (rather than maximum-in-time) values occurs because climate-induced deviations vary about zero (Section 2.4.3). Despite the difference between the relative climate and pumping impacts based on time-averaged and maximum-in-time values, these diagnostics, collectively, provide a basis for judging climate and pumping contributions in a generalised manner to inform future USB management.

The modelling approach adopted in the current study can also be extended to provide insight into the cause of recent (i.e., since 2009) partial groundwater recovery. That is, a decrease in ΔS_c and ΔS_p between January 2009 and December 2012 is apparent, equal to 5.4×10^3 and 2.9×10^3 ML, respectively (Figure 2.16). Hence, the contribution of climate to recent groundwater storage recovery is approximately twice that of pumping. The models used as a basis for this investigation also provide insight into USB groundwater behaviour more generally. For example, the flow divide that is apparent in the QL aquifer (Figure 2.7) indicates that the northwest portion of the current USB extent should be included within the Coffin Bay QL aquifer boundary, and that the USB aquifer, as defined by the extent of saturated QL sediments, is significantly smaller than previously thought (e.g., Zulfic et al., 2007). This has implications for the management of USB given that groundwater extraction allocations have been based traditionally on a proportion of recharge volumes, which take into account the saturated area of the QL aquifer (DFW, 2012).

The primary limitation associated with model-based approaches such as those adopted in the current study is the requirement of reliable climate and anthropogenic stress data and observation data (on which calibration is based), spanning both natural and disturbed periods. Because long-term and well-constrained stress and observation data are often scarce, the application of spatially distributed models are sometimes avoided for preference of simpler, less data-intensive methods (e.g., Helsel and Hirsch, 2002; Hamed, 2008). However, the uncertainty associated with stress (e.g., groundwater extraction; Kerrou et al., 2013) and observation data (as considered in the current study) can be represented comprehensively through the use of highly parameterised models (e.g., Doherty et al., 2011). This avoids the need for highly simplified approaches (which preclude representation of physically based, spatially and temporally distributed hydrological response) because of data constraints.

2.7. Conclusions

There is a global need for quantification of the relative contributions of natural and human stresses to groundwater depletion to inform water resource management. To date, spatially distributed models, which provide a basis for representing many of the complexities associated with field-scale groundwater processes, have not been applied in this context. In this paper, we extend previous modelling studies through use of a highly parameterised, spatially distributed groundwater model to distinguish between the relative impacts of climate variability and pumping within a regional setting.

Results show that, for USB, the contribution of pumping to aquifer depletion is 2.9 and 1.4 times that of climate, based on the time-averaged and maximum-in-time basin-wide volumetric storage changes, respectively. Pumping impacts are shown to exceed climate impacts between 1978 (i.e., two years after pumping initiation) and 2012, and over the majority (80%) of the spatial area of USB. These results serve as a response to the recent parliamentary enquiry (NRC, 2013) into the cause of groundwater decline by establishing that pumping impacts are generally greater than climate impacts. Future USB groundwater management is expected to be enhanced through adaption of pumping rates to better align with the condition of the aquifer attributable to both climate and pumping impacts.

This study demonstrates the applicability of a relatively simple modelling strategy, in combination with a highly parameterised, spatially distributed groundwater model, for quantifying the relative contribution of natural and human aquifer impacts. Use of spatially distributed models within this framework allow for the distinction to be made between the contributions of individual stresses to hydrological system response on a high-resolution in space and time. In this way, specific causal factors of groundwater response can be identified. This approach can be extended to quantify the relative impacts of stresses for a wide range of hydrological settings, which may involve different data types (e.g., stream flow, solute concentrations), and where human influences such as land-use change (e.g., Zhang et al., 2013) and reservoir regulation (e.g., Wanders and Wada, in press) are of interest. Furthermore, this approach can be applied within a scenario-testing framework whereby the contributions of various potential climate change and future water management scenarios are investigated.

Acknowledgements

The author M.J. Knowling was supported by an Australian Postgraduate Award, and top-up scholarships from the Goyder Institute for Water Research and the National Centre for Groundwater Research and Training (NCGRT), a collaborative initiative of the Australian Research Council and the National Water Commission. The authors gratefully acknowledge Ty Watson for assistance with PEST and model post-processing, and the following people for their contribution to this work: Carlos Ordens, John Hutson, Etienne Bresciani, Julie McClements, Darren Alcoe, and Le Dang. The authors also thank Marco Maneta, and two anonymous reviewers for their valuable comments.

Chapter 3

3. Estimability of recharge through groundwater model calibration: Insights from a highly parameterised field-scale steady-state example

This chapter is based on the following paper:

Knowling, M.J., Werner, A.D., 2016. Estimability of recharge through groundwater model calibration: Insights from a field-scale steady-state example. Journal of Hydrology, doi: 10.1016/j.jhydrol.2016.07.003.

3.1. Abstract

The ability of groundwater models to inform recharge through calibration is hampered by the correlation between recharge and aquifer parameters such as hydraulic conductivity (*K*), and the insufficient information content of observation datasets. These factors collectively result in non-uniqueness of parameter estimates. Previous studies that jointly estimate spatially distributed recharge and hydraulic parameters are limited to synthetic test cases and/or do not evaluate the effect of non-uniqueness. The extent to which recharge can be informed by calibration is largely unknown for practical situations, in which complexities such as parameter

heterogeneities are inherent. In this study, a systematic investigation of recharge, inferred through model calibration, is undertaken using a series of numerical experiments that include varying degrees of hydraulic parameter information. The analysis involves the use of a synthetic reality, based on a regional-scale, highly parameterised, steady-state groundwater model of Uley South Basin, South Australia. Parameter identifiability is assessed to evaluate the ability of parameters to be estimated uniquely. Results show that a reasonable inference of recharge (average recharge error <10%) requires a surprisingly large number of preferred value regularisation constraints (>100 *K* values across the 129 km^2 study area). The introduction of pumping data into the calibration reduces error in both the average recharge and its spatial variability, whereas submarine groundwater discharge (as a calibration target) reduces average recharge error only. Nonetheless, the estimation of steady-state recharge through inverse modelling may be impractical for real-world settings, limited by the need for unrealistic amounts of hydraulic parameter and groundwater level data. This study provides a useful benchmark for evaluating the extent to which field-scale groundwater models can be used to inform recharge subject to practical data-availability limitations.

3.2. Introduction

Reliable estimates of recharge and its variability in time and space are often a requirement for effective groundwater management. This is particularly the case where recharge estimates are used as a basis for allocating groundwater extraction rates (Werner et al., 2011). Traditionally, recharge is inferred from groundwater-level changes (e.g., Crosbie et al., 2005), solute concentrations (e.g., Wood and

Sanford, 1995), unsaturated zone numerical modelling (e.g., Keese et al., 2005) and/or remote sensing techniques (e.g., Jackson, 2002). The reliability of these approaches depends on many factors, including site characteristics (e.g., climate, vegetation, soil, topography, depth to watertable) and the availability and accuracy of field data (Scanlon et al., 2002). Estimates derived from field measurements also require consideration of the representative spatial and temporal scales (Scanlon et al., 2002). For these reasons, recharge is widely regarded as one of the most challenging water-balance components to quantify (Dripps and Bradbury, 2007).

An alternative method of recharge estimation is inverse modelling using numerical groundwater models. This involves the inference of recharge through calibration or "history matching" (i.e., minimising the discrepancy between field observations and corresponding model-generated outputs). The application of groundwater models in this context is appealing because of their ability to account for important nonlinear interactions between recharge, discharge, evapotranspiration and changes in groundwater storage (Sanford, 2002). Additionally, recharge values estimated from groundwater models typically apply to regional scales (e.g., >25 km^2), which are directly applicable to management decision making. Stoertz and Bradbury (1989) presented one of the first attempts at informing recharge by groundwater model calibration. They produced a map of recharge and discharge zones, and provided a range of spatially averaged net recharge rates for a basin in Wisconsin (USA). Groundwater model calibration has since been used frequently to inform recharge (e.g., D'Agnese et al., 1999; Essaid et al., 2003; Lubczynski and Gurwin, 2005; Palma and Bentley, 2007; Yidana, 2011; Hashemi et al., 2013). However, the

uncertainty associated with groundwater model-based recharge estimates is rarely quantified.

The primary challenge associated with recharge estimation through groundwater model calibration lies in the correlation between recharge and hydraulic properties such as hydraulic conductivity (*K*) and specific yield (Scanlon et al., 2002). Parameter correlation leads to non-uniqueness of estimated parameter values (Carrera and Neuman, 1986; McKenna et al., 2003). That is, multiple parameters can be varied in such a way that a model's outputs (such as the heads used in a calibration objective function) are unaffected. Non-uniqueness arises from the limited information content of observation datasets used as a basis for calibration. In the case where only groundwater-level observations are available, a unique solution for recharge and *K* cannot be obtained; only their ratio can be estimated uniquely (Scanlon and Cook, 2002). Efforts to overcome parameter non-uniqueness for the purposes of estimating recharge through calibration have involved several approaches (Sanford, 2002), including the use of flux observations (e.g., stream base flow; Arnold et al., 2000; Hunt et al., 2006) or seepage velocities (e.g., based on groundwater age; Portniaguine and Solomon, 1998; Sanford et al., 2004), either as prescribed stresses or as calibration targets.

An alternative means to address groundwater model non-uniqueness is by imparting expert knowledge of aquifer hydraulic properties, such as *K*, through regularisation (Carrera et al., 2005). Regularisation refers to the stabilisation of the parameter estimation process such that a unique solution to an otherwise ill-posed inverse problem can be obtained (Engl et al., 1996). For example, Tikhonov regularisation

schemes involve the specification of preferred parameter values and/or relationships based on expert knowledge to supplement the information contained within the calibration dataset. Then, any deviation in parameters from preferred values will occur only to the extent that is necessary to achieve a target level of model-to-measurement misfit (Tikhonov and Arsenin, 1977; Doherty and Skahill, 2006). Theoretically, a unique estimate of recharge may be achieved providing adequate constraints are imparted on *K* parameter values and relationships; however, the extent to which this is the case, for practical model calibration applications, is not presently understood. Aquifer hydraulic parameter distributions are often assumed to be homogeneous or piecewise homogeneous for the purposes of informing recharge (e.g., Sanford et al., 2004; Gómez et al., 2010; Cao et al., 2013). These approaches assume that any errors in recharge arising from the simplification of hydraulic parameter distributions are small (e.g., Ireson and Butler, 2013).

The uncertainty in various model input parameters, in particular recharge and *K*, has led to "simultaneous" or "joint" estimation of such parameters. For example, Hendricks Franssen et al. (2004) jointly estimated spatially variable transmissivity and recharge for a simple synthetic aquifer in their study of the impact of transmissivity and recharge uncertainty on well capture zones. They highlighted the importance of regularisation for obtaining reasonable parameter estimates. This was also demonstrated by Friedel (2005) in an investigation of the effect of variations in calibration dataset information on parameter non-uniqueness for a synthetic vadose zone model. Despite these efforts, Hendricks Franssen et al. (2009) subsequently asserted that very few studies estimate multiple parameter types simultaneously

within complex, regional-scale groundwater models, and that this constituted an important direction for future research.

A number of studies that jointly estimate recharge and aquifer hydraulic parameters for real-world settings have been presented. For example, Hendricks Franssen et al. (2008) used a stochastic calibration procedure to estimate spatial distributions of recharge and *K* in a groundwater model of the Chobe region, Botswana (Africa). They evaluated the value of remote sensing data and other data combinations by considering changes to ensemble recharge and *K* distribution statistics. However, their study did not investigate the influence of *K* constraints on calibrated estimates of recharge, given that stochastic methods are less likely to display parameter surrogacy (i.e., the process whereby parameters take on spurious values to compensate for other model errors; Clark and Vrugt, 2006; White et al., 2014). Hayley et al. (2014) estimated spatial recharge and *K* distributions simultaneously in their calibration of a highly parameterised groundwater model of north-eastern Alberta (Canada) using cloud computing. However, their calibrated recharge and *K* fields were evaluated separately in terms of reasonableness. That is, the effect of parameter non-uniqueness was not considered in their discussion of calibration outcomes. Carniato et al. (2015) jointly estimated the spatial distribution of recharge, *K* and specific yield as part of their development of a reactive transport model of a groundwater contamination field site in northern Belgium. They demonstrated that accurate predictions of plume concentrations and reduced uncertainty associated with recharge and *K* were obtained by incorporating both groundwater level and concentration data into the calibration. However, the degree to which their inverse

modelling outcomes were influenced by parameter non-uniqueness was not evaluated.

Recently, Erdal and Cirpka (2016) applied the Ensemble Kalman filter (Evensen, 1994) to jointly estimate spatially distributed recharge and *K* on the basis of head observations for a highly simplified synthetic aquifer. They showed that accurate values of recharge and *K* can be estimated simultaneously providing that a sufficient degree of prior information is available (i.e., the geostatistics of the true (initial ensemble) recharge- and *K*-fields are well-known). However, they discussed that their findings regarding the amount of prior information required for accurate recharge estimation may not be transferable to real-world applications, and recommended that more realistic cases be investigated in future studies to provide insights into the estimability of recharge and hydraulic parameters for practical situations.

Presently, the extent to which groundwater models can be used to estimate recharge through calibration, for complex, field-scale aquifer settings, is largely unknown. This is despite the generally accepted notion that groundwater models can be used to estimate recharge provided that groundwater level distributions and other model parameters such as *K* are known "well enough", as advocated by Sanford (2002).

The objective of this study is to evaluate the ability of groundwater model calibration to inform recharge in the face of non-uniqueness by using a systematic inverse modelling approach involving a regional-scale, highly parameterised groundwater model of Uley South Basin (USB), South Australia. We intentionally use the most

common approach to groundwater model calibration, namely the application of gradient methods such as those adopted in PEST (Doherty, 2016) and UCODE (Poeter and Hill, 1998), so that the findings offer guidance to groundwater practitioners. A quasi-hypothetical modelling strategy, whereby a reference model that serves as a synthetic reality for comparing subsequent model results, is adopted. The current assessment focuses on the spatial distribution of time-averaged recharge and basin-wide recharge volumes. This is achieved based on steady-state modelling, as is considered in many previous investigations of recharge (e.g., Gómez-Hernández and Gorelick, 1989; Zhang, 2014). The USB is chosen as a case study because of the well-developed understanding of the regional flow system based on extensive observation datasets, and the availability of an existing highly parameterised groundwater model developed by Knowling et al. (2015). Also, spatially distributed recharge rates are available from field-based estimates and one-dimensional unsaturated zone modelling (Ordens et al., 2012; Ordens, 2014).

3.3. Theoretical background

Two-dimensional Dupuit-Forcheimer flow within a heterogeneous aquifer under steady-state conditions is described by the diffusion equation:

$$\frac{\partial}{\partial x}\left(K_x b \frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y b \frac{\partial h}{\partial y}\right)+W=0 \tag{3.1}$$

where x and y [L] are Cartesian coordinates, h [L] is the hydraulic head, b [L] is the saturated aquifer thickness (which varies with h in unconfined aquifer systems), W [L/T] is the volumetric source/sink term per unit area that expresses the combined

effects of groundwater recharge, pumping, leakage, etc. (positive W values indicate inflow), and K_x and K_y [L/T] are the hydraulic conductivities in the x and y directions, respectively.

Equation (3.1) is solved in a piecewise manner by groundwater models that use numerical methods such as finite-difference or finite-element approximations (e.g., Harbaugh et al., 2000; Diersch, 2005). At the node scale, the computation of h is a function of the ratio of recharge-to-K (i.e., W/K, considering recharge as the only sink/source component for simplicity). It follows that a unique estimate of recharge and K may not be attainable based on h observations alone, and that only the ratio W/K may be determinable without additional information pertaining to K or groundwater fluxes (e.g., $Kb\frac{\partial h}{\partial x}$). Non-uniqueness at the scale of the model domain is complicated by such aspects as the spatial patterns in K, flow directions, boundary conditions, etc., and is therefore more challenging to characterise for practical situations, which are hampered by observation-data scarcity. The current study focuses on the estimability of recharge in the face of non-uniqueness for practical situations, in response to the knowledge gap identified by Hendricks Franssen et al. (2009).

3.4. Methodology

3.4.1. Study area

The USB (34º76' S, 135º56' E) is a topographically enclosed coastal groundwater basin of area ~129 km^2, and is located in the southern Eyre Peninsula, South

Australia. The region has a semi-arid climate, which is characterised by winter-dominant rainfall (May-October), and hot, dry summer months (November-April) (Harrington et al., 2006). Long-term average annual rainfall and pan evaporation rates are 560 and 1547 mm/y, respectively (Bureau of Meteorology, 2010). The land surface is composed mainly of skeletal soils of sandy and clayey loam (Evans, 1997) and exposed calcrete, where a widespread distribution of sinkholes serve to redistribute surface water into the unsaturated zone (Ordens et al., 2012). USB's vegetation consists primarily of various scrub species, Mallee trees, drooping she-oak and significant areas of sparse grassland (Bresciani et al., 2014). Groundwater in USB occurs predominantly within an unconfined Quaternary sand and limestone aquifer (QL), which is underlain by a discontinuous Tertiary clay aquitard (TC) and a semi-confined Tertiary aquifer comprising silty and clayey sand (TS) (Zulfic et al., 2007). Groundwater abstraction occurs from the QL aquifer, which constitutes the primary freshwater supply for the Eyre Peninsula region (Harrington et al. 2006). Historically, allowable extraction rates have been based on a fixed proportion of volumetric recharge estimates (EPNRM, 2006). Despite the karst nature of some of the USB sediments, the aquifer behaves mostly as a single-porosity (i.e., equivalent porous medium) setting, at least at scales of interest to water resources management and regional-scale modelling (Werner, 2014).

3.4.2. Groundwater model

The groundwater model of USB used as a basis for the current study is described in detail by Knowling et al. (2015). MODFLOW (Harbaugh et al., 2000) is used to simulate groundwater flow within USB under steady-state conditions. The

assumption of steady-state conditions is considered a reasonable approximation of the average groundwater levels during two distinct time intervals: 1960–1976 (i.e., pre-development) and 2005–2009 (i.e., a period of relatively stable water levels). The pre-development steady-state condition is adopted in the reference USB model. The contemporary condition (2005-2009), which involves significant groundwater extraction, serves as a basis for developing an alternative steady-state model. This model is used to explore whether or not the addition of prescribed fluxes (i.e., groundwater extraction) assists in constraining the estimation of recharge. Also, considering two alternative steady-state models allows for an evaluation of whether or not the estimability of recharge is specific to a particular steady-state condition. Table 3.1 lists the main features of the reference USB model. Figure 3.1 shows a schematic of the reference model and its parameter distributions.

Table 3.1. Summary of reference model characteristics.

Model characteristic	**Description**
Layer(s) type	Unconfined upper layer (QL aquifer); Convertible confined/unconfined (i.e., depending on hydraulic head relative to layer top elevation) lower layer (TS aquifer); TC aquitard simulated implicitly, i.e., a "quasi-3D approach" (Chiang and Kinzelbach, 1998).
Spatial discretisation	100 m × 100 m model-grid cells (245 rows, 132 columns; Figure 3.1a).
K	Estimated through calibration (Figure 3.1b and c) and constrained by field data (e.g., pumping test results) or literature values. Vertical *K* assumed to be 0.1 times horizontal *K*.
Ocean boundary	Constant-head boundary using density corrections following Morgan et al. (2012).
Interaction with adjacent basins	General-head boundaries (GHB; Harbaugh et al., 2000) used. GHB forcing heads are based on time-averaged groundwater levels in adjacent basins. GHB conductances (*C*) estimated through calibration (Figure 3.1d), and constrained by knowledge of *K* and aquifer thickness.
Spatially distributed recharge	Applied to uppermost active layer. Recharge distributions based on the modelling by Ordens (2014) (e.g., Figure 3.1e). Spatially averaged recharge for pre-development conditions is 100 mm/y, and for contemporary conditions is 69 mm/y.
Evapotranspiration (ET)	ET package (EVT; Harbaugh et al., 2000) simulates shallow groundwater ET, whereas unsaturated zone ET is accounted for in the recharge modelling of Ordens (2014).
Groundwater pumping	Pumping rates at each well based on metered data and extrapolation (where metered data are unavailable) (Werner et al., 2011). No pumping occurs during pre-development conditions. Total pumping rate for contemporary conditions is 6.98×10^6 m^3/y.

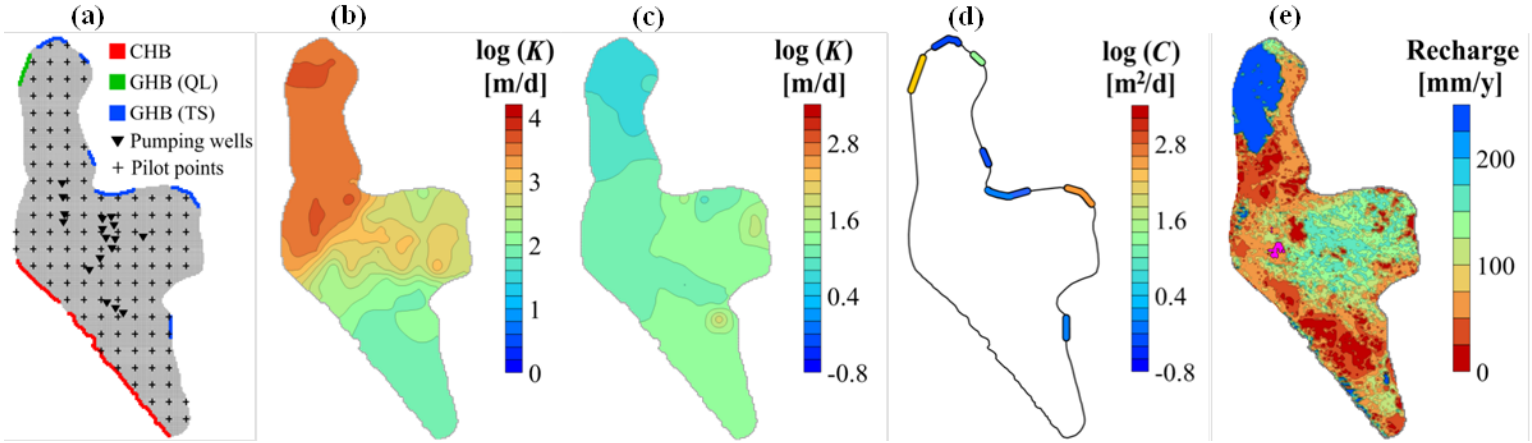

Figure 3.1. Reference model characteristics: (a) model domain, grid, boundary conditions ("CHB" denotes constant-head boundary, and "GHB" denotes general-head boundary), extraction well and pilot point locations; (b) log(*K*) [m/d] of the QL aquifer; (c) log(*K*) [m/d] of the TS aquifer; (d) log(*C*) [m^2/d] for each GHB reach, and (e) spatially distributed recharge rates [mm/y] (temporally averaged over the pre-development period 1960–1976) from Ordens (2014).

The spatial variability of *K* is achieved using 125 pilot points (de Marsily et al., 1984) distributed in a regular grid configuration (Figure 3.1a). Cell-by-cell *K* variability is specified by kriging based on an isotropic, exponential variogram of log(*K*), with a sill of one, a range of 1200 m, and a nugget of zero. A single boundary conductance (*C*) is assigned to each GHB reach (Figure 3.1a). It should be noted that the reference model *K* and *C* distributions (Figure 3.1b, c and d) differ slightly from those obtained by Knowling et al. (2015). They estimated *K* and *C* parameters through joint calibration of the pre-development and contemporary steady-state models, whereas reference model *K* and *C* distributions are estimated on the basis of only the pre-development model in the current study.

3.4.3. Model calibration

Estimation of groundwater model parameters is undertaken using PEST (Doherty, 2016). BeoPEST (Schreüder, 2009), a version of PEST that allows model-run parallelisation, is used to reduce computational times. PEST adopts a gradient-based algorithm (i.e., the Gauss-Marquardt-Levenberg (GML) method; Levenberg, 1944) to minimise a weighted least-squares objective function, Φ. For parameter estimation problems involving Tikhonov regularisation, Φ is given by:

$$\Phi = (\mathbf{Xp} - \mathbf{h})^{t}\mathbf{Q}_{m}(\mathbf{Xp} - \mathbf{h}) + \mu(\mathbf{Zp} - \mathbf{d})^{t}\mathbf{Q}_{r}(\mathbf{Zp} - \mathbf{d}) \quad (3.2)$$

where **p** is a vector containing model parameter values, **h** is a vector containing measurement observation values, **X** is a sensitivity (or "Jacobian") matrix containing derivatives of model outputs (for which there are corresponding measurement observations **h**) with respect to **p**, **d** is a vector containing Tikhonov regularisation

"observations" (representing preferred parameter values and/or relationships), **Z** is a Jacobian matrix containing derivatives of model-generated counterparts to regularisation observations **d** with respect to **p**, μ is a regularisation weight factor (discussed below), and $\mathbf{Q}_m$ and $\mathbf{Q}_r$ are matrices whose diagonal elements are the square of the weights assigned to corresponding measurement and regularisation observations, respectively. The value of μ controls the trade-off between fitting observation data and the incorporation of regularisation constraints, explained in more detail below.

Traditional calibration approaches, which involve the inference of a small number of parameters, minimise the measurement objective function, Φ_m, which is the term $(\mathbf{Xp}-\mathbf{h})^t\mathbf{Q}_m(\mathbf{Xp}-\mathbf{h})$ in equation (3.2). Where the inverse problem is "well-posed" or "over-determined", minimisation of Φ_m yields a unique parameter set. The regularisation objective function, Φ_r, given by the term $(\mathbf{Zp}-\mathbf{d})^t\mathbf{Q}_r(\mathbf{Zp}-\mathbf{d})$ in equation (3.2), encapsulates the deviation of parameter values from those representing a preferred condition. A common application of Tikhonov regularisation is to include the difference between neighbouring pilot-point parameter values in **d**. By specifying the corresponding elements of **d** equal to zero, a preferred homogeneity constraint is enforced (Doherty, 2003). The desired level of model-to-measurement misfit (Φ_m) is user-specified as the "target measurement objective function" Φ_m^l, which is obtained by adjusting μ. Φ_r is thereby minimised under the constraint that:

$$\Phi_m \leq \Phi_m^l \tag{3.3}$$

This process constitutes a constrained minimisation problem (e.g., deGroote-Hedlin and Constable, 1990), whereby Φ_m^l provides a user-specified basis for controlling the

degree to which the problem is regularised. This term is prescribed in accordance with the uncertainty associated with observation measurements. The reader is referred to Fienen et al. (2009) for a detailed discussion on $\Phi_m^{\,l}$ and other PEST-specific regularisation control variables.

For regularised (non-linear) model calibration problems, a parameter upgrade vector, $\Delta\mathbf{p}$, as calculated iteratively using the GML procedure, is given by (Doherty and Skahill, 2006):

$$\Delta\mathbf{p} = (\mathbf{X}^t\mathbf{Q}_m\mathbf{X} + \mu\mathbf{Z}^t\mathbf{Q}_r\mathbf{Z} + \lambda\mathbf{I})^{-1}(\mathbf{X}^t\mathbf{Q}_m\mathbf{r} + \mu\mathbf{Z}^t\mathbf{Q}_r\mathbf{s}) \tag{3.4}$$

where λ is the Marquardt lambda parameter, **I** is the identity matrix, **r** is a vector of the model-to-measurement discrepancies (**Xp**−**h**), and **s** is a vector of the regularisation discrepancies (**Zp**−**d**). The λ term is often increased such that $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$ becomes invertible to prevent instability in the calculation of the parameter upgrade. Despite the ability of high λ values to impart stability, robust parameter estimation requires some form of regularisation (Tonkin and Doherty, 2005). The reader is referred to Doherty and Hunt (2010) and references cited therein for a more detailed description of pertinent calibration methodologies.

In this study, the sensitivities of model outputs to parameter values within the **X** and **Z** matrices are calculated using central finite differences with a 1% perturbation of current parameter values following, e.g., James et al. (2009). The maximum parameter value upgrade factor between optimisation iterations is 20% of current values. The specification of $\Phi_m^{\,l}$ is dependent on initial modelling results. Weighting of different regularisation observation groups is achieved on the basis of total

composite observation-group sensitivities, which is the sum of the (weighted) sensitivities of all observations within a particular group to all adjustable parameters (Hill, 1998; 2010). Individual regularisation observations (i.e., within a group) are weighted uniformly. The initial λ value and the factor by which λ is adjusted during parameter upgrade testing is set to 10 and -3, respectively, as suggested by Doherty (2016).

3.4.4. Parameter identifiability

Parameters are evaluated in terms of their identifiability (i.e., their ability to be estimated uniquely through calibration). The set of parameter values obtained through model calibration $\underline{\mathbf{p}}$ is represented by (Doherty et al., 2010):

$$\underline{\mathbf{p}} = \mathbf{Gh} \tag{3.5}$$

where **G** is a matrix (sometimes referred to as the "parameter solution matrix") expressing the means by which the estimated parameter values are obtained from measurement observations **h**. The formulation of **G** is therefore dependent on the regularisation method employed within the calibration process. In this study, **G** is given by the following (Doherty, 2016):

$$\mathbf{G} = (\mathbf{X}^t\mathbf{Q}_m\mathbf{X} + \mu\mathbf{Z}^t\mathbf{Q}_r\mathbf{Z} + \lambda\mathbf{I})^{-1}\mathbf{X}^t\mathbf{Q}_m \tag{3.6}$$

Equation (3.6) applies to the situation of calibrated parameter values, and assumes that **r** and **s** are sufficiently small that they can be ignored. The relationship between

a set of measurement observations **h** and a set of true (i.e., real-world) system parameters **p′** is represented by (Doherty et al., 2010):

$$\mathbf{h} = \mathbf{X}\mathbf{p}' + \boldsymbol{\varepsilon} \tag{3.7}$$

where **ε** is a vector expressing the measurement noise associated with measurements **h**. This assumes that the (linear) model **X** is free from structural defects. Substitution of (7) into (5) yields (Moore and Doherty, 2006):

$$\underline{\mathbf{p}} = \mathbf{G}\mathbf{X}\mathbf{p}' + \mathbf{G}\boldsymbol{\varepsilon} = \mathbf{R}\mathbf{p}' + \mathbf{G}\boldsymbol{\varepsilon} \tag{3.8}$$

where **R** (= **GX**) is a resolution matrix. Where measurement noise is small enough such that it can be neglected (i.e., the **Gε** term on the right-hand side of (8) equals zero), the **R** matrix describes the relationship between estimated parameter values $\underline{\mathbf{p}}$ and their real-world counterparts **p′**. Each row of **R** contains the weights by which real-world parameter values are multiplied to produce a particular estimated parameter (commonly referred to as "integration kernel"; Vasco et al., 1997; Paradis et al., 2015). Note that $\underline{\mathbf{p}}$ and **p′** are equivalent if **R** equals **I** (and **ε** equals zero). The identifiability of parameter *i* (d_i) is defined generally as the ability of model calibration to constrain parameter *i*, and is given by (Doherty and Hunt, 2009):

$$d_i = [\mathbf{R}]_{i,i} \tag{3.9}$$

where subscript *i*,*i* designates the *i*th diagonal element of **R**. The value of d_i can vary between zero (indicating that the parameter in question is completely unidentifiable) and one (indicating that the parameter in question is completely identifiable). Note that a d_i value of one does not necessarily mean that its estimated value will be error-

free, due to the effects of measurement uncertainty. However, statistics such as parameter error variance (e.g., Moore and Doherty, 2005) are not reported in this current study because measurement uncertainty is assumed to be negligible (given the model-generated nature of the synthetic calibration dataset). To explore the influence of model non-linearity, two Jacobian matrices (which are required for the calculation of **R** and **G**) are considered based on both initial and calibrated parameter values.

3.4.5. Inverse modelling cases

Heads generated on the basis of the reference model are used as "observations" for the calibration dataset in subsequent calibration cases. Heads are taken at pilot point locations where the QL and TS sediments are saturated, resulting in a total of 217 head observations derived from the reference model. The use of coincident observations and estimable parameters facilitates the comparison of reference model- and calibration-based recharge and *K* distributions at discrete locations (i.e., at pilot points), without the complication of kriging effects (i.e., intra-pilot point recharge and *K* variability). Head observations are weighted uniformly.

Parameter value constraints in the form of upper and lower bounds, and initial values are given in Table 3.2. Constraints on hydraulic parameters are consistent with those used for the reference model. *K* and *C* parameters are log-transformed within the calibration process. The influence of initial *K* and *C* values on estimated values is explored using an alternative set of initial values equal to reference-model parameters.

Table 3.2. Parameterisation set-up for inverse modelling cases. [a]Alternative initial recharge rates of 80 and 120 mm/y, and initial hydraulic parameter values from the reference model are used to assess the influence of initial parameter values; [b]A recharge rate of 69 mm/y is used for the contemporary steady-state model; [c]Acronyms relate to the linkages through which USB is connected to adjacent basins (see Knowling et al. (2015) for more details): "UE" refers to Uley East, "UW" refers to Uley Wanilla, "CB" refers to Coffin Bay, and "SE" refers to south-east.

Model parameters/stresses	**Parameterisation method (number of parameters)**	**Bounds**	**Initial values[a]**
Recharge [mm/y]	Pilot points (125)	0–300	100[b]
K [m/d]			
QL	Pilot points (125)	$1–10^4$	100
TS	Pilot points (125)	0.132–1320	13.2
C [m^2/d][c]			
UE (TS)	Single value	0.36–3600	36
UW1 (TS)	Single value	0.054–540	5.4
UW2 (TS)	Single value	0.059–590	5.9
UW3 (TS)	Single value	0.23–2300	23
CB1 (TS)	Single value	0.29–2900	29
CB2 (TS)	Single value	0.26–2600	26
CB3 (QL)	Single value	0.18–1800	18
SE (TS)	Single value	0.07–770	7.7

The spatial variability of recharge is achieved using pilot point parameters co-located with those used for the estimation of *K*. An isotropic variogram of exponential type is used to produce cell-by-cell recharge variability, as is the case for *K* variability.

A spatially uniform recharge rate of 100 mm/y, which is equal to the spatially averaged (and temporally averaged over the pre-development period) value of Ordens (2014), is prescribed as an initial value (i.e., prior to calibration). This allows the degree to which the spatial distribution of recharge can be inferred through calibration to be investigated systematically. Spatially uniform recharge rates of 80 and 120 mm/y are also considered to explore the effect of initial recharge values. For the contemporary steady-state model, a spatially uniform recharge rate of 69 mm/y is

used, which is equal to the spatially averaged value of Ordens (2014) during the contemporary period.

Because the reference-model recharge distribution varies on a cell-by-cell basis compared to calibrated recharge distributions, which are based on kriging of pilot-point values, unavoidable discrepancies exist between these distributions. For the purpose of generating a reference recharge distribution that is directly comparable to those obtained through calibration, a smoothed representation of the reference recharge distribution is produced by kriging pilot-point values (using a consistent variogram) in such a way that the difference between the original reference distribution and its kriged variant are minimised in a least-squares sense. The reader is referred to Moore (2006) for a detailed description of this approach. Figure 3.2 shows the smoothed representation of the reference recharge distribution.

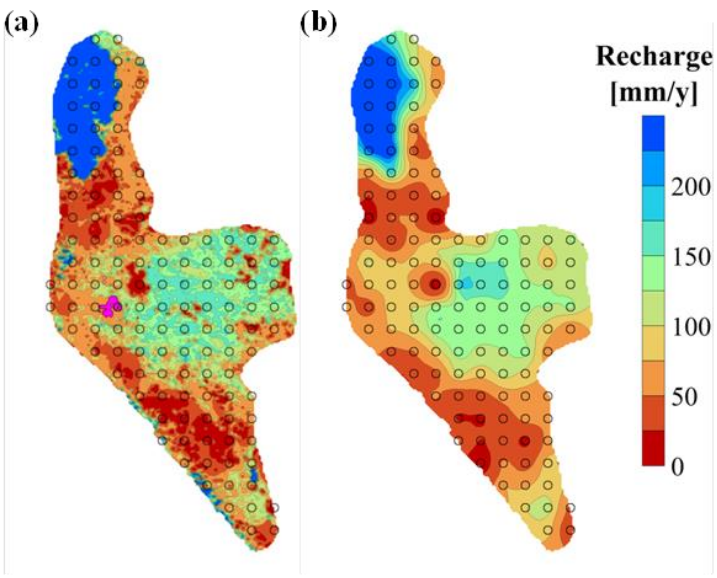

Figure 3.2. (a) Reference-model recharge distribution, and (b) its counterpart obtained from kriging pilot-point values that are calculated such that cell-by-cell recharge differences with (a) are minimised. The circles depict pilot point locations.

Three inverse modelling cases are considered initially, as summarised in Table 3.3. The first case ("Case 1") involves the estimation of recharge subject to fixed hydraulic parameter values from the reference model. Regularisation is not used for Case 1, and hence, the calibration process involves the minimisation of Φ_m only. Case 1 represents a best-case scenario given that all *K* and *C* parameter values are known. Although Case 1 constitutes an unrealistic situation in practice (i.e., an excessive degree of parameter knowledge), it provides a framework for demonstrative and comparative purposes.

Table 3.3. Inverse modelling cases.

ID	Estimable parameters	Description	Hydraulic parameter information	Regularisation used?
Case 1	Recharge	Perfect knowledge of hydraulic parameters	*K* and *C* fixed as per reference model	No
Case 2	Recharge, *K*, *C*	No prior knowledge of hydraulic parameters	Parameter bounds only	No
Case 3a to 3k	Recharge, *K*, *C*	Varying degree of knowledge regarding hydraulic parameters	Preferred *K* and *C* values (from reference model); preferred *K*-homogeneity	Yes

Case 2 involves the estimation of both recharge and hydraulic parameters in the absence of regularisation constraints (Table 3.3). Case 2 represents a worst-case scenario, whereby no expert knowledge regarding *K* and *C* is available. Because the Case 2 calibration involves correlated parameters that are largely unconstrained, it is expected that considerable parameter compensation will occur. Nevertheless, this case serves as a useful demonstration and for comparing subsequent cases.

In Case 3, the estimation of recharge and hydraulic parameters is subject to varying degrees of hydraulic parameter information, imparted by regularisation. The Case 3 calibration therefore involves the minimisation of both Φ_m and Φ_r. The Φ_m^l is specified such that a relatively comparable level of model-to-reference model misfit is obtained for all cases. This is achieved by prescribing a value for Φ_m^l that slightly exceeds Φ_m obtained for Cases 1 and 2 given that regularisation is expected to produce higher Φ_m values (due to the minimisation of both Φ_m and Φ_r).

Case 3 includes eleven variants (Case 3a to 3k), which vary according to the number of Tikhonov regularisation constraints imposed. The first five cases (3a to 3e) employ preferred *K* and *C* values, which accord with the reference model. The number of preferred values is based on different percentages of the total number of

calibrated parameters (i.e., 258 parameters: 125 QL *K*, 125 TS *K*, and eight *C*), as follows: 10% (25 *K* and one *C*); 25% (62 *K* and two *C*); 50% (125 *K* and four *C*); 75% (188 *K* and six *C*); and 90% (225 *K* and seven *C*). Figure 3.3 shows the location of preferred *K* and *C* values for each sub-case. Uniform weights are assigned to all preferred parameter values. Cases 3f to 3j are the same as Cases 3a to 3e except with the additional constraint of preferred homogeneity in *K*. The use of preferred *K*-homogeneity is considered reasonable here given that the generation of reference-model *K* distributions involved preferred *K*-homogeneity. An additional case (3k) is considered whereby only preferred *K*-homogeneity is employed (i.e., no preferred values). Weights on preferred homogeneity constraints are also uniform, and are equivalent to those assigned to preferred values.

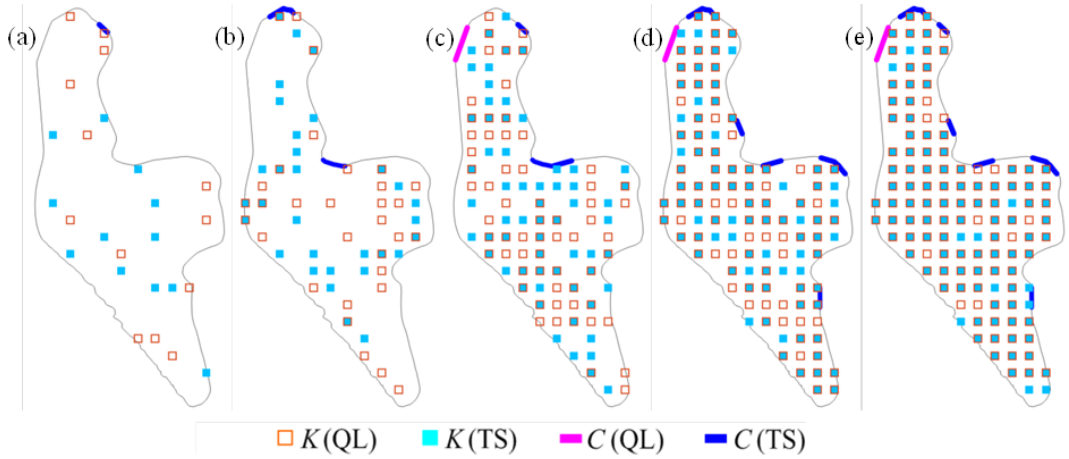

Figure 3.3. Distribution of preferred *K* and *C* values used for Cases 3a to 3e (and for Cases 3f to 3j). The number of preferred values is based on the following percentages of the total number of adjustable hydraulic parameters: (a) 10%; (b) 25%; (c) 50%; (d) 75%; and (e) 90%.

Each of the cases described above are re-evaluated to consider an observation of total submarine groundwater discharge (SGD) to assess whether or not the inclusion of a groundwater-flux calibration target allows for an improved estimation of recharge.

The "observed" SGD (2.698×10^7 m^3/y) is obtained from the reference model. A relative observation weight of 0.01 is prescribed to account for differences between flux and head units. All cases are also re-evaluated in terms of recharge estimability for the contemporary model (which includes pumping), and using different initial recharge and hydraulic parameter values for the pre-development model.

3.5. Results

3.5.1. Case 1

Table 3.4 lists model-to-reference model misfit statistics (including values of Φ_m and the corresponding calibration root-mean-squared-error (RMSE)) for all inverse modelling cases. The level of fit obtained for Case 1 (Φ_m of 1.36 m^2, RMSE of 0.008 m) is considered adequate.

Table 3.4. Model-to-reference model misfit statistics for all inverse modelling cases. [a]Contribution to Φ_m from the model-to-reference model head misfit only; [b]RMSE for the model-to-reference model head misfit and flux misfit, respectively; and [c]Level of fit considered to be inadequate for the purpose of comparing recharge estimations.

		Additional inverse modelling cases				
	Original	**Use of SGD calibration target**	**Contemporary model**	**Initial recharge of 80 mm/y**	**Initial recharge of 120 mm/y**	**Initial *K*, *C* from reference model**
	Φ_m [m^2] (RMSE [m])	**Φ_m [m^2][a] (RMSE [m], [m^3/d])[b]**	**Φ_m [m^2] (RMSE [m])**	**Φ_m [m^2] (RMSE [m])**	**Φ_m [m^2] (RMSE [m])**	**Φ_m [m^2] (RMSE [m])**
Case 1	**1.36 (0.008)**	1.73 (0.009, 0.9)	4.2 (0.014)	0.12 (0.002)	0.11 (0.002)	-
Case 2	**0.94 (0.007)**	23 (0.032, 0.6)	0.58 (0.005)	6.2 (0.017)	0.68 (0.006)	0.12 (0.002)
Case 3a	**136 (0.079)[c]**	15 (0.026, 5.0)	262 (0.110)[c]	2.2 (0.010)	102 (0.068)[c]	1.4 (0.008)
Case 3b	**1.9 (0.009)**	5.8 (0.016, 1.4)	2.2 (0.010)	2.2 (0.010)	2.4 (0.010)	0.58 (0.005)
Case 3c	**2.3 (0.010)**	9.3 (0.021, 0.3)	3.9 (0.013)	2.2 (0.010)	2.2 (0.010)	2.2 (0.010)
Case 3d	**2.2 (0.010)**	2.2 (0.010, 0.1)	6.1 (0.017)	4.5 (0.014)	2.3 (0.010)	2.0 (0.009)
Case 3e	**2.2 (0.010)**	1.9 (0.009, 1.4)	3.6 (0.013)	2.2 (0.010)	2.1 (0.010)	1.8 (0.009)
Case 3f	**2.3 (0.010)**	2.5 (0.011, 0.1)	4.7 (0.015)	2.3 (0.010)	2.4 (0.010)	2.3 (0.010)
Case 3g	**2.4 (0.011)**	3.2 (0.012, 1.2)	2.8 (0.012)	2.3 (0.010)	3.3 (0.012)	2.3 (0.010)
Case 3h	**3.3 (0.012)**	7711 (0.596, 31)[c]	112 (0.072)[c]	420 (0.139)[c]	510 (0.153)[c]	3.1 (0.012)
Case 3i	**5.9 (0.016)**	9.0 (0.020, 0.1)	2.6 (0.011)	280 (0.113)[c]	9.1 (0.021)	2.7 (0.011)
Case 3j	**2.2 (0.010)**	2.3 (0.010, 0.4)	156 (0.085)[c]	2.2 (0.010)	2.2 (0.010)	2.5 (0.011)
Case 3k	**285 (0.115)[c]**	13,804 (0.799, 100)[c]	5047 (0.482)[c]	7045 (0.570)[c]	13,958 (0.802)[c]	43 (0.045)[c]

Figure 3.4 shows the calibrated recharge ($\underline{R}$) values, and their comparison to the kriged reference model recharge (R'_{krig}) values, for Case 1. A reasonable correlation is evident between pilot point $\underline{R}$ and R'_{krig} values (Figure 3.4a), as evidenced by a regression line-of-best-fit slope and coefficient of determination (r^2) equal to 0.96 and 0.66, respectively. The regression-line slope approaching unity indicates reasonable agreement between $\underline{R}$ and R'_{krig}, on average, whereas the mid-range r^2 value suggests a moderate degree of scatter. The error in $\underline{R}$ (i.e., $\underline{R}$–R'_{krig}) at each pilot point varies from -139 to 171 mm/y, with an average error and average absolute error of 11 mm/y (model overestimation) and 34 mm/y, respectively. The volumetric $\underline{R}$ error percentage is -1.1% (a small model underestimation). The inconsistency between the average pilot point $\underline{R}$ error (overestimation) and volumetric $\underline{R}$ error (underestimation) is a consequence of kriging, as evidenced by the average $\underline{R}$ error obtained on a model cell-by-cell basis (-0.02 mm/y).

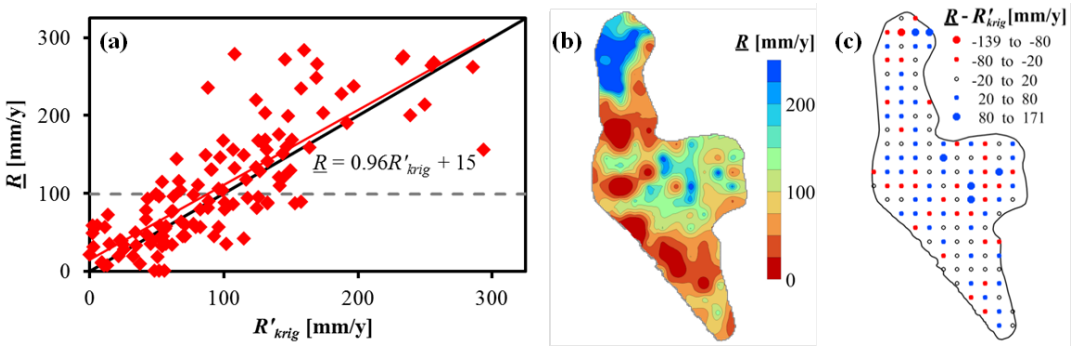

Figure 3.4. (a) $\underline{R}$-versus-R'_{krig} scatterplot (the black line represents the 1:1 line, the red line represents the regression line-of-best-fit, as given by the equation shown, and the dashed line represents the initial uniform $\underline{R}$ value of 100 mm/y), (b) $\underline{R}$ distribution [mm/y], (c) $\underline{R}$ error distribution [mm/y], for Case 1.

The reasonable agreement between $\underline{R}$ and R'_{krig} is further illustrated by the similarity of their distributions (compare Figures 3.4b to 3.2b). The $\underline{R}$ field largely resembles

the R'_{krig} distribution. The spatial distribution of $\underline{R}$ errors (Figure 3.4c) displays complicated patterns that are not attributable to physical factors related to recharge.

Figure 3.5 shows the identifiability of recharge parameters employed in Case 1. Identifiability values are highly variable (ranging between zero and one). Just 23 of the 125 parameters are considered to be identifiable on the basis of the calibration dataset, indicated by identifiability values of greater than 0.8. Of the remaining parameters, 44 display mid-range values (0.2-0.8), and 58 display low values (<0.2).

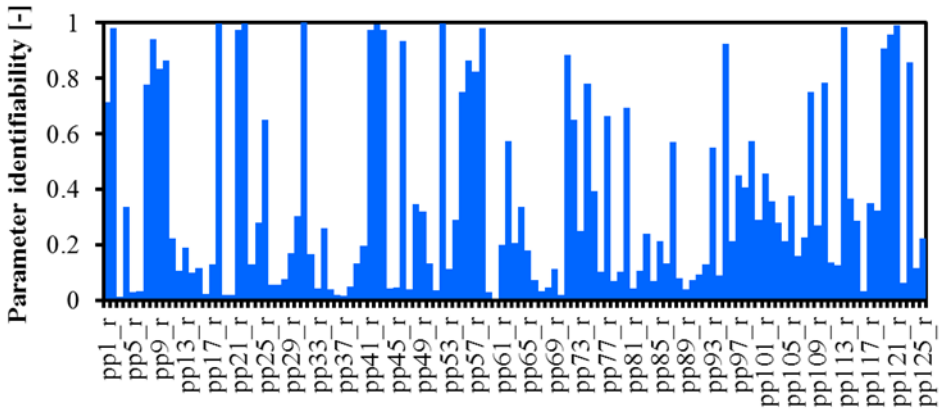

Figure 3.5. Recharge parameter identifiability for Case 1. The x-axis labels are names assigned to the 125 pilot points.

3.5.2. Case 2

The model-to-reference model misfit of Case 2 (Φ_m of 0.94 m^2, RMSE of 0.007 m) is smaller than that of Case 1. This occurs because, in the absence of regularisation, the calibration has greater freedom to modify parameters relative to Case 1. Case 2 represents the best fit to head observations relative to other cases that exclude flux data and that adopt an initial recharge value of 100 mm/y (Table 3.4).

Figure 3.6 shows $\underline{R}$ values, and their comparison to R'_{krig} values, for Case 2. A correlation between $\underline{R}$ and R'_{krig} is lacking, as evidenced by the regression-line slope of 0.06 (Figure 3.6a). The majority of $\underline{R}$ values are only marginally modified by calibration (67% of values lie between 80 and 120 mm/y). The average error and average absolute error of $\underline{R}$ at each pilot point is -14 mm/yr (model underestimation) and 54 mm/y, respectively. The volumetric $\underline{R}$ error percentage is -15%, which is consistent with the average cell-by-cell-based error (-15 mm/y).

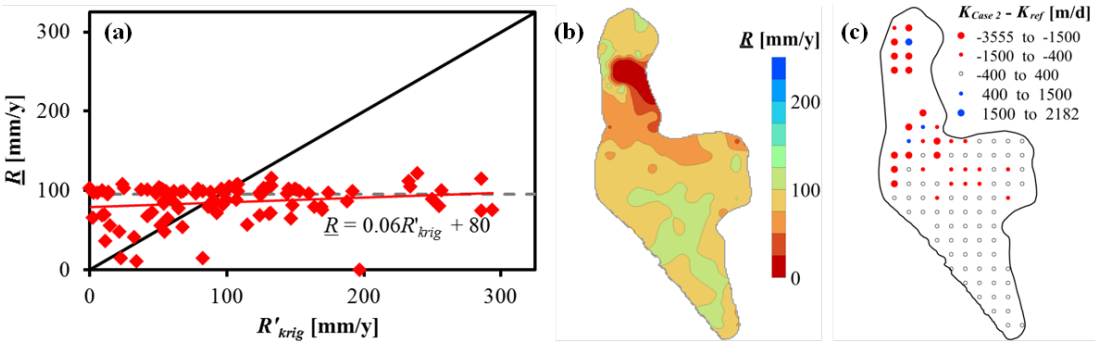

Figure 3.6. (a) $\underline{R}$-versus-R'_{krig} scatterplot, (b) $\underline{R}$ distribution [mm/y], and (c) K (QL) error distribution [m/d], for Case 2. In (c), the K-error is not displayed at some pilot points because the simulated head in the QL aquifer is below the cell-bottom elevation.

The $\underline{R}$ distribution (Figure 3.6b) does not display consistent patterns compared to those of R'_{krig} (Figure 3.2b). For example, the high recharge rates associated with the north-west (>200 mm/y) and central (>150 mm/y) regions of USB are not reflected in the $\underline{R}$ distribution. These inconsistencies are a result of the compensatory roles assumed by hydraulic parameters, i.e., spurious hydraulic parameter values are obtained to compensate for errors in $\underline{R}$ to maintain an equivalent level of model-to-reference model fit. For example, in the aforementioned areas where $\underline{R}$ is less than R'_{krig}, calibrated QL K values are lower than those from the reference model, as

evidenced by negative *K* (QL) errors (Figure 3.6c). The average *R* error of -15 mm/y is compensated by an average pilot-point *K* (QL) error of -472 m/d.

Figure 3.7 shows the identifiability of hydraulic and recharge parameters employed in Case 2. While many of the hydraulic parameters display mid-range identifiability values (Figure 3.7a, b and c), only seven of the 270 hydraulic parameters are deemed identifiable (indicated by values of >0.8). The identifiability of all but three recharge parameters (which display values of 1.0, 0.64, 0.52) is low (<0.2), with the majority of parameters displaying values of approximately zero (Figure 3.7d).

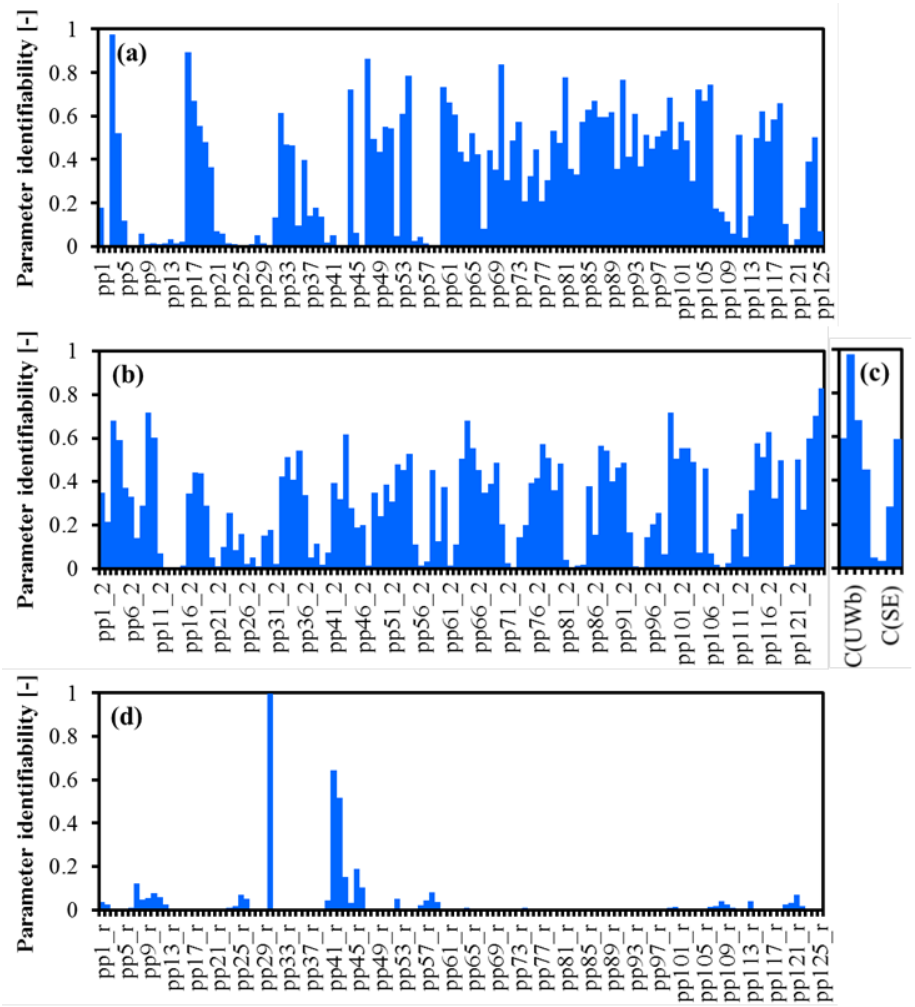

Figure 3.7. Identifiability of (a) QL *K*, (b) TS *K*, (c) *C*, and (d) recharge parameters for Case 2. The x-axis labels are names assigned to the pilot point parameters.

3.5.3. Case 3

A $\Phi_m^{\ l}$ value of 2.17 m^2 is specified for Case 3 such that the corresponding calibration RMSE value (0.01 m) slightly exceeds that obtained for Cases 1 and 2. The model-to-reference model misfit for seven of the eleven cases (3b, c, d, e, f, g and j) is within ~15% of the target (Φ_m values < 2.50 m^2, RMSE values <0.0106 m) (Table 3.4), therefore allowing for reliable comparisons of recharge estimates to be made between the Case 3 variants. The levels of misfit for Cases 3h and 3i are higher (Φ_m values of 3.3 and 5.9 m^2, RMSE values of 0.012 and 0.016 m, respectively), but nonetheless are considered sufficient for comparative purposes. Acceptable calibration results, in terms of Φ_m, could not be obtained for Cases 3a and 3k. Higher Φ_m values are attributable to the role of regularisation within the calibration process, whereby Φ_m is penalised by the requirement to minimise Φ_r.

Table 3.5 lists average deviations from preferred *K* and *C* values, and average differences between neighbouring pilot-point *K* values. Generally, the deviation from preferred hydraulic parameter values for Cases 3a to 3g is considered insignificant. For Cases 3f to 3i, the extent to which estimated hydraulic parameter values deviate from preferred values generally increases due to the introduction of preferred homogeneity. This occurs because the preferred smoothness counteracts the ability of inverse modelling to reproduce the preferred values of *K*. The average difference between neighbouring pilot-point *K* values for Cases 3f to 3k is considered to be relatively small.

Table 3.5. Average deviation from preferred *K* values [m/d] and *C* values [m^2/d], and average difference between neighbouring pilot-point *K* values (for cases involving preferred homogeneity only; denoted "*K*-u" below) [m/d] for inverse modelling cases involving regularisation (Case 3a to 3k).

				Additional inverse modelling cases													
	Original			**Use of SGD calibration target**			**Contemporary model**			**Initial recharge of 80 mm/y**			**Initial recharge of 120 mm/y**			**Initial *K*, *C* from reference model**		
	K	***C***	***K*-u**	***K***	***C***	***K*-u**	***K***	***C***	***K*-u**	***K***	***C***	***K*-u**	***K***	***C***	***K*-u**	***K***	***C***	***K*-u**
Case 3a	**0.73**	**0.008**	**-**	31	1.2	-	1.4	0.001	-	0.003	0	-	41	0.76	-	0.004	0	-
Case 3b	**0.003**	**0**	**-**	83	0.003	-	0.001	0	-	0	0	-	0.007	0	-	0	0	-
Case 3c	**0.004**	**0**	**-**	133	4.2	-	16	0.71	-	0	0	-	0	0	-	0.002	0	-
Case 3d	**0.068**	**0.001**	**-**	72	0.66	-	3.9	0.42	-	32	0.26	-	0.014	0	-	0.004	0	-
Case 3e	**0.044**	**0**	**-**	86	1.1	-	1.3	0.03	-	0.005	0	-	0.003	0	-	0	0	-
Case 3f	**0.12**	**0**	**1.8**	0.19	0.001	1.9	0.74	0	4.6	0.14	0	1.8	0.12	0	1.8	0.042	0	1.8
Case 3g	**0.30**	**0**	**1.7**	12	0	1.8	1.4	0.01	5.3	0.31	0	1.7	0.34	0	1.7	0.22	0	1.7
Case 3h	**2.0**	**0.008**	**1.7**	2.3	0	1.5	6.6	0.27	3.9	2.2	0.002	1.6	3.2	0.002	1.6	1.3	0.001	1.7
Case 3i	**7.5**	**0.21**	**1.7**	12	0.22	1.8	13	0.03	3.0	10	0.56	1.2	19	0.80	1.2	4.0	0.009	1.2
Case 3j	**6.1**	**0.005**	**1.7**	62	0.44	1.8	18	0.28	1.8	6.0	0.002	1.7	6.2	0.006	1.7	6.5	0.004	1.7
Case 3k	**-**	**-**	**1.4**	-	-	1.1	-	-	1.0	-	-	1.1	-	-	1.2	-	-	1.6

Figure 3.8 shows the $\underline{R}$ values, in comparison to R'_{krig} values, for Case 3a, c, e, f, h, j and k (other cases are omitted for brevity). The degree to which $\underline{R}$ and R'_{krig} are correlated increases from Case 3a to 3e and from Case 3f to 3j, i.e., with the number of preferred value constraints, as evidenced by the regression-line slopes given in Figure 3.8. A higher variability in $\underline{R}$ values is apparent for Cases 3f, h and j compared to 3a, c and e, as indicated by larger coefficients of variation (*CV*; i.e., ratio of standard deviation to mean) values (0.21, 0.44, and 0.58 compared to 0.11, 0.22, and 0.38, respectively).

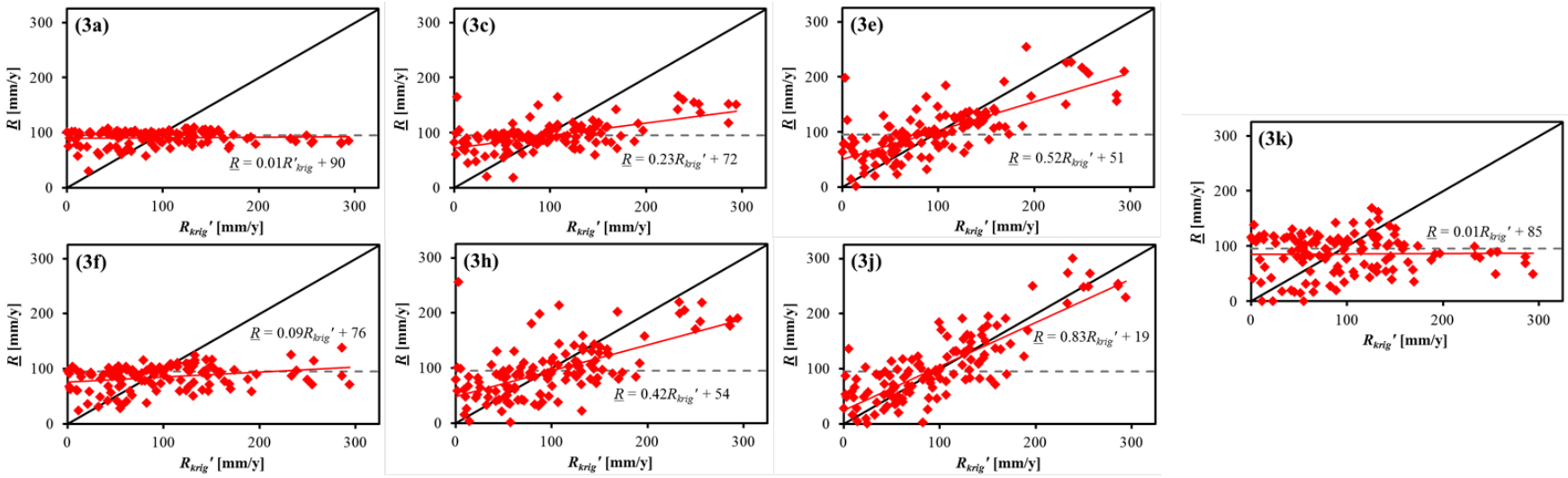

Figure 3.8. $\underline{R}$-versus-R'_{krig} scatterplots for Case 3a, c, e, f, h, j and k.

Figure 3.9a and b show the average error and average absolute error of $\underline{R}$ estimates (based on cell-by-cell values) for all cases. Cell-by-cell recharge values are assessed given that average errors based on these values display consistency with volumetric recharge error. In general terms, a reduction in $\underline{R}$ error is evident as the number of preferred hydraulic parameter values increases, as expected given that preferred values guide the calibration process closer to the parameter set of the reference case. A clear error reduction trend is apparent between Cases 3b and 3e, and Cases 3f and 3k. Inconsistencies in this trend are evident for Case 3a, occurring as a result of a less-than-sufficient calibration fit. For cases where an adequate fit is obtained (all except 3a and k), the average error is 3% smaller. The average absolute error is 18% smaller, on average, with preferred *K*-homogeneity. Although errors based on Cases 1 and 2 cannot be compared directly to those of Case 3 (given the difference in the calibration approaches), they display consistent patterns in error reduction with increasing degrees of hydraulic parameter information.

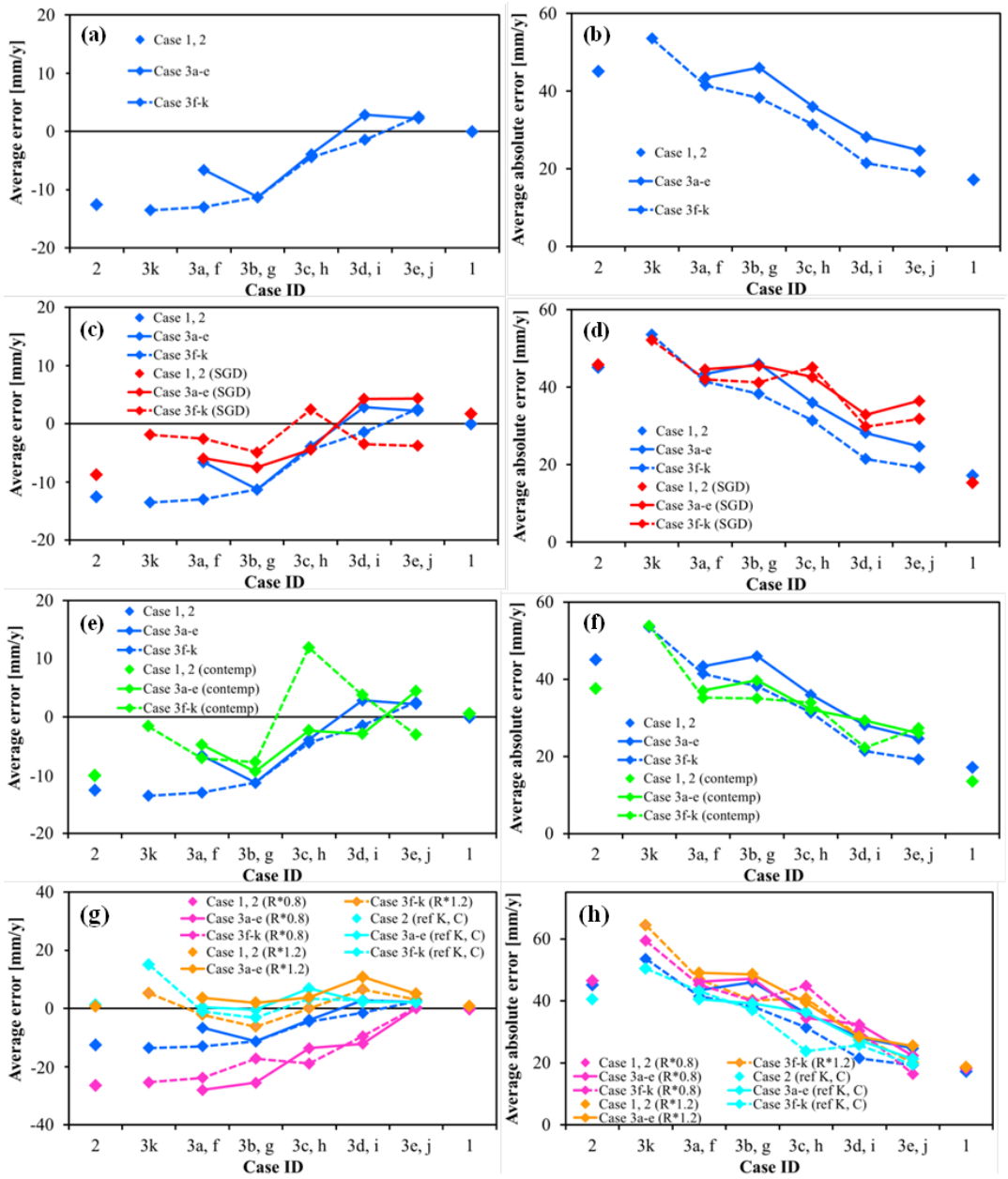

Figure 3.9. Average error and average absolute error of *R* for (a)-(b) Cases 1 to 3, (c)-(d) Cases 1 to 3 with a SGD calibration target, (e)-(f) Cases 1 to 3 for the contemporary steady-state model ("contemp"), and (g)-(h) Cases 1 to 3 with initial recharge values of 80 mm/y ("R*0.8") and 120 mm/y ("R*1.2"), and initial hydraulic parameter values equal to those of the reference-model ("ref K, C"). Errors are based on cell-by-cell recharge values. The blue markers and lines in (g) and (h) are results from the original cases shown in (a) and (b). Note the different y-axis scale on (g).

3.5.4. Use of SGD as a calibration target

The thirteen inverse modelling cases (1 to 3k) presented above are re-evaluated with the addition of an SGD calibration target. The model-to-measurement misfit for eleven of the thirteen cases (1, 2, 3a, b, c, d, e, f, g, i and j) is considered to be sufficiently low to allow for reliable comparisons of recharge estimates between cases (RMSE values of <0.032 m (for head) and <5 m^3/d (for flux); Table 3.4). The deviation from both preferred *K* and *C* values is considered large for all cases except 3f (Table 3.5). Cases 3a, b, c, d, e and j, in particular, display very large deviations from preferred *K* values. This highlights that the inclusion of the flux target produces generally worse estimates of spatially distributed hydraulic parameters (and therefore recharge, as discussed below and in Section 3.6) due to the addition of SGD to the calibration objective function.

Figure 3.9c and d show that a reduction in $\underline{R}$ error is generally apparent with increasing number of preferred hydraulic parameter values, with or without the flux target. Errors for Cases 3a, e, h and j do not display consistent patterns as a result of either a less-than-sufficient model fit (for Case 3h), or large deviations from preferred values (for Cases 3b, e and j). The amount by which the error is reduced through addition of preferred values, however, is smaller where the flux target is applied. In particular, the average recharge error is not reduced from Case 3f to 3j. This is at least partly due to the smaller average errors obtained where fewer preferred values are used (Cases 3a, b, f and g) (note the inadequate level of fit obtained for Case 3a without the flux target, therefore precluding a reliable comparison between Case 3a with and without the flux target). Of the ten cases where a reliable comparison can be made between cases with and without the inclusion of the flux target (i.e., where sufficient model fits are obtained; all cases

except 3a, h and k), the average error is smaller for cases 2, 3b, f and g (the average error is 25% smaller, on average, across the ten cases), and the average absolute error is smaller for cases 1 and 3b (the average error is 14% larger, on average, across the ten cases) where the flux target is used.

3.5.5. Application of an alternative steady-state condition

Inverse modelling cases 1 to 3k are re-evaluated for the contemporary steady-state model. The model-to-reference model misfit for nine of the thirteen cases (1, 2, 3b, c, d, e, f, g, and i) is considered adequate for the purpose of comparing recharge estimates between cases (RMSE values <0.017 m; Table 3.4). The deviation from both preferred *K* and *C* values is considered insignificant for Case 3b only (Table 3.5). The difference between neighbouring *K* values, on average, for Cases 3f to 3j is relatively small (Table 3.5).

Figure 3.9e and f show the (cell-by-cell-based) average error and average absolute error of $\underline{R}$ for all cases involving the contemporary steady-state model. A reduction in $\underline{R}$ error is apparent with increasing number of preferred hydraulic parameter values, similar to the error trends identified from the pre-development model. The error-reduction trend is particularly apparent for Case 3b to 3e and 3g to 3j (note that, while the average error of Cases 3g and 3h are of opposite sign, the average error magnitudes and the average absolute errors are comparable). For cases where a sufficient model fit is obtained for both the pre-development and contemporary models (all except 3a, h, j and k), the average error is 18% smaller, and the average absolute error is 9% smaller, on average, where the contemporary model is used.

3.5.6. Effect of initial recharge and hydraulic parameter values

Cases 1 to 3k are re-evaluated using different initial recharge and hydraulic parameter values. The model-to-reference model misfit for 31 of the 38 cases (thirteen cases for an initial recharge rate of 80 mm/y, thirteen cases for an initial recharge rate of 120 mm/y, and twelve cases for initial hydraulic parameter values specified as reference-model values) is considered adequate for comparing recharge estimates between different cases (RMSE values <0.021 m; Table 3.4). The deviation from preferred hydraulic parameter values is large only for Cases 3h to 3j for each of the different initial parameter value sets (Table 3.5). For Cases 3f to 3j, the deviation from preferred *K* values generally increases (Table 3.5) because the imposition of preferred *K*-homogeneity counteracts that of preferred values.

Figure 3.9g and h show the (cell-by-cell-based) average error and average absolute error of $\underline{R}$ for cases with different initial recharge and hydraulic parameter values. The error-reduction trend (with increasing number of preferred hydraulic parameter values) is generally apparent for all cases. Average $\underline{R}$ values are larger where the initial recharge value of 120 mm/y is applied, and smaller where the initial recharge value of 80 mm/y is applied, compared to the initial recharge value of 100 mm/y (Figure 3.9g). Relative to the original cases, the average error is 100% larger, on average, with an initial recharge value of 80 mm/y (excluding Cases 3a, h, i and k due to inadequate model-fit), and 33% smaller, on average, with an initial recharge value of 120 mm/y (excluding Cases 3a, h and k due to inadequate model-fit). The average absolute error is 2% and 8% larger, on average, for initial recharge values of

80 and 100 mm/y, respectively, compared to the original cases. Both the average error and average absolute error is smaller with the initial hydraulic parameter values from the reference model compared to the original cases (61% and 6% smaller, on average, for all cases except 3a and k, where poor fits are obtained).

3.6. Discussion

The level of model fit (i.e., in terms of head RMSE) and the accuracy of estimated recharge values displays a poor correlation amongst the various cases. This is evidenced by r^2 values of 0.02 and 0.03 for regression lines representing RMSE-average recharge error and RMSE-average absolute recharge error, respectively (for cases where an adequate model fit is obtained). This is consistent with previous studies (e.g., McKenna et al., 2003; Hill and Tiedeman, 2007) that show that the head match is not a proxy for accurate parameter values. The attainment of inaccurate recharge values despite low model misfit exemplifies the compensatory roles assumed by hydraulic parameters in offsetting errors in recharge, as illustrated by, e.g., the underestimation of *K* values where recharge values are lower than those of the reference model (Figure 3.6c).

Inverse modelling results demonstrate that the ability of the calibration process to inform recharge varies significantly with the degree to which hydraulic parameters are constrained, in agreement with Sanford (2002). Where all hydraulic parameters are known, it is shown that spatially averaged recharge can be estimated accurately (average cell-by-cell error of -0.02%), whereas spatially distributed recharge is less well-constrained, as evidenced by the moderate degree of scatter about the regression

line (r^2 value of 0.66; Figure 3.4). That is, an accurate basin-scale estimate of recharge does not guarantee that a reasonable recharge distribution has been obtained from model calibration. While failure of the calibration process to identify recharge spatial variability may not impact the model's predictability (e.g., to estimate drawdown) directly, there are important implications for the distribution of *K* values, which are more likely to influence the prediction of pumping impacts (e.g., Freeze and Cherry, 1979). Where aquifer parameters are known, discrepancies in the calibrated recharge distributions are caused by compensation between recharge parameters, as evidenced by low recharge identifiability values (Figure 3.5).

The amount of hydraulic parameter information required to obtain a reasonable estimate of recharge, in particular its spatial variability, on the basis of the current findings, appears to be excessive from a practical perspective. For cases with initial recharge values based on the average reference-model recharge rate and literature-derived initial hydraulic parameter values (Table 3.2), a reasonable spatially averaged recharge estimate generally requires at least 129 of the 258 reference-model hydraulic parameter values to be set as regularisation constraints (<10% average error is obtained for 23 of the 31 cases where an adequate model fit is obtained; Figure 3.9a, c and e). While average recharge can be estimated with a reasonable degree of accuracy for these cases, the spatial variability of recharge requires higher levels of hydraulic parameter information. This is evidenced by the requirement for at least 232 preferred hydraulic parameter values to achieve average absolute recharge errors of <20% (average absolute error <20% is obtained for 4 of the 31 cases where an adequate model fit is obtained; Figure 3.9b, d and f).

The incorporation of preferred *K*-homogeneity is shown to be of benefit to the estimation of recharge for some cases. In particular, the imposition of *K*-homogeneity is shown to produce, on average, improved estimation of average recharge (average error is 14% lower), and spatial variability in recharge (average absolute error is 15% lower), across all cases that involve regularisation where an adequate fit is obtained (total of 18 cases). This is because smoother *K*-fields lead to enhanced spatial variability in recharge, due to the effects of parameter compensation and the heterogeneous nature of the reference model (Figure 3.9b). However, the benefit of preferred *K*-homogeneity to recharge estimability for other real-world situations is yet to be demonstrated.

The use of SGD, which can be inferred in practice from environmental tracers and other methods (e.g., Burnett et al., 2006; Houghham et al., 2008), is shown to be of variable worth as a calibration target in the estimation of recharge for the inverse modelling cases considered in the current study. On average, the inclusion of SGD improves the estimation of spatially averaged recharge (25% smaller average error; Figure 3.9c), but not the estimation of spatial recharge variability (14% larger average absolute errors; Figure 3.9d). The improvement in average recharge estimability occurs because of the relationship between SGD and recharge (Cheng and Ouazar, 1999). Such improvements occur only where relatively few (<129) preferred hydraulic parameter values are employed. This indicates that the constraint on average recharge imparted by having a large number of preferred values distributed over the basin is similar to that imparted by a basin-wide flux target (e.g., SGD). The reduction in the estimability of spatially variable recharge with the addition of the SGD target is consistent with the large deviations from preferred

hydraulic parameter values obtained with the SGD target (Table 3.5). The use of distributed flux data (e.g., spatial variability in SGD), while not considered here, is likely to improve the constraint on the spatial variability in recharge.

Results from the contemporary steady-state model are, in general terms, similar to those obtained from the pre-development steady-state model. This indicates that trends in recharge estimability are not specific to a particular steady-state condition (Figure 3.9e and f). The findings of the current investigation are therefore somewhat validated, notwithstanding that both models share many similarities. The simulation of pumping is shown to produce, on average, lower recharge errors than the pre-development model (average error is 18% lower and average absolute error is 9% lower), given that the prescribed flux provides an additional constraint to the inverse problem (Zhang, 2014). In general terms, the results from both the addition of SGD as a calibration target and pumping suggest that the ability of groundwater fluxes to constrain the estimation of recharge is complicated and problem-dependent. This is supported by Voss (2011), who advised that the incorporation of alternative data types (i.e., besides head) to assist in constraining the uncertainty of groundwater model parameters may not always be as effective as commonly thought.

The degree to which hydraulic parameters must be constrained to achieve reasonable recharge estimates is shown to be highly dependent on initial recharge and hydraulic parameter values (e.g., <10% average error is obtained for initial recharge values of 120, 100 and 80 mm/y using 0%, 50% and 90% of reference model-based preferred values, respectively; Figure 3.9g). The role of initial parameter values in constraining the calibration process is highlighted by the larger average recharge error where a

larger initial recharge value is specified (120 mm/y), compared to the smaller average recharge error where a smaller initial recharge value is specified (80 mm/y) (Figure 3.9g). This is in agreement with previous studies that demonstrate the importance of initial parameter values within traditional groundwater model calibration problems (e.g., Bravo et al., 2002; Bahremand and De Smedt, 2008). Differences in estimated parameter values due to different initial values occur because of the presence of multiple objective function minima (e.g., Hill and Tiedeman, 2007).

The analysis of parameter identifiability provides insight into the degree of non-uniqueness between parameters of both the same type and different types on a spatially distributed basis. For example, significantly lower recharge identifiability is evident where hydraulic parameters are also estimated as part of the calibration process (110 of the 125 recharge identifiability values <0.05; Figure 3.7c) compared to where hydraulic parameters are known (22 of the 125 recharge identifiability values <0.05; Figure 3.5). This indicates that the ability of the calibration process to reduce the potential for wrongness in recharge estimates is small, due to the need to jointly estimate hydraulic parameters on the basis of the same dataset. The spatial variability in identifiability values display complicated patterns that are related primarily to the proximity of boundary conditions (e.g., highest *K*-identifiabilities are located adjacent to the fixed-head ocean boundary) and the extent of aquifers (e.g., highest recharge-identifiabilities are located where recharge occurs within the TS aquifer, underlying unsaturated QL sediments). The distribution of observation data typically dominates such patterns (e.g., Fienen et al., 2010; Wallis et al., 2014); this is not the case here because of the evenly distributed head targets. Confidence in

these results is enhanced given that the effect of model non-linearity on identifiability values is relatively minor (values differ by 8% on average). A strong inverse correlation between parameter identifiability and post-calibration parameter error variance (r^2 values of >0.99; not shown for brevity) is obtained given the assumption of negligible measurement errors, which is considered reasonable for a synthetic reference model (Section 3.4.4). Therefore, the identifiability values reported in this study can be considered as a proxy for post-calibration parameter uncertainty (e.g., Moore and Doherty, 2005).

In this study, the ability of the calibration process to inform parameters is evaluated primarily with respect to the error associated with parameter estimates, which can be quantified through the application of a reference model. In practice, however, estimated parameter values cannot be evaluated in terms of error. The reasonableness of parameter values must therefore be assessed in terms of whether or not they can be considered plausible on the basis of field data such as aquifer test-derived *K* estimates (e.g., Langevin and Zygnerski, 2013; Knowling et al., 2015). Such assessments are limited by the need to consider the scale-dependency of hydraulic properties (e.g., Sánchez-Vila et al., 2006; Li et al., 2011).

In studies that apply unsaturated zone models to estimate recharge, the parameters of recharge models are constrained using expert knowledge related to, e.g., land-surface and unsaturated-zone processes (Jyrkama et al., 2002; Chemingui et al., 2015). In the current study, recharge is estimated solely through the application of the regional-scale groundwater model, and recharge modelling serves merely to provide the reference-case recharge spatial variability. We focus on aspects of the groundwater

model, aside from expert knowledge of recharge itself, in attempting to inform recharge independent of other recharge estimation techniques. Further work, including coupling of the recharge model with MODFLOW, is required to investigate the extent to which recharge can be informed by groundwater model calibration with varying degrees of information pertaining to both recharge (e.g., spatial zones of land use, soil type, etc.) and aquifer hydraulic properties.

While this study shows that the estimation of recharge through calibration of field-scale steady-state groundwater models may be limited, the estimability of recharge through calibration of transient groundwater models, which involves considerably more observations than steady-state models, yet requires the additional consideration of storage parameter uncertainty, remains unknown for practical situations. The degree to which field-scale transient model calibration can inform recharge and its variability in time is the subject of future work.

It is expected that the findings of this study are to some degree dependent on the unique characteristics of the USB reference model (e.g., the lack of groundwater-dependent streams). Application of similar analyses to other field sites, for example, temperate climates where surface water and groundwater are often hydraulically connected, may reveal an enhanced estimability of recharge due to the availability of stream discharge data, which may reduce parameter non-uniqueness. Therefore, the investigation of recharge estimability for other field sites is warranted to extend the current findings to a wider range of situations, in particular where measurements of groundwater flux are more readily attainable than is possible for USB.

3.7. Conclusions

Non-uniqueness poses one of the primary challenges to the estimation of groundwater model parameters. The current study focuses on the ramifications of non-uniqueness between recharge and aquifer hydraulic parameter values in terms of recharge estimability through groundwater model calibration. Presently, the degree to which recharge can be informed through calibration is largely unknown for complex field-scale situations. This study evaluates the ability of calibration to inform recharge within a regional setting using a series of highly parameterised steady-state inverse modelling experiments containing varying degrees of hydraulic parameter constraints.

Inverse modelling results show that the ability of the calibration process to inform recharge varies significantly with the degree to which hydraulic parameters are constrained. For the field-scale groundwater model considered, reasonable estimates of recharge are shown to require large amounts of hydraulic parameter information. For example, an average error of <10% requires that 50% of the reference-model hydraulic parameters are included in the calibration using preferred-value regularisation (>100 *K* values across the 129 km^2 study area). The estimation of the spatial variability of recharge requires a higher level of hydraulic parameter information (e.g., >200 *K* values needed for an average absolute recharge error <20%). The use of both SGD (as a calibration target) and pumping data are shown to be of benefit to the estimation of average recharge, particularly where relatively few preferred hydraulic parameter values are included. The addition of pumping data also produces improved estimates of the spatial variability in recharge.

This study offers guidance to groundwater modelling practitioners by highlighting the potential effects of non-uniqueness in terms of recharge estimability in a real-world modelling scenario, and by providing insight into scenarios under which recharge may (or may not) be informed through model calibration in practical situations. The findings of this study provide a useful benchmark for evaluating the extent to which field-scale groundwater models can be used to inform recharge under practical data-availability limitations. Future work in this area should evaluate the estimability of transient recharge through calibration, and consider field sites that display different characteristics (e.g., surface water-groundwater interactions).

Acknowledgements

The author M.J. Knowling was supported by an Australian Postgraduate Award, and top-up scholarships from the Goyder Institute for Water Research and the National Centre for Groundwater Research and Training (NCGRT), a collaborative initiative of the Australian Research Council and the National Water Commission. The authors wish to acknowledge Etienne Bresciani who reviewed an early version of the manuscript, and Daniel Partington for programming assistance. The authors also acknowledge Michael Fienen, Hossein Hashemi and the anonymous reviewers whose comments helped improve the manuscript.

Chapter 4

4. Time-varying recharge estimability through field-scale groundwater model calibration

This chapter is based on the following manuscript, which is currently under review at Journal of Hydrology:

Knowling, M.J., Werner, A.D. Transient recharge estimability through field-scale groundwater model calibration.

4.1. Abstract

The estimation of recharge through groundwater model calibration is hampered by the non-uniqueness of recharge and aquifer parameter values. It has been shown recently that the estimability of spatially distributed recharge through calibration of steady-state models for practical situations (i.e., real-world, field-scale aquifer settings) is limited by the need for excessive amounts of hydraulic parameter and groundwater level data. However, the extent to which temporal recharge variability can be informed through transient model calibration, which requires the additional consideration of storage parameters compared to steady-state models, is presently unknown for practical situations. In this study, time-varying recharge estimates, inferred through calibration of a field-scale highly parameterised groundwater model, are systematically investigated subject to changes in (1) the degree to which

hydraulic parameters including hydraulic conductivity (K) and specific yield (S_y) are constrained, (2) the number of water-level calibration targets, and (3) the temporal resolution on which recharge is estimated. The analysis involves the use of a synthetic reality, i.e., a reference model, based on a groundwater model of Uley South Basin, South Australia. Identifiability statistics are used to evaluate the ability of recharge and hydraulic parameters to be estimated uniquely. Results show that reasonable estimates of monthly recharge (<30% recharge root-mean-squared error) require a considerable amount of transient water-level data, and that the spatial distribution of K is known (i.e., through joint recharge-and-S_y estimation). The joint estimation of recharge, S_y and K, however, precludes reasonable inference of recharge and hydraulic parameter values. This study suggests that the estimation of temporal recharge variability through calibration may be impractical for real-world settings, limited by the requirement for both a significant amount of water-level data and well-constrained hydraulic parameter (in particular K) distributions.

4.2. Introduction

Effective groundwater management strategies often require reliable estimates of recharge and its variability in time and space. However, recharge is widely regarded as one of the most difficult water balance components to quantify given that it cannot be measured directly, and hence, must be inferred typically from the application of multiple methods (Scanlon et al., 2002). An increasingly popular recharge estimation method is the application of a numerical groundwater model within an inverse modelling framework (e.g., Essaid et al., 2003; Hashemi et al., 2013). This approach involves the inference of recharge through calibration or history matching (i.e.,

minimising the discrepancy between field measurements and corresponding model-generated outputs).

The estimation of recharge through groundwater model calibration is hampered by the non-uniqueness between estimated values of recharge and aquifer hydraulic properties (Sanford, 2002). Non-uniqueness arises from the limited information content within field measurements and the correlation between model parameters (Carrera and Neuman, 1986). Recently, Erdal and Cirpka (2016) applied the Ensemble Kalman filter (Evensen, 1994) to jointly estimate spatially distributed recharge and hydraulic conductivity (*K*) on the basis of head observations for a highly simplified synthetic aquifer. They showed that accurate values of recharge and *K* can be estimated simultaneously providing that a sufficient degree of prior information is available (i.e., the geostatistics of the true (initial ensemble) recharge- and *K*-fields are well-known). Knowling and Werner (2016) extended the work of Erdal and Cirpka (2016) by jointly estimating recharge and hydraulic parameters (i.e., *K*) for a more realistic test case (i.e., a synthetic reality based on a highly parameterised steady-state groundwater model of a real-world field site) for the purpose of evaluating the estimability of recharge and its spatial distribution for practical situations. The results of their study indicate that the estimation of steady-state recharge through field-scale model calibration requires unrealistic amounts of hydraulic parameter (*K*) information and groundwater level data. This suggests that previous steady-state groundwater modelling investigations that attempt to infer regional-scale, spatially variable recharge through model calibration likely produce non-unique and potentially erroneous estimates.

The investigations of Erdal and Cirpka (2016) and Knowling and Werner (2016) did not examine the extent to which the temporal variability of recharge can be informed, in the face of non-uniqueness, through groundwater model calibration. While transient groundwater models capture considerably more observations than steady-state models, thereby reducing non-uniqueness, transient model calibration requires the additional consideration of storage parameters, such as specific yield (S_y).

A small number of studies have inferred time-varying recharge through transient model calibration. For example, Lubczynski and Gurwin (2005) estimated recharge variability in their calibration of a transient groundwater model of the Sardon catchment (Spain). Their recharge estimates were obtained using a "trial and error" approach (i.e., model performance was evaluated with respect to incremental adjustments in recharge values), and were subject to fixed storativity and K distributions from preceding automated calibration of the model in which recharge was fixed. Liu et al. (2008) estimated time-varying recharge from their calibration of a transient groundwater model of the North China Plain. Their recharge rates were estimated simultaneously with K and boundary conductance, but were subject to a fixed S_y value. Dickinson et al. (2004) inferred the time-variability of mountain-front recharge to idealised representations of alluvial basins within the south-western United States from long-term groundwater levels by inverting a one-dimensional analytical model developed by Townley (1995). The effect of parameter non-uniqueness on temporal recharge estimability has not been evaluated in previous studies. Therefore, the extent to which the temporal variability of recharge can be informed through groundwater model calibration, in particular for complex field-scale aquifer settings, is presently unknown.

The objective of the current study is to address this knowledge gap, and extend the work of Knowling and Werner (2016) by evaluating the ability of groundwater model calibration to inform the temporal variability of recharge for practical, regional-scale situations. This is undertaken using a systematic, quasi-hypothetical inverse modelling approach involving a field-scale, highly parameterised groundwater model of Uley South Basin (USB), South Australia. Here, we adopt the most common approach to groundwater model calibration, i.e., the application of gradient optimisation methods (e.g., Poeter and Hill, 1999; Doherty, 2016), such that the findings offer guidance to groundwater modelling practitioners. The USB is selected as a case study because of the availability of extensive observation datasets, which provide a comprehensive depiction of the spatial and temporal trends in groundwater behaviour. These formed the basis for an existing highly parameterised groundwater model by Knowling et al. (2015), and the estimation of spatially and temporally variable recharge rates using field-based methods and one-dimensional unsaturated zone modelling (Ordens et al., 2012; Ordens, 2014).

4.3. Theoretical background

Transient two-dimensional flow within an unconfined heterogeneous aquifer is given by the Boussinesq equation (Boussinesq, 1904):

$$\frac{\partial}{\partial x}\left(K_x b \frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y b \frac{\partial h}{\partial y}\right)+W=S_y \frac{\partial h}{\partial t} \tag{4.1}$$

where x and y [L] are Cartesian coordinates, h [L] is the hydraulic head, b [L] is the saturated aquifer thickness (which varies with h), W [L/T] is the volumetric source/sink term per unit area expressing the combined effects of groundwater recharge, pumping, etc. (positive values indicate inflow), K_x and K_y [L/T] are the hydraulic conductivities in the x and y directions, respectively, and t [T] is time.

Groundwater flow modelling codes solve equation (4.1) in a piecewise manner using numerical methods such as finite-difference or finite-element approximations (e.g., Harbaugh et al., 2000; Diersch, 2005). The computation of h at the node scale is a function of the ratios W/K and W/S_y. Therefore, a unique estimate of recharge (or K or S_y) is not attainable based on observations of h alone; only recharge-to-K and recharge-to-S_y ratios may be determined in the absence of additional information regarding K, S_y, recharge and/or groundwater fluxes. Temporal groundwater-level fluctuations are governed primarily by recharge, discharge (e.g., pumping) and S_y in climate-driven aquifers (e.g., Maréchal et al., 2006; Ordens et al., 2012; Zhou et al., 2014). Non-uniqueness at the scale of the model domain is complicated by aspects such as the spatial patterns in aquifer parameters, flow directions, boundary conditions, etc., and hence, is more challenging to characterise for practical situations in which observation-data scarcity is a limiting factor. The current study focuses on the estimability of transient recharge in the face of non-uniqueness, i.e., uncertainty in S_y and K, for practical situations.

4.4. Methodology

4.4.1. Study area

The USB (~129 km^2) is a coastal, topographically enclosed groundwater basin located in the southern Eyre Peninsula, South Australia (34º76' S, 135º56' E). The climate of the region is characterised by winter-dominant rainfall (May-October), and hot, dry summer months (November-April) (Harrington et al., 2006). Average annual rainfall and pan evaporation rates are 560 and 1547 mm/y, respectively (Bureau of Meteorology, 2010). Groundwater supplies are obtained from an unconfined Quaternary sand and limestone aquifer (QL), which overlies a discontinuous Tertiary aquitard and a semi-confined Tertiary aquifer comprised of silty and clayey sand (TS) (Zulfic et al., 2007). Despite evidence of preferential pathways through surface calcrete, the QL aquifer exhibits diffuse flow behaviour (Werner, 2014). Groundwater abstraction from the QL aquifer constitutes the primary freshwater resource for the Eyre Peninsula region (Werner et al., 2011).

4.4.2. Groundwater model

The groundwater model of USB developed by Knowling et al. (2015) constitutes the "reference model" that is used as a basis for analysing the inverse modelling results of the current study. MODFLOW (Harbaugh et al., 2000) is used to simulate groundwater flow within USB spanning the period January 2003 to December 2012. Table 4.1 lists the main features of the reference USB model. Figure 4.1 provides a schematic of the reference model and its parameter distributions that are pertinent to the current study, i.e., the hydraulic properties of the QL aquifer. The parameter distributions of the underlying TS aquifer (i.e., specific storage, K) are not estimated in this study as a means of reducing computation times associated with the parameter

estimation process, and are set based on the calibration results of Knowling et al. (2015).

Table 4.1. Summary of reference model characteristics (adapted from Knowling and Werner (2016)).

Model characteristic	**Description**
Layer(s) type	Unconfined upper layer (QL aquifer), convertible confined/unconfined lower layer (TS aquifer). Intervening aquitard simulated implicitly (Chiang and Kinzelbach, 1998).
Spatial discretisation	100 m × 100 m model-grid cells (Figure 4.1a).
Temporal discretisation	Monthly stress periods; 18 time steps per stress period.
S_y (QL aquifer)	Estimated through calibration (Figure 4.1b), constrained by field data and literature values.
Specific storage (S_s) (TS aquifer)	Estimated through calibration, constrained by field data and literature values.
K (QL and TS aquifers)	Estimated through calibration (QL K; Figure 4.1c), constrained by field data and literature values. Vertical K assumed 0.1 times horizontal K.
Ocean boundary	Constant-head boundary with density corrections (Morgan et al., 2012).
Interaction with adjacent basins	General-head boundaries (GHB; Harbaugh et al., 2000). GHB conductances (C) estimated through calibration, constrained by K and aquifer thickness information.
Spatially and temporally variable recharge	Applied to uppermost active layer. Spatial and temporal recharge variability based on the modelling of Ordens (2014) (Figure 4.2). Spatial distribution is held constant-in-time. Spatially and temporally averaged recharge rate is 78 mm/y.
Evapotranspiration (ET)	ET package (Harbaugh et al., 2000) simulates shallow groundwater ET (unsaturated zone ET accounted for in the recharge rates of Ordens (2014)).
Groundwater pumping	Individual pumping-well rates based on metered data and extrapolation (where metered data are unavailable) (Werner et al., 2011).

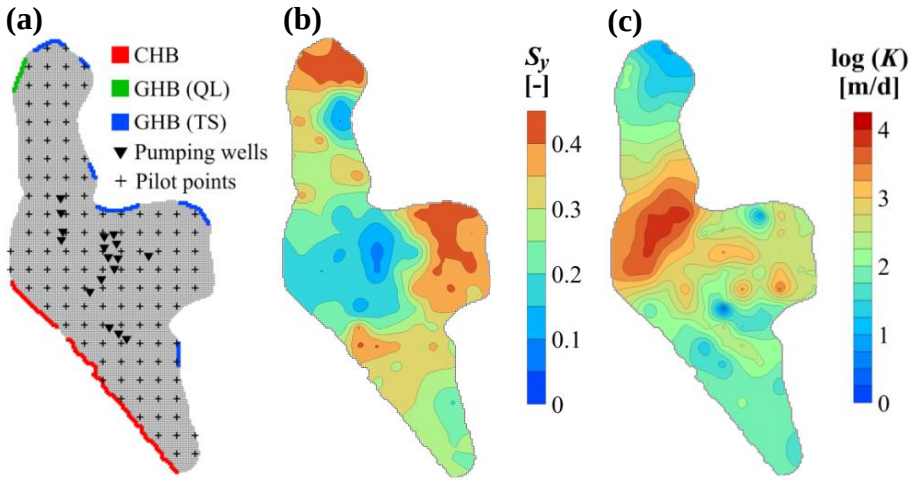

Figure 4.1. Reference model characteristics (adapted from Knowling and Werner (2016)): (a) grid, boundary conditions ("CHB" denotes constant-head boundary, and "GHB" denotes general-head boundary), pumping well and pilot point locations; (b) S_y of the QL aquifer; and (c) log(K) of the QL aquifer.

The spatial variability of S_y and K is achieved using pilot points (de Marsily et al., 1984), distributed in a regular grid configuration (Figure 4.1a). An isotropic exponential variogram (of S_y and log(K)), with a range of 1200 m and a nugget of zero, produces cell-by-cell S_y (Figure 4.1b) and K (Figure 4.1c) variability. A single boundary conductance (C) value (=195 m^2/d in the reference model) is used for the QL GHB boundary reach.

The spatial distribution of recharge, obtained from temporally averaged rates of Ordens (2014) over the period 2003–2012, is fixed in a relative sense in the reference transient model (Figure 4.2a), such that the ratio of recharge values in neighbouring cells remains unchanged. That is, while spatial patterns of recharge produced by Ordens (2014) vary in time, only transient variability is modified in the current study, and spatial variability is fixed, at least in relative terms. Temporal recharge variability is achieved by scaling the time-averaged recharge distribution such that

the total recharge to USB is equivalent to monthly totals obtained by Ordens (2014), for the period 2003–2012 (Figure 4.2b). Fixing the relative spatial distribution of recharge in the reference case in this way allows for a significantly reduced computational burden during calibration, but nonetheless allows for analyses of the estimability of temporal recharge signals.

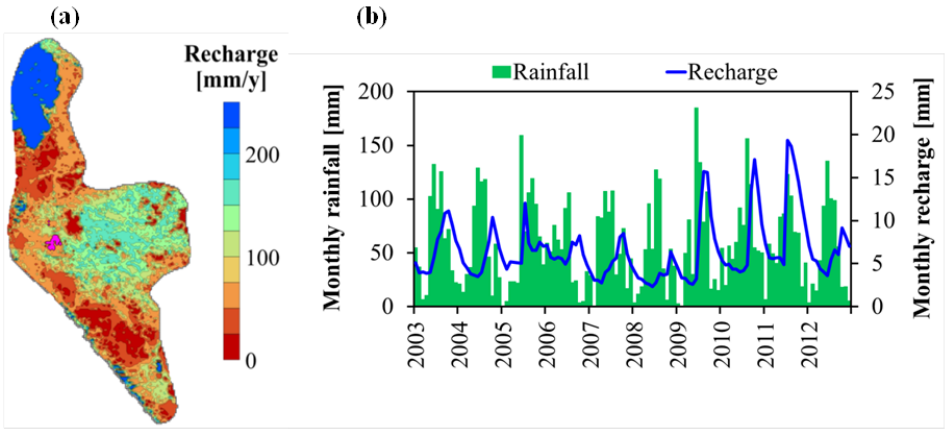

Figure 4.2. (a) Spatially distributed recharge rates (averaged over the period 2003–2012), and (b) time series of monthly recharge (and rainfall) rates applied to the reference model.

4.4.3. Model calibration

Model parameters are estimated using PEST (Doherty, 2016). To reduce computation times, BeoPEST (Schreüder, 2009), a version of PEST that allows model-run parallelisation, is used. PEST adopts the Gauss-Marquardt-Levenberg method to minimise a least-squares objective function Φ. For parameter estimation problems that employ Tikhonov regularisation, Φ is given by:

$$\Phi = (\mathbf{Xp} - \mathbf{h})^t \mathbf{Q}_m (\mathbf{Xp} - \mathbf{h}) + \mu(\mathbf{Zp} - \mathbf{d})^t \mathbf{Q}_r (\mathbf{Zp} - \mathbf{d}) \quad (4.2)$$

where **p** is a vector of model parameter values, **h** is a vector of field observations, **X** is a matrix containing sensitivities of model outputs (for which there are corresponding observations **h**) with respect to **p**, **d** is a vector of Tikhonov regularisation "observations", which represent preferred parameter values and/or relationships, **Z** is a matrix containing sensitivities of model-generated counterparts to regularisation observations **d** with respect to **p**, μ is a regularisation weight factor, and $\mathbf{Q}_m$ and $\mathbf{Q}_r$ are diagonal matrices containing the squares of weights assigned to measurement and regularisation observations, respectively.

The degree to which the inverse problem is regularised is controlled using a user-specified "target measurement objective function" Φ_m^l (Fienen et al., 2009). The regularised inverse problem is solved via minimisation of Φ_r under the constraint that $\Phi_m \leq \Phi_m^l$. The reader is directed to Doherty and Skahill (2006), Doherty and Hunt (2010) and references cited therein for a detailed description of pertinent calibration methodologies.

Sensitivities of model outputs with respect to parameter values (comprising the **X** and **Z** matrices) are computed using central finite differences with a 1% perturbation of parameter values, following James et al. (2009). The maximum parameter value upgrade factor used is 0.2. The specification of Φ_m^l is dependent on initial inverse modelling results. The reader is referred to Knowling and Werner (2016) for a more detailed description of the optimisation variables adopted.

4.4.4. Parameter identifiability

Parameters are evaluated in terms of their identifiability (i.e., their ability to be estimated uniquely through the process of calibration). The identifiability of parameter *i*, d_i, is given by (Doherty and Hunt, 2009):

$$d_i = [\mathbf{R}]_{i,i} \tag{4.3}$$

where **R** is a matrix that describes the relationship between estimated and real-world parameter values (commonly referred to as the "resolution matrix") in the absence of significant measurement uncertainty (Moore and Doherty, 2006; Paradis et al., 2015), and the subscript *i*,*i* designates the *i*th diagonal element. The construction of **R** is dependent on the regularisation technique employed; see Doherty (2016) for different regularisation-specific **R** formulations. For problems involving Tikhonov regularisation, **R** is given by (Doherty, 2016):

$$\mathbf{R} = (\mathbf{X}^t\mathbf{Q}_m\mathbf{X} + \mu\mathbf{Z}^t + \lambda\mathbf{I})^{-1}\mathbf{X}^t\mathbf{Q}_m\mathbf{X} \tag{4.4}$$

where λ is the Marquardt lambda parameter and **I** is the identity matrix. The value of d_i can vary between zero and one, which respectively indicate that the parameter under investigation is entirely unidentifiable and entirely identifiable. A d_i value of one does not mean that its estimated value will be free from error due to measurement uncertainty. Statistics such as parameter error variance (Moore and Doherty, 2005) are nevertheless not reported in the current study because measurement uncertainty is assumed negligible.

4.4.5. Inverse modelling cases

The inverse modelling methodology can be subdivided into the following phases (Table 4.2): (1) the estimation of time-varying recharge subject to fixed hydraulic parameter (S_y, K and C) values from the reference model ("Case 1"), (2) the joint estimation of time-varying recharge and spatially distributed S_y ("Case 2"), and (3) the joint estimation of time-varying recharge and the spatial distribution of S_y and K, and C ("Case 3"). Case 1 represents a best-case scenario in terms of recharge estimability given that all other parameter values are "known". Despite the use of error-free hydraulic parameter distributions within Cases 1 and 2, which constitutes an unrealistic situation in practice, these cases provide a framework for demonstrative and comparative purposes. The potential for parameter compensation is explored in Cases 2 and 3 (i.e., whereby correlated parameters are largely unconstrained).

Table 4.2. Inverse modelling cases.

ID	**Recharge**	**S_y**	**K, C**
Case 1	Estimated	Fixed as per reference model	Fixed as per reference model
Case 2	Estimated	Estimated	Fixed as per reference model
Case 3	Estimated	Estimated	Estimated

The calibration dataset for inverse modelling cases involving the estimation of recharge and S_y (Cases 2 and 3) comprises deviation-from-the-mean reference-model heads taken at pilot point locations within the upper layer, except at places where QL sediments are entirely unsaturated, resulting in a total of 94 sampling locations. Temporal head differences are used such that the inference of time-varying recharge and aquifer storage parameters is enhanced (Hill and Tiedeman, 2007; Ackerman et al., 2010). All head-difference targets are assigned uniform weights. For cases involving the estimation of recharge, S_y, K and C (Case 3), an average reference-

model head at each of the 94 pilot point locations is added to the calibration dataset given that the estimation of *K* and *C* parameters require time-averaged groundwater levels. Each average head target is assigned an equal weight, which is 100 times that of head-difference targets.

The number of head-difference targets is varied to examine the effect of the information content of the calibration dataset on estimated parameter values. The following numbers of head-difference targets are considered: (1) 11,280 (head-difference target at each sampling time, i.e., one every month; herein referred to as "calibration dataset A"), (2) 4980 (approximately half of the sampling times; "calibration dataset B"), and (3) 1222 (approximately one-tenth of the sampling times; "calibration dataset C"). Figure 4.3 illustrates the reference-model head sampling intervals.

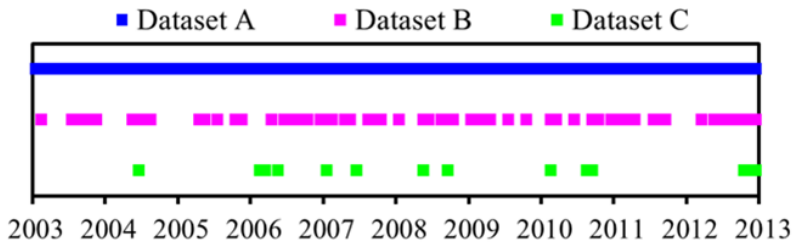

Figure 4.3. The three different reference-model head target sampling intervals considered.

Table 4.3 lists the constraints on parameter values in the form of both upper and lower bounds, and initial values. Constraints on hydraulic parameters are equivalent to those of the reference model. Hydraulic parameters are log-transformed within the calibration.

Table 4.3. Model parameterisation of inverse modelling cases.

Model parameters	Parameterisation method (number of parameters)	Bounds	Initial values
Recharge [mm/y]	Varies from case-to-case (Table 4.3)	0–300	78
S_y [-]	Pilot points (127)	0–0.41	0.2
K [m/d]	Pilot points (125)	$1–10^4$	100
C [m^2/d]	Single value	0.18–1800	18

A temporally uniform recharge rate of 78 mm/y (equal to the spatially and temporally averaged value of the reference-model recharge; Section 4.4.2) is prescribed prior to calibration. This means that deviations from the initial time-averaged value of recharge by calibration provide insight into the degree to which measurements inform the temporal recharge signal. The spatial distribution of recharge is equal to that of the reference model (Section 4.4.2).

The temporal variability of recharge is parameterised in a number of ways, as listed in Table 4.4. Recharge is estimated on both a monthly (i.e., an adjustable recharge parameter is assigned to all 120 months within the ten-year simulation period; referred to as "R_{120}" herein) and yearly basis (i.e., an adjustable recharge parameter is assigned to every year within the simulation period; "R_{10}"). Recharge is also estimated on a combined monthly-yearly basis (i.e., an adjustable recharge parameter is assigned to each of the twelve months January-December and to each year within the simulation period; "R_{22}"). A variant of both the R_{120} and R_{22} parameterisations is considered to ensure that a reasonable degree of month-to-month recharge variability is obtained, giving rise to two additional cases "R_{120U}" and "R_{22U}", respectively. This is achieved by imposing Tikhonov regularisation in the form of preferred parameter uniformity (described below).

Table 4.4. Temporal recharge parameterisation schemes.

Temporal recharge parameterisation			
ID	**Description**	**No. of recharge parameters**	**Preferred uniformity?**
R_{120}	Monthly recharge parameters	120	No
R_{22}	Combined monthly-yearly recharge parameters	22	No
R_{10}	Yearly recharge parameters	10	No
R_{120U}	Equivalent to R_{120}, except with preferred recharge uniformity constraints applied	120	Yes
R_{22U}	Equivalent to R_{22}, except with preferred recharge uniformity constraints applied	22	Yes

Preferred uniformity of recharge rates between subsequent months is used as a means of achieving reasonableness of calibrated recharge values. This represents a mode of imparting expert knowledge regarding temporal recharge variability (e.g., recharge rates in July are more likely to be equal to those in June and August than they are in May and September, March and November, etc.), without the need for "hard" data pertaining to relative monthly recharge rates. Where preferred-uniformity is employed, the $\Phi_m^{\,l}$ is set on a case-by-case basis such that the variability of calibrated recharge in-time appears to be reasonable, while also ensuring that adequate head matches are obtained.

4.5. Results

4.5.1. Case 1

The degree of head misfit is considered to be low for all calibrations (root-mean-squared-error (RMSE) values <0.03 m). Head-RMSE ($RMSE_h$) values range between 0.007 and 0.027 m. For cases involving preferred uniformity (R_{120U} and R_{22U}), Φ_m is within reasonable proximity to $\Phi_m^{\,l}$ (78% to 101%). Larger $RMSE_h$ values are

apparent where fewer recharge parameters are estimated through calibration (e.g., $RMSE_h$ values of 0.013, 0.018 and 0.023 m are obtained, on average, for R_{120}, R_{22} and R_{10}, respectively), which is appropriate given the well-known relationship between the number of adjustable parameter employed and the degree of fit attainable, except for cases adopting preferred recharge uniformity on the basis of datasets B ($RMSE_h$ is 0.021 and 0.017 m for R_{120U} and R_{22U}) and C ($RMSE_h$ is 0.027 and 0.024 m for R_{120U} and R_{22U}). The smaller $RMSE_h$ values for R_{22U} compared to R_{120U} occur because the increase in $RMSE_h$ arising from the imposition of preferred uniformity is significantly greater for cases employing a large number of parameters where relatively few water-level data are available ($RMSE_h$ is 64% higher for R_{120U} relative to R_{120}, compared to 9% higher for R_{22U} relative to R_{22}, averaged across datasets B and C). Larger $RMSE_h$ values are also apparent where datasets containing fewer water-level targets are used ($RMSE_h$ is, on average, 0.0166, 0.0168 and 0.0242 m on the basis of dataset A, B and C, respectively).

Figure 4.4 shows time series of calibrated recharge values for Case 1, and reference-model recharge (R'). With a decreasing number of water-level targets (Figure 4.4, (a) to (c)), the general agreement between calibrated recharge and R' is reduced, as expected. This is evidenced by an RMSE in recharge ($RMSE_R$) of 28, 35 and 48 mm/y, on average, for calibrations employing datasets A, B and C, respectively.

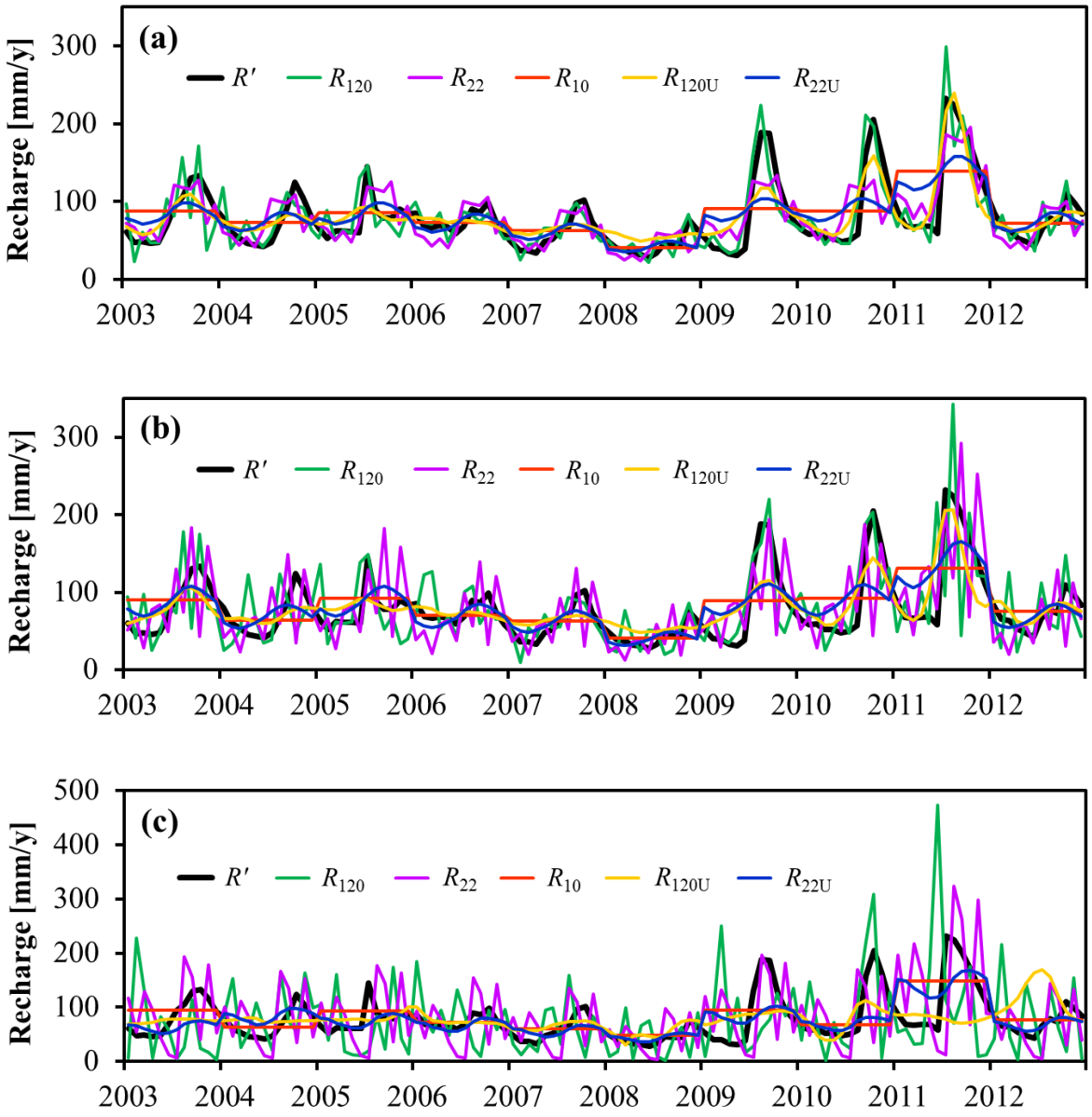

Figure 4.4. Time series of calibrated recharge values using datasets (a) A (11,280 water-level targets), (b) B (4980 targets), and (c) C (1222 targets), compared to that of *R'*, for Case 1. Note the different scale on the y-axis of (c).

The influence of the number of water-level targets on $RMSE_h$ and $RMSE_R$ is illustrated by Figure 4.5a. $RMSE_h$ and $RMSE_R$ values obtained on the basis of dataset C (shown in green) are, on average, located towards the top (i.e., higher $RMSE_R$) and to the right (i.e., higher $RMSE_h$) of Figure 4.5a, whereas $RMSE_h$ and $RMSE_R$ values for dataset A (shown in blue) are, on average, located towards the

bottom (i.e., lower $RMSE_R$) and left (i.e., lower $RMSE_h$) of Figure 4.5a. Relatively accurate estimates of recharge are obtained during periods in which water-level targets are present (e.g., on average, $RMSE_R$ is 35 mm/y for times at which water-levels are present, compared to 52 mm/y for times at which no water-levels are present, where calibration dataset C is employed; Figure 4.4c).

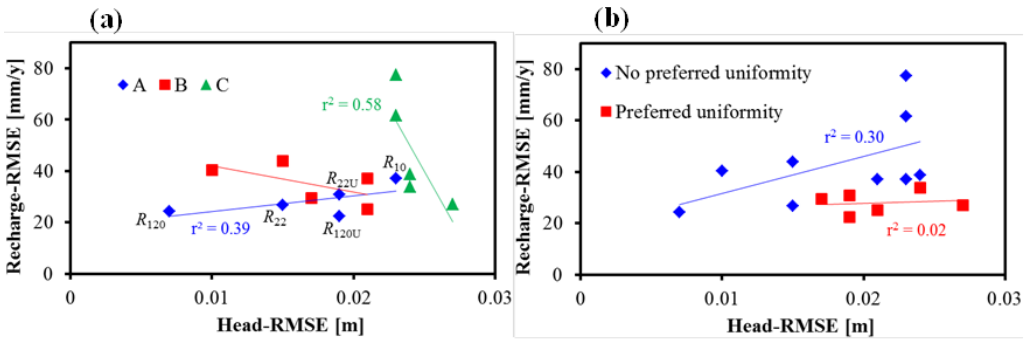

Figure 4.5. Scatterplots (and regression lines-of-best-fit) of recharge-RMSE versus head-RMSE values, grouped according to (a) the different calibration datasets (A, B and C) and (b) the use or non-use of preferred recharge uniformity, for Case 1.

The following observations can be made from calibrated recharge values where dataset A is used (Figure 4.4a). The most distinct features of the R' time series are best represented by calibrated R_{120} values (e.g., R' peaks during 2009 and 2010 are captured by only R_{120}). However, the R_{120} time series displays non-physical temporal recharge variability, in the oscillating behaviour most evident in 2003, 2004 and 2011. The preferred temporal recharge uniformity that was imposed in the R_{120U} calibration removes these oscillations, but subsequently results in underestimation of peak values in the R' time series. The $RMSE_R$ for R_{120U} (22 mm/y) is therefore only slightly smaller than that for R_{120} (24 mm/y). The R_{22} time series also suffers from recharge-value oscillation (displaying smaller amplitudes, but greater frequencies relative to R_{120}), and fails to represent the timing of R' events (e.g., there is an offset

in the peaks of the R_{22} time series, that sometimes precedes or lags R'). However, the average-annual and average-monthly variability of R' is generally well-represented by R_{22} ($RMSE_R$ values of 5 and 18 mm/y, respectively). Preferred uniformity produces a smoother time series (R_{22U} versus R_{22}) and thereby eliminates the artificial oscillations, but in doing do creates a weaker match to the original values ($RMSE_R$ for R_{22U} is 31 mm/y, compared to 27 mm/y for R_{22}). While calibrated R_{10} values provide a reasonable match to average annual R' values ($RMSE_R$ of 6 mm/y), the R_{10} time series is simply unable to match the dynamics of the R' time series ($RMSE_R$ of 37 mm/y over the whole time-series).

Collectively, these results indicate that smaller $RMSE_R$ values are obtained where a larger number of adjustable parameters are employed, as expected, due to the relationship between the degrees of freedom within the calibration and the level of model head fit attainable, the latter of which is related to recharge error for Case 1, given that S_y and K are known (i.e., recharge is the only unknown). This is illustrated by the positive correlation between $RMSE_h$ and $RMSE_R$ values for dataset A (shown in blue), with a regression line-of-best-fit coefficient of determination (r^2) value of 0.39 (Figure 4.5a). The general inverse relation between the RMSE, in terms of both head and recharge, and the number of parameters is also illustrated in Figure 4.5a. Data points pertaining to cases with few parameters are generally located toward the upper-right of Figure 4.5a, whereas cases employing a relatively large number of parameters are generally located toward the bottom-left (see marker labels for case descriptors).

Where dataset B is used for calibration (Figure 4.4b), recharge-error trends between different parameterisations are similar to those identified on the basis of dataset A, notwithstanding the generally larger errors obtained where fewer water-level targets are used. In particular, the $RMSE_R$ value for R_{120} (40 mm/y) is smaller than that of R_{22} (44 mm/y), and the $RMSE_R$ value for R_{120U} (25 mm/y) is smaller than that of R_{22U} (29 mm/y), which further highlights the relationship between the $RMSE_R$ and the number of adjustable parameters employed. However, the $RMSE_R$ for R_{10} (39 mm/y) is smaller than that of both R_{120} and R_{22}, despite the higher $RMSE_h$ obtained for R_{10}, due to the non-physical oscillation in recharge shown by R_{120} and R_{22}. This is illustrated by Figure 4.5a, which shows that the $RMSE_h$ versus $RMSE_R$ regression line for dataset B displays a negative correlation with an r^2 value of 0.36 (shown in red).

Where dataset C is used (Figure 4.4c), the relationship between $RMSE_R$ and the number of parameters is also apparent for cases adopting preferred uniformity, i.e., $RMSE_R$ for R_{120U} (27 mm/y) is smaller than that for R_{22U} (34 mm/y); however, this relationship is not apparent for cases without preferred uniformity on the basis of dataset C, as evidenced by the $RMSE_R$ for R_{120} (77 mm/y) exceeding that of R_{22} (62 mm/y), which exceeds that of R_{10} (39 mm/y). This is illustrated by the $RMSE_h$ versus $RMSE_R$ regression line for dataset C, which displays a correlation with a negative slope ten times larger than that on the basis of dataset B, and an r^2 value of 0.58 (shown in green; Figure 4.5a). Noteworthy is that the differences in $RMSE_R$ for R_{10} is relatively small across the different datasets (37, 37 and 39 mm/y for datasets A, B and C, respectively), which highlights that models with fewer parameters have lower data-requirement needs.

The benefit of preferred uniformity in terms of recharge estimation is significantly larger for datasets B and C compared to dataset A, as illustrated by changes in $RMSE_R$ values (positive values indicate reduction in $RMSE_R$ afforded by preferred uniformity), for R_{120U} versus R_{120}, of 15, 50, and 2 mm/y, and, for R_{22U} versus R_{22}, 14, 28, and -4 mm/y (this negative value indicates that preferred uniformity does not improve the match of R_{22} to R' on the basis of dataset A) for datasets B, C and A, respectively. This is because of the inadequate number of water-level targets within datasets B and C to constrain recharge, which, without preferred uniformity, results in oscillatory values. Conversely, where an adequate number of water-levels are used for calibration, the preferred uniformity is of less or no benefit to the estimation of recharge. The increase in $RMSE_R$ for R_{22U}, compared to R_{22}, on the basis of dataset A occurs because the imposition of preferred uniformity counteracts the expression of information within water-level targets. The influence of preferred uniformity on $RMSE_h$ and $RMSE_R$ values is shown in Figure 4.5b. Relative to the cases without preferred uniformity (shown in blue), the cases that employ preferred uniformity (shown in red) are clustered toward the lower-right part of Figure 4.5b (i.e., preferred uniformity generally produces lower $RMSE_R$ values and higher $RMSE_h$ values).

Figure 4.6 shows the identifiability of recharge parameters for R_{120U}, R_{22U} and R_{10}. The identifiability of monthly recharge parameters (R_{120U}) varies between 0.02 and 1.0 (Figure 4.6a). Of the 120 recharge parameters, 34 are considered to be identifiable (as indicated by identifiability values >0.8) on the basis of dataset A, whereas only 12 and 11 are considered identifiable on the basis of datasets B and C, respectively. The identifiability of average-monthly and yearly recharge parameters

(R_{22U}) is equal to 1.0, regardless of the dataset adopted (Figure 4.6b). The identifiability of yearly recharge parameters (R_{10}) is equal to 1.0 where datasets A and B are used, and varies between 0.45 and 1.0 where dataset C is used (Figure 4.6c). Clearly, recharge parameters that represent larger time scales (R_{22U} and R_{10}) generally display higher identifiability. This is because the amount of information within the calibration dataset pertaining to each recharge parameter is larger in a relative sense as the temporal parameterisation resolution is reduced. Identifiability values are also shown to be inversely related to R' variability, to a decreasing degree with fewer water-level targets. This is illustrated by the R_{120U} identifiability-versus-R' scatterplots, which display regression-line r^2 values of 0.23, 0.17 and 0.02 where datasets A, B and C are used, respectively (Figure 4.6d).

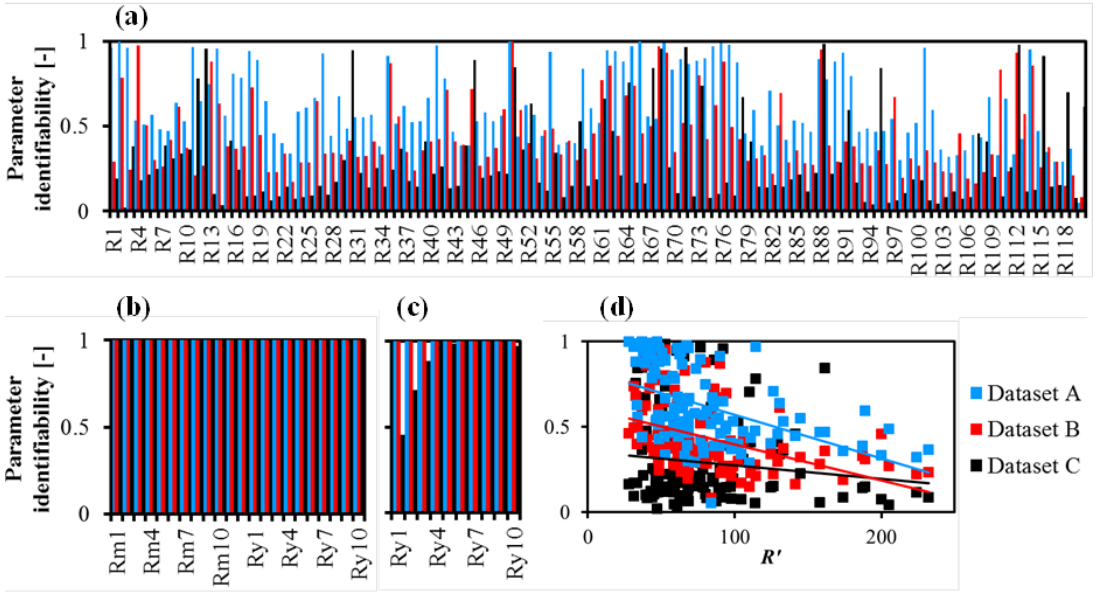

Figure 4.6. Identifiability of recharge parameters on the basis of datasets A, B and C for (a) R_{120U}, (b) R_{22U} and (c) R_{10}, and (d) scatterplot (and regression lines-of-best-fit) of R_{120U}-identifiability versus R', for Case 1. The regression lines display r^2 values of 0.23, 0.17 and 0.02 for datasets A, B and C, respectively.

4.5.2. Case 2

Only the recharge parameterisations that include preferred uniformity (R_{120U} and R_{22U}) are evaluated in the following given that they generally display superior performance in Case 1 (in terms of recharge-estimate accuracy) relative to those that do not employ preferred uniformity (R_{120}, R_{22} and R_{10}). Additionally, the benefit afforded by preferred uniformity in terms of recharge estimation (i.e., avoidance of oscillatory values) is expected to be larger where hydraulic parameters are being estimated simultaneously (Cases 2 and 3).

The degree of head misfit is considered to be low for all Case 2 calibrations ($RMSE_h$ values <0.03 m). $RMSE_h$ values range between 0.013 and 0.027 m. In a similar manner to the results of Case 1, larger $RMSE_h$ values are apparent where fewer recharge parameters are estimated, as expected, on the basis of dataset A, but not on the basis of datasets B and C. This is because the increase in $RMSE_h$ values due to the use of preferred uniformity is significantly greater where a large number of parameters are being estimated on the basis of relatively few water-level targets. Larger $RMSE_h$ values are also apparent where fewer water-levels are used ($RMSE_h$ is, on average, 0.016, 0.021 and 0.027 m for datasets A, B and C, respectively). The $RMSE_h$ values obtained for four cases (R_{120}, dataset A; R_{22U}, dataset A; R_{120U}, dataset B; R_{120U}, dataset C) are below or equal to those of Case 1. The $RMSE_h$ values of the two remaining cases (R_{22U}, dataset B and R_{22U}, dataset C) for Case 2 (0.021 and 0.026 m) are larger than those of Case 1 (0.017 and 0.024 m).

Figure 4.7a-c compares the time series of calibrated recharge values for Case 2 to R'. Similar to the results obtained for Case 1, the agreement between calibrated recharge

and R' values generally diminishes with a decreasing number of water-level targets (Figure 4.7, (a) to (c)). This is evidenced by $RMSE_R$ values of 25, 27 and 40 mm/y, on average, where datasets A, B and C are used, respectively. The variability-in-time of recharge estimate accuracy is related to the water-level sampling interval, in agreement with the results of Case 1.

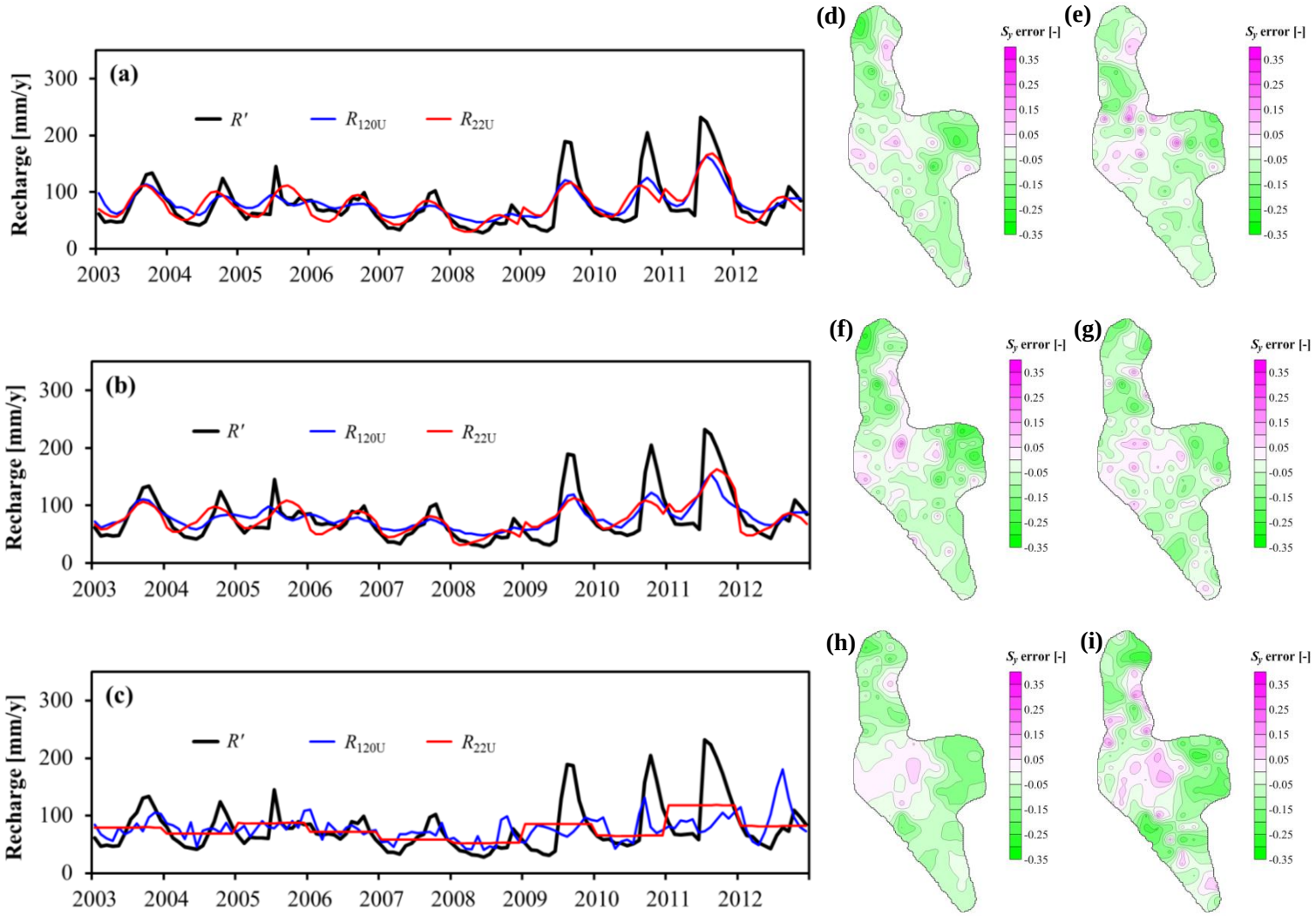
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
Recharge [mm/y]
300
200
100
0
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
R′
R120U
R22U
Sy error [-]
0.35
0.25
0.15
0.05
-0.05
-0.15
-0.25
-0.35

Figure 4.7. (pg. 126) Time series of calibrated recharge values using datasets (a) A, (b) B, and (c) C, compared to that of R', and S_y error distributions with dataset A for (d) R_{120U} and (e) R_{22U}, dataset B for (f) R_{120U} and (g) R_{22U}, and dataset C for (h) R_{120U} and (i) R_{22U}, for Case 2.

The superior performance of R_{120U} relative to R_{22U} on the basis of datasets A and B is consistent with the results of Case 1 (Figure 4.7a and b). That is, R_{120U} yields lower $RMSE_R$ values (24 and 26 mm/y) compared to those of R_{22U} (26 and 27 mm/y) for datasets A and B, respectively, due to the greater ability of R_{120U} to capture the variability in R'. However, in contrast to the results of Case 1 is the lesser performance of R_{120U} compared to R_{22U} for dataset C, as evidenced by the $RMSE_R$ value for R_{120U} (41 mm/y) exceeding that of R_{22U} (38 mm/y) (Figure 4.7c). This is because of the failure of R_{120U} to accurately reproduce the variability of R' (due to the sparse water-level sampling of dataset C), despite that R_{22U} values do not display monthly variability, and that only annual recharge variability is well-represented ($RMSE_R$ of 10 mm/y) (due to both the sparse water-level record of dataset C and the use of preferred uniformity).

The match between R_{120U} and R' for Case 2 is weaker than that of Case 1 where datasets A, B and C are used ($RMSE_R$ values of 24 mm/y (dataset A), 26 mm/y (dataset B) and 41 mm/y (dataset C) for Case 2, compared to 22 mm/y (dataset A), 25 mm/y (dataset B) and 27 mm/y (dataset C) for Case 1). The match between R_{22U} and R' for Case 2 is also weaker than that of Case 1 on the basis of dataset C ($RMSE_R$ is 38 mm/y for Case 2 compared to 34 mm/y for Case 1). Larger $RMSE_R$ values for Case 2 are a result of the compensatory roles assumed by S_y parameters. That is, spurious S_y values are obtained in offsetting errors in calibrated recharge values to maintain an equivalent level of model-head fit. For example, the variability

in R_{120U} is less than that of R' for dataset A (e.g., the standard deviation of R_{120U} values is 22 mm/y, compared to that of R' of 42 mm/y), and to compensate for this, calibrated S_y values are, on average, lower than those of the reference model (average S_y error of -0.07), as illustrated by the dominance of negative values in Figure 4.7d (shown in green). For other cases (i.e., where R_{22U} is estimated on the basis of datasets A and B), the match between calibrated recharge and R' values for Case 2 is stronger than that of Case 1, as evidenced by $RMSE_R$ values of 26 mm/y (dataset A) and 27 mm/y (dataset B) for Case 2, and 31 mm/y (dataset A) and 29 mm/y (dataset B) for Case 1. This occurs because of the inclusion of preferred S_y-homogeneity constraints into the regularisation objective function for Case 2, which reduces the extent to which recharge uniformity constraints are imposed relative to Case 1, thereby allowing for enhanced representation of R' peaks and troughs in Case 2 (e.g., the standard deviation of R_{22U} values for Case 2 (28 mm/y) exceeds that for Case 1 (25 mm/y)).

The spatially averaged errors in calibrated S_y values, similar to those of recharge values, are larger with a diminishing number of water-level targets (e.g., S_y-RMSE values of, on average, 0.089, 0.098 and 0.107 where datasets A, B and C are used, respectively; Figure 4.7d-i). Calibrated $RMSE_R$ and S_y-RMSE values display a reasonable correlation, as evidenced by the $RMSE_R$-to-S_y-RMSE regression line r^2 of 0.39. This is a consequence of non-uniqueness, whereby the attainment of a sufficient level of model head fit through the process of calibration requires that only the ratio of recharge and S_y be sufficiently accurate (i.e., heads can be matched with recharge and S_y values that are proportionately correct, but individually in error).

Figure 4.8 shows the identifiability of recharge and S_y parameters for Case 2. Only eight and four of the R_{120U} parameters are considered identifiable (>0.8) on the basis of datasets A and B, respectively, whereas none are considered identifiable on the basis of C (Figure 4.8a). Of the R_{22U} parameters, 22, 20 and 9 are considered identifiable where datasets A, B and C are used, respectively (Figure 4.8b). The annual recharge parameters within R_{22U} display higher identifiabilities than monthly parameters, highlighting that parameters which represent larger time scales display larger identifiability. Recharge identifiability values of Case 2 are, on average, less than those of Case 1 (0.54 and 0.42, respectively), due to the simultaneous estimation of S_y on the basis of the same datasets. The identifiability of spatially distributed S_y parameters (from R_{120U} and R_{22U} calibrations) varies between zero and 0.95, zero and 0.83, and zero and 0.05 for the calibrations employing dataset A, B, and C, respectively (Figure 4.8(c) and (d)). A total of just three, one and zero of the S_y parameters are considered identifiable on the basis of dataset A, B and C, respectively.

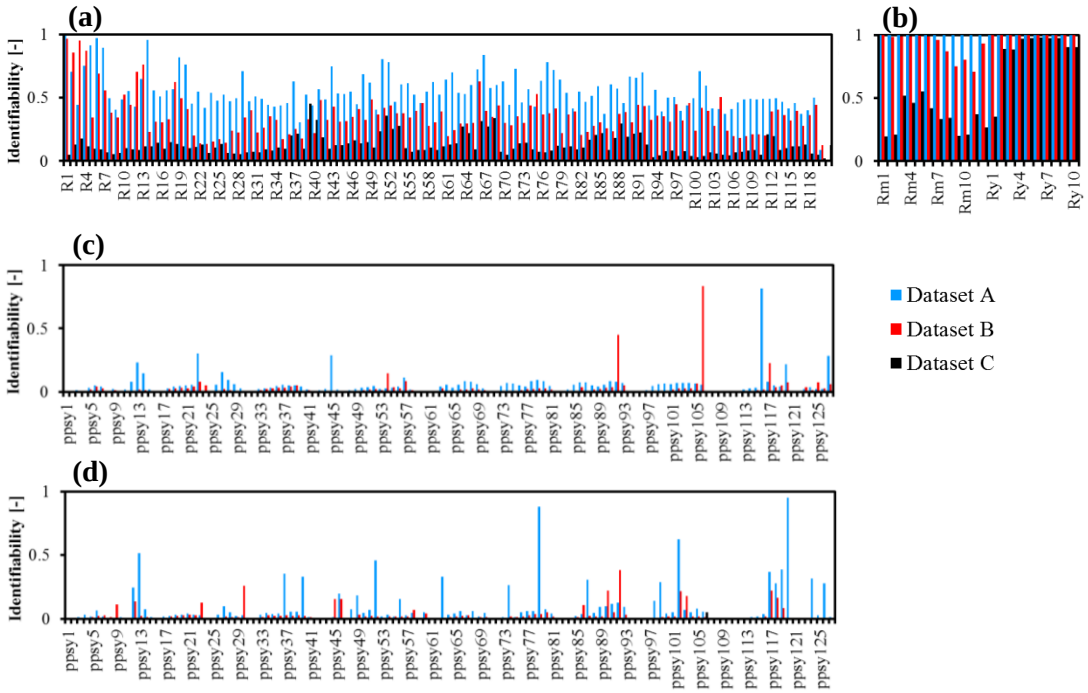

Figure 4.8. Identifiability of recharge for (a) R_{120U} and (b) R_{22U}, and S_y parameters for (c) R_{120U} and (d) R_{22U} on the basis of datasets A, B, and C, for Case 2.

4.5.3. Case 3

The degree of head misfit is considered to be sufficiently low only for the case where R_{120U} and dataset A are used for Case 3. A Φ_m value of 42 m^2 is obtained for this case, which is made up of temporal head-difference misfits (12 m^2), and time-averaged water-level misfits (30 m^2). The corresponding head-difference and average water-level $RMSE_h$ values are 0.033 and 0.006 m, respectively. This head-difference $RMSE_h$ is considerably larger than those on the basis of R_{120U} and dataset A for Case 1 (0.019 m) and Case 2 (0.013 m), as a result of the inclusion of time-averaged water-levels in the calibration dataset. Much larger misfits are obtained for the other cases (head-difference $RMSE_h$ values range from 0.094 to 0.222 m). A number of different relative weights between average-head and head-difference targets were trialled (1, 5, 10, 50) in an attempt to achieve a better match to head differences, but

were ultimately unsuccessful. We therefore evaluate only the results of the aforementioned case herein.

Figure 4.9a shows the R_{120U} time series on the basis of dataset A for Case 3 and R'. The match between R_{120U} and R' for Case 3 is significantly weaker than those of Cases 1 and 2 (that employ R_{120U} and dataset A). This is evidenced by a recharge-RMSE value of 37 mm/y for Case 3, compared to those obtained for Case 1 (22 mm/y) and Case 2 (24 mm/y). This highlights the compensatory roles assumed by hydraulic parameters in offsetting errors in recharge, similar to the results of Case 2. A larger degree of compensation is apparent in Case 3 relative to Case 2. This is evidenced by the underestimation of spatially averaged K (-404 m/d; Figure 4.9c), the average (in time) recharge error of -15 mm/y, the underestimation of spatially averaged S_y (-0.044; Figure 4.9b), and the R_{120U}-variability standard deviation of 24 mm/y below that of R'.

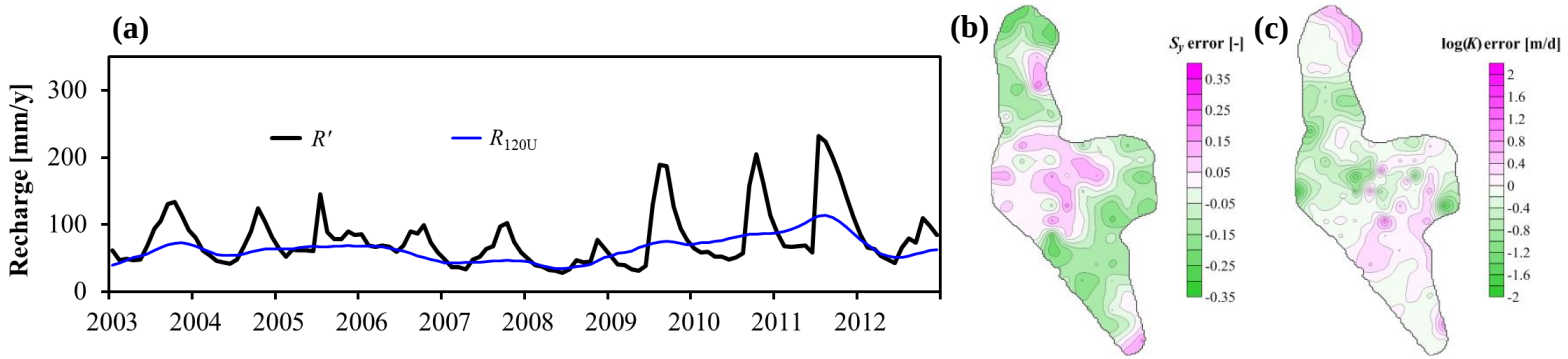

Figure 4.9. (a) Time series of R_{120U} recharge, compared to that of R', (b) S_y error distribution, and (c) K error distribution, where R_{120U} and dataset A are used for Case 3.

Figure 4.10 shows the identifiability of recharge and hydraulic parameters for Case 3. Eight R_{120U} parameters are deemed identifiable (Figure 4.10a), compared to none of the S_y parameters (note a maximum value of 0.79 is obtained; Figure 4.10b), and 79 of the K parameters (Figure 4.10c). An identifiability value of 0.84 is obtained for the C parameter. The identifiability of R_{120U} parameters for Case 3 is generally smaller than those for Cases 1 and 2 (on the basis of dataset A), as evidenced by the average recharge identifiability of 0.49 for Case 3, compared to 0.63 and 0.55 of Case 1 and 2, respectively. This is expected, given that recharge is jointly estimated with both S_y and K (and C) in Case 3. S_y-identifiability values are, on average, approximately equal (0.07) for Cases 2 and 3. The spatial variability in identifiability values of S_y and K is related primarily to the QL aquifer geometry, i.e., highest S_y-identifiabilities are located in close proximity to model boundaries, whereas high K-identifiability values are located across the basin, except in proximity to unsaturated regions of QL sediments.

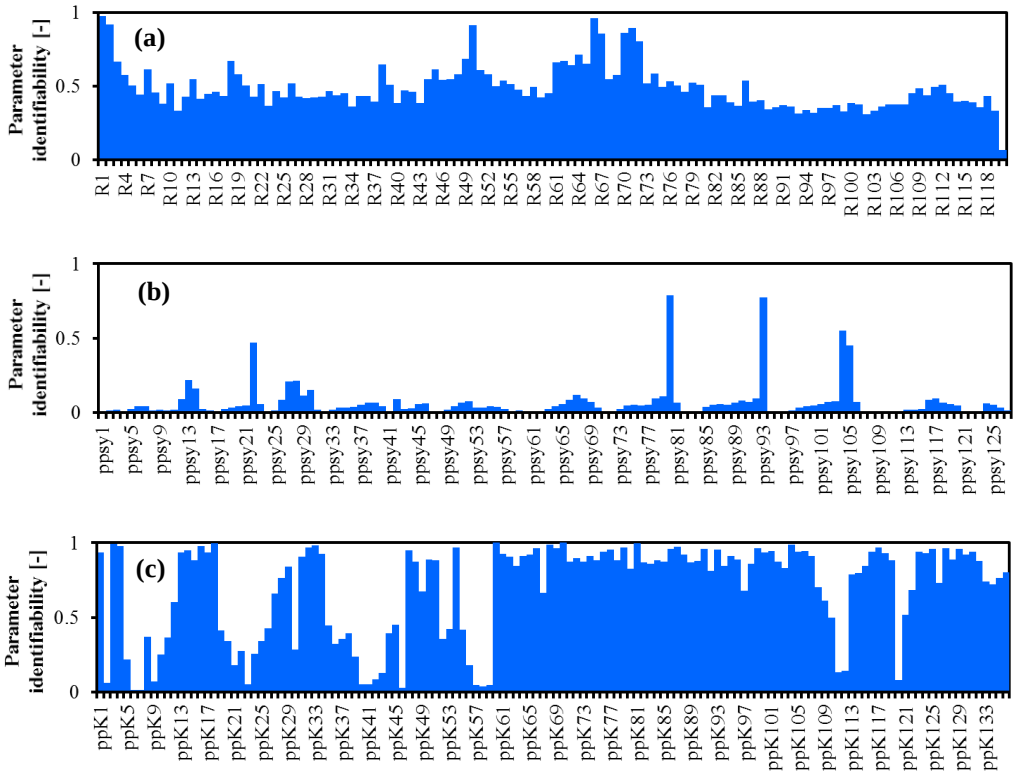

Figure 4.10. Identifiability of (a) R_{120U}-recharge, (b) S_y, and (c) K parameters on the basis of dataset A for Case 2.

4.6. Discussion

A significant correlation is evident between the level of model fit (i.e., head-RMSE) and the accuracy of recharge estimates among cases where recharge and S_y are estimated jointly, as evidenced by an r^2 value of 0.78 for a regression line representing head-RMSE versus recharge-RMSE. Such a correlation is expected only for cases where hydraulic parameters are known (i.e., where recharge is the only unknown), as demonstrated by the head-RMSE versus recharge-RMSE regression-line for Case 1, displaying a moderate r^2 value of 0.39 (where dataset A is used) (Figure 4.5). This correlation (for Case 2) suggests that the level of model fit obtained via calibration is related to the accuracy of parameter estimates, which is in contrast to previous studies that show that the level of model fit (especially those

where the calibration dataset is comprised of heads only) is not a proxy for accurate parameter estimates (e.g., McKenna et al., 2003; Pool et al., 2015; Knowling and Werner, 2016). This indicates that the constraints imposed through use of preferred-S_y homogeneity and preferred recharge uniformity are adequate in guiding the calibration process toward reference-model parameter values. Correlation between head and recharge errors is not expected for other cases where different parameter types are being estimated jointly and where regularisation constraints are less informed.

The temporal variability of recharge is shown to be, in practical terms, of limited estimability through calibration (>22% recharge-RMSE), even for cases where all spatially distributed hydraulic parameters are known and where monthly transient water-level data are available for calibration (which can be considered as "best-case" scenarios) (RMSE values of 36% on average; Figure 4.4a). The estimability of recharge is, as expected, lower still where spatially variable S_y values are simultaneously estimated (>24% recharge-RMSE). Larger recharge-RMSE values for joint recharge-and-S_y estimations exemplify the compensatory roles assumed by S_y parameters (e.g., where calibrated recharge variability in time is underestimated, calibrated S_y values are lower than those of the reference model, on average; Figure 4.7d), as supported by the lower recharge identifiability values obtained for the joint estimations (19% of recharge parameters are considered identifiable for joint estimations, compared to 29% of recharge parameters for recharge-only estimations). These are somewhat unsatisfying results, and have important implications for model-based groundwater management decision-making. For example, these results highlight the need for management decisions to be made on the basis of a suite of

models that collectively account for the uncertainty associated with model parameters, rather than a single, minimum error variance parameter set (e.g., Tonkin and Doherty, 2009; Doherty, 2015), which, as shown here, is likely to exhibit erroneous parameter values due to non-uniqueness, even where extensive observation datasets and unrealistic levels of parameter knowledge are present.

The amount of transient water-level data used for calibration in some of the cases (e.g., 120 consecutive monthly targets) is at the upper limit of what can be considered reasonable from a practical perspective (e.g., monthly water-level measurements spanning >10 years have been reported by Ahmadi and Sedghamiz (2007), Shamsudduha et al. (2009), Mack et al. (2013), Cao et al. (2013)). However, assumptions of known hydraulic parameter distributions (*K* for Case 2, and S_y and *K* for Case 3) constitute less realistic scenarios. The requirement for well-constrained parameter (in particular *K*) distributions, as evident from the results of the current study, poses a significant challenge. For example, the estimation of spatially distributed *K* through calibration of, e.g., steady-state groundwater models for the purpose of subsequently estimating recharge through calibration of transient models is limited by the need for the spatial distribution of recharge to be sufficiently constrained. This highlights the importance of alternative (e.g., field-based) means to estimate spatial *K* distributions.

Both the application of low-resolution (e.g., annual) recharge parameters and preferred recharge uniformity constraints constitute lumping mechanisms, and are therefore both of benefit to the estimation of recharge where a relatively small amount of water-level data is available. This is because of the greater number of

water-level observations on which time-varying recharge estimates are based, as evidenced by, e.g., high identifiability values for annual recharge parameters (Figure 4.5). In particular, the preferred uniformity constraint appears to have the effect of bridging the gap between sparse (in time) water-levels (while also avoiding oscillatory values; Figure 4.4). Conversely, both of these mechanisms are generally of detriment to recharge estimation where a relatively large amount of water-level data is available and where S_y and K are known. This is because of the reduced extent to which information contained within the water-level dataset can be expressed due to the insufficient flexibility within these parameterisation schemes. In general terms, these results are consistent with previous studies that demonstrate the lower data-requirement needs of models with simpler parameterisation schemes (e.g., Jakeman and Hornberger, 1993; Hill, 2006).

The insufficient level of head fit obtained for five of the six cases where K is jointly estimated along with recharge and S_y highlights that the inclusion of time-averaged water-level targets extends significantly the calibration objective function, and the challenges facing the process of parameter estimation where excessive parameter correlation exists (e.g., Doherty, 2015). Nevertheless, for the case where a sufficient head fit is obtained (involving all water-level data and preferred uniformity), the estimability of time-varying recharge via the joint estimation of recharge, S_y and K is significantly reduced relative to cases where K, and S_y and K, are known (recharge-RMSE value is 68% and 54% larger, respectively; Figure 4.9). This is in accordance with the larger degree of parameter compensation apparent for Case 3 (e.g., the underestimation of both K (-404 m/d) and S_y (-0.044) in conjunction with the

underestimation of time-averaged recharge (-15 mm/y) and recharge variability in time).

The extent to which time-varying recharge can be reliably estimated with respect to the factors discussed above is expected to be dependent on initial recharge and hydraulic parameter values, following the results of Knowling and Werner (2016). The importance of initial parameter values within traditional groundwater model calibration problems, i.e., in constraining the gradient-based optimisation problem where multiple objective function minima are present (e.g., Hill and Tiedeman, 2007), has been demonstrated by many studies (e.g., Bravo et al., 2002; Bahremand and De Smedt, 2008; Kannan et al., 2008).

The analysis of parameter identifiability provides a means to investigate the level of non-uniqueness, and its ramifications in terms of parameter estimability, without the need to evaluate parameter values directly. While identifiability values indicate the degree to which the calibration process can reduce the potential for wrongness in estimates of recharge in general terms, the identifiability values presented in the present study can also be considered a proxy for post-calibration parameter uncertainty given the assumption of negligible measurement error due to the synthetic modelling approach adopted (e.g., Moore and Doherty, 2005). The moderate inverse correlation between recharge identifiability and reference-model recharge where hydraulic parameters are known (in particular where a sufficient amount of data is used for calibration; Figure 4.6d) indicates that large recharge events are relatively difficult to identify through calibration, as differentiated from

small events, where groundwater-level responses are less dynamic, allowing for better identification of recharge.

A limitation of the present study is that the spatial variability of recharge (of both the reference model and the models undergoing calibration) is fixed in time. This assumption is considered to be necessary here given that the evaluation of calibrated recharge variability in both time and space would require an impractical amount of computational time. For example, the estimation of 125 pilot-point recharge parameters (as considered in Knowling and Werner (2016)), at each of the 120 monthly stress periods considered in the current study, would alone require 15,000 model runs per optimisation iteration (using two-point derivatives). This number of model runs per iteration is approximately 40 times larger than that of Case 3, which required a computational time considered to be excessive (approximately five days, using eight Intel i7-2600 CPUs running at 3.40 GHz, with a 64-bit operating system and 16.0 GB RAM). While the assumption of a constant spatial distribution of recharge in time does not hold in practice, it nonetheless allows for the estimability of the time component of recharge to be evaluated. Future work considering both spatially and temporally variable recharge estimation would benefit from the use of computationally efficient optimisation methodologies (e.g., hybrid methods that combine the computational savings afforded by subspace regularisation techniques with the increased stability achieved through the imposition of Tikhonov regularisation; Tonkin and Doherty, 2005).

The use of a synthetic reference model allows for the calibration process to be evaluated in terms of its ability to inform parameters with respect to error. However,

estimated parameter values cannot be assessed in terms of error in practice; instead, parameter values must be evaluated in terms of whether or not they can be considered reasonable on the basis of field data (e.g., Langevin and Zygnerski, 2013; Knowling et al., 2015). Such assessments are complicated by the need to consider, e.g., the scale-dependency of hydraulic properties (e.g., Sánchez-Vila et al., 2006).

The findings of this study are likely to be somewhat dependent on the unique characteristics of the USB reference model (e.g., the lack of surface water-groundwater interaction). The application of similar analyses to field sites located in, e.g., temperate climates where surface water and groundwater are often hydraulically connected, may indicate that the degree to which recharge can be informed is enhanced given the availability of stream discharge data (Sanford, 2002; Hunt et al., 2006). The investigation of recharge estimability for other field sites is warranted in order to extend the present findings to a wider range of situations, in particular where groundwater flux measurements are readily attainable.

4.7. Conclusions

The current study extends the steady-state investigation of Knowling and Werner (2016) by evaluating the extent to which time-varying recharge can be informed through calibration of field-scale transient groundwater models. The methodology adopted involves the use of a series of highly parameterised inverse modelling experiments that contain (1) varying degrees of hydraulic parameter (S_y, K) constraints, (2) different numbers of water-level calibration targets, and (3) different temporal recharge parameterisation schemes. The analysis involves the use of a

synthetic reality (i.e., a reference model), based on a transient groundwater model of USB.

Results show that the estimability of transient recharge is highly dependent on the degree to which the spatial distribution of S_y and K are constrained, the number of water-level targets available for calibration, and the temporal recharge parameterisation scheme adopted (e.g., temporal parameter resolution). Reasonable estimates of recharge variability in time (<30% recharge-RMSE) are shown to require, for the field-scale groundwater model applied, a large amount of transient water-level calibration data, and that the spatial distribution of K is known. Even for cases representing a "best-case" scenario whereby a large amount of water-level data is available and the spatial distribution of both S_y and K are known, estimates of time-varying recharge are of only moderate quality (>22% recharge-RMSE). The extent to which temporally variable recharge can be determined is significantly diminished where recharge, S_y and K are jointly estimated. The use of larger time scale recharge parameters and preferred recharge uniformity are shown to be of potential benefit to recharge estimation where water-level data are lacking.

This study exemplifies the potential ramifications of non-uniqueness in terms of recharge estimability, and offers insight into scenarios under which recharge can be reliably informed for practical real-world situations, thereby offering guidance to groundwater modelling practitioners. This study, in combination with Knowling and Werner (2016), suggests that the estimation of recharge through model calibration may be impractical for real-world settings, due to the requirement for excessive amounts of both water-level data, distributed in both space (for the purpose of

informing spatially distributed recharge, K and S_y) and time (for the purpose of informing time-varying recharge and S_y), and hydraulic parameter data. The findings presented here extend upon the preliminary benchmark of Knowling and Werner (2016) for evaluating the extent to which field-scale groundwater model calibration can inform recharge subject to practical data-availability limitations.

Acknowledgements

The author M. J. Knowling was supported by an Australian Postgraduate Award, and top-up scholarships from the Goyder Institute for Water Research and the National Centre for Groundwater Research and Training (NCGRT), a collaborative initiative of the Australian Research Council and the National Water Commission. The authors wish to acknowledge Michael Fienen whose comments helped improve the manuscript.

Chapter 5

5. Conclusions

This thesis focuses on the application of highly parameterised groundwater modelling strategies for addressing practical hydrology problems. The three studies in this thesis investigate in particular: (1) the quantification of climate and human impacts on regional aquifer settings; (2) the estimability of recharge through field-scale steady-state groundwater model calibration; and (3) the estimability of time-varying recharge through field-scale transient groundwater model calibration. The key findings of these studies are summarised below.

The first study extends previous modelling approaches for disentangling climate and human impacts through the application of a highly parameterised groundwater model. The USB is used as a case study following the conjecture surrounding the causal factors of aquifer depletion, i.e., climate variability and/or groundwater pumping. Results show that, while both climate and pumping impacts vary significantly both spatially and temporally, the contribution of pumping to aquifer depletion is 2.7 and 2.2 times that of climate, based on time-averaged and maximum-in-time impacts, respectively. Pumping impacts are shown to exceed climate impacts between 1978 and 2012, and over the majority (80%) of the spatial extent of USB. These results serve as a response to the recent South Australian parliamentary enquiry (NRC, 2013) into the cause of aquifer depletion. Future management of USB is expected to be enhanced through the adaption of pumping rates in alignment with

the condition of the aquifer in response to both climate and pumping impacts. This study demonstrates the application of a highly parameterised model, in combination with a relatively simple modelling framework, for quantifying the relative contribution of natural and human aquifer impacts.

The second study evaluates the extent to which recharge and its spatial distribution can be informed through calibration of field-scale steady-state groundwater models. The ability of the calibration process to inform recharge is shown to vary significantly with the degree to which hydraulic parameters are constrained. Inverse modelling results show that, for the highly parameterised model of USB, reasonable estimates of recharge require excessive amounts of hydraulic parameter information. For example, an average error of <10% requires that 50% of the reference-model hydraulic parameters (>100 *K* values across the 129 km^2 study area) be included in the calibration using preferred-value regularisation. While both SGD (as a calibration target) and pumping data are shown to be of benefit to the estimation of spatially averaged recharge, pumping data also produces improved estimates of the spatial variability in recharge. The findings of this study suggest that the estimation of recharge through calibration of steady-state models may be impractical for real-world settings, limited by the need for unrealistic amounts of groundwater level and hydraulic parameter data. This study provides guidance to groundwater modelling practitioners by highlighting the potential ramifications of non-uniqueness in terms of recharge estimability for a real-world case, and by offering insight into scenarios under which recharge may (or may not) be informed through calibration.

The third study assesses the extent to which the temporal variability of recharge can be informed through calibration of field-scale transient groundwater models. Results show that the estimability of time-varying recharge varies highly with the degree to which the spatial distribution of S_y and K are constrained, the number of water-level targets available for calibration, and the temporal recharge parameterisation scheme adopted (e.g., the temporal resolution on which recharge is being inferred). Reasonable estimates of recharge variability in time (<30% recharge-RMSE) are shown to require, for the transient groundwater model of USB, a considerable amount of transient water-level targets, and that the spatial distribution of K is known (i.e., via joint recharge-and-S_y estimation). The extent to which time-varying recharge can be informed is significantly diminished where K is estimated jointly with recharge and S_y. The use of preferred uniformity is shown to be of benefit to recharge estimation for cases where relatively few water-level calibration targets are available. This study, in combination with the second study, suggests that the estimation of recharge through model calibration may be impractical for real-world settings, due to the requirement for excessive amounts of spatially distributed and transient water-level data, in addition to hydraulic parameter data.

References

Alcolea, A., Carrera, J., Medina, A., 2006. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. Adv. Water Resour. 29: 1678-1689.

Ackerman, D.J., Rousseau, J.P., Rattray, G.W., Fisher, J.C., 2010. Steady-state and transient models of groundwater flow and advective transport, Eastern Snake River Plain aquifer, Idaho National Laboratory and Vicinity, Idaho. U.S. Geological Survey Scientific Investigations Report, 2010-5123, 236 pp.

Ahmadi, S.H., Sedghamiz, A., 2007. Geostatistical analysis of spatial and temporal variations of groundwater level. Environ. Monit. Assess. 129: 277-294, doi: 10.1007/s10661-006-9361-z.

Arnold, J.G., Muttiah, R.S., Srinivasan, R., Allen, P.M., 2000. Regional estimation of base flow and groundwater recharge in the Upper Mississippi basin. J. Hydrol. 227: 21-40, doi: 10.1016/S0022-1694(99)00139-0.

Aster, R.C., Borchers, B., Thurber, C.H., 2005. Parameter estimation and inverse problems. Elsevier Academic Press, Amsterdam, Netherlands, 301 pp.

Bahremand, A., De Smedt, F., 2008. Distributed hydrological modeling and sensitivity analysis in Torysa watershed, Slovakia. Water Resour. Manage. 22: 393-408, doi: 10.1007/s11269-007-9168-x.

Barnett, S.R., 1978. Eyre Peninsula groundwater survey – Uley South basin. Department of Mines and Energy, Report 78/22, Government of South Australia, Adelaide, 45 pp.

Bear, J., 1972. Dynamics of fluids in porous media. Elsevier, New York, USA, 764 pp.

Bestland, E.A., 2010. Geology of the sea cliffs along Uley South, Eyre Peninsula. A report for the Natural Resource Management Board, School of the Environment, Flinders University of South Australia, 36 pp.

Blasone, R-S., Madsen, H., Rosbjerg, D., 2008. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. J. Hydrol. 353: 18-32.

Boussinesq, J., 1904. Recherches théoriques sur l'écoulement des nappes d'eau infiltrées dans le sol et sur le débit des sources. J. Math. Pures Appl. 10: 5-78.

Bravo, H. R., Jiang, F., Hunt, R.J., 2002, Using groundwater temperature data to cotstrain parameter estimation in a groundwater flow model of a wetland system. Water resour. Res. 38(8): 1153, doi: 10.1029/2000WR000172.

Bresciani, E., Ordens, C.M., Werner, A.D., Batelaan, O., Guan, H., Post, V.E.A., 2014. Spatial variability of chloride deposition in a vegetated coastal area: Implications for groundwater recharge estimation. J. Hydrol. 519: 1177-1191, doi: 10.1016/j.jhydrol.2014.08.050.

Brunner, P., Doherty, J., Simmons, C.T., 2012. Uncertainty analysis and implications for data acquisition in support of integrated hydrologic models. Water Resour. Res. 48, W07513, doi: 10.1029/2011WR011342.

Buishand, T.A., 1982. Some methods for testing the homogeneity of rainfall records. J. Hydrol. 58: 11-27.

Bureau of Meteorology, 2010. SILO database. http://www.bom.gov.au/silo/. Cited 1 February 2010.

Burnett, W.C., Aggarwal, P.K., Aureli, A., Bokuniewicz, H., Cable, J.E., Charette, M.A., et al., 2006. Quantifying submarine groundwater discharge in the coastal

zone via multiple methods. Sci. Total Environ. 367: 498-543, doi: 10.1016/j.j.scitotenv.2006.05.009.

Burrows, W., Doherty, J., 2014. Efficient calibration/uncertainty analysis using paired complex/surrogate models. Ground Water 53(4), 531-541, doi: 10.1111/gwat.12257.

Cao, G., Zheng, C., Scanlon, B.R., Liu, J., Li, W., 2013. Use of flow modeling to assess sustainability of groundwater resources in the North China Plain. Water Resour. Res. 49(1): 159-175, doi: 10.1029/2012WR011899.

Candela, L., von Igel, W., Elorza, F.J., Aronica, G., 2009. Impact assessment of combined climate and management scenarios on groundwater resources and associated wetland (Majorca, Spain). J. Hydrol. 376: 510-527, doi: 10.1016/j.jhydrol.2009.07.057.

Carniato, L., Schoups, G., van de Giesen, N., Seuntjens, P., Bastiaens, L., Sapion, H., 2015. Highly parameterized inversion of groundwater reactive transport for a complex field site. J. Contam. Hydrol. 173: 38-58, doi: 10.1016/j.jconhyd.2014.12.001.

Carrera, J., Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady-state conditions. 2. uniqueness, stability, and solution algorithms. Water Resour. Res. 22(2): 211-227, doi: 10.1029/WR022i002p00211.

Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., Slooten, L.J., 2005. Inverse problem in hydrogeology. Hydrogeol. J. 13: 206-222, doi: 10.1007/s10040-004-0404-7.

Chemingui, A., Sulis, M., Paniconi, C., 2015. An assessment of recharge estimates from stream and well data and from a coupled surface-water/groundwater model for the des Anglais catchment, Quebec (Canada). Hydrogeol. J. 1-13, doi: 10.1007/s10040-015-1299-1.

Cheng, A. H. D., Ouazar, D., 1999. Analytical solutions. Ch. 6. In: Seawater intrusion in coastal aquifer – concepts, methods and practices. Eds. Bear, J., Cheng, A.H.D, Sorek, S., Ouazar, D., Herrera, I., Dordrecht, The Netherlands, Kluwer Academic Publishers.

Chiang, W-H., Kinzelbach, W., 1998. Processing MODFLOW – A simulation system for modelling groundwater flow and pollution. Hamburg, 342 pp. http://www.pmwin.net. Cited 5 May, 2012.

Christensen, S., Doherty, J., 2008. Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration. Adv. Water Resour. 31: 674-700.

Clark, M.P., Vrugt, J.A., 2006. Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters. Geophys. Res. Lett. 33, doi: 10.1029/2005GL025604.

Clarke, D.S., Berens, V., Dennis, K.J., 2003. Uley South – Coffin Bay observation well network review. Department of Water, Land and Biodiversity Conservation, Report 2003/04, Government of South Australia, Adelaide, 118 pp.

Cong, Z., Yang, D., Gao, B., Yang, H., Hu, H., 2009. Hydrological trend analysis in the Yellow River basin using a distributed hydrological model. Water Resour. Res. 45: W00A13, doi: 10.1029/2008WR006852.

Crosbie, R.S., Binning, P., Kalma, J.D., 2005. A time series approach to inferring groundwater recharge using the water table fluctuation method. Water Resour. Res. 41: W01008, doi: 10.1029/2004WR003077.

Custodio, E., 2002. Aquifer overexploitation: what does it mean? Hydrogeol. J. 10: 254-277.

Cuthbert, M.O., 2014. Straight thinking about groundwater recession. Water Resour. Res. 50, 2407-2424, doi: 10.1002/2013WR014060.

D'Agnese, F.A., Faunt, C.C., Hill, M.C., Turner, A.K., 1999. Death Valley regional ground-water flow model calibration using optimal parameter estimation methods and geoscientific information systems. Adv. Water Res. 22(8): 777-790, doi:10.1016/S0309-1708(98)00053-0.

Dausman, A.M., Doherty, J., Langevin, C.D., Dixon, J., 2010a. Hypothesis testing of buoyant plume migration using a highly parameterized variable-density groundwater model at a site in Florida, USA. Hydrogeol. J. 18: 147-160.

Dausman, A.M., Doherty, J., Langevin, C.D., Sukop, M.C., 2010b. Quantifying data worth toward reducing predictive uncertainty. Ground Water 48(5): 729-740.

Davis, A., Munday, T., Somaratne, N., 2013. Characterisation of a coastal aquifer system in the Eyre Peninsula, South Australia, using nuclear magnetic resonance methods. In: Wetzelhuetter, C. (Ed.), 2013, Groundwater in the coastal zones of Asia-Pacific, Coastal Research Library Vol. 7, Springer, Heidelberg, p. 89-120.

deGroote-Hedlin, C., Constable, S., 1990. Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data. Geophysics 55: 1613-1624, doi: 10.1190/1.1442813.

de Marsily, G.H., Lavedan, G. Boucher, M., Fasanino, G., 1984. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In: Verly, G. et al. (Eds.), 1984, Geostatistics for natural resources characterization, Part 2, D. Reidel Pub. Co., p. 831-849.

Demissie, Y.K., Valocchi, A.J., Minsker, B.S., Bailey, B.A., 2009. Integrating a calibrated groundwater flow model with error-correcting data-driven models to improve predictions. J. Hydrol. 364: 257-271.

Dickinson, J.E., Hanson, R.T., Ferré, T.P.A., Leake, S.A., 2004. Inferring time-varying recharge from inverse analysis of long-term water levels. Water Resour. Res. 40: W07403, doi: 10.1029/2003WR002650.

Diersch, H.-J.G., 2005. FEFLOW reference manual, WASY Ltd., Berlin, 292 pp.

Doherty, J., 2003. Ground water model calibration using pilot points and regularization. Ground Water 41: 170-177, doi: 10.1111/j.1745-6584.2003.tb02580.x.

Doherty, J., Skahill, B.E., 2006. An advanced regularization methodology for use in watershed model calibration. J. Hydrol. 327: 564-577, doi: doi:10.1016/j.jhydrol.2005.11.058.

Doherty, J., Hunt, R.J., 2009. Two statistics for evaluating parameter identifiability and error reduction. J. Hydrol 366: 119-127, doi: 10.1016/j.jhydrol.2008.12.018.

Doherty, J.E., Hunt, R.J., 2010. Approaches to highly parameterized inversion: A guide to using PEST for groundwater-model calibration. U.S. Geological Survey Scientific Investigations Report, 2010-5169, 60 pp.

Doherty, J.E., Hunt, R.J., Tonkin, M.J., 2010. Approaches to highly parameterized inversion: A guide to using PEST for model-parameter and predictive-uncertainty analysis. U.S. Geological Survey Scientific Investigations Report, 2010–5211, 71 pp.

Doherty, J.E., Fienen, M.F., Hunt, R.J., 2010. Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions. U.S. Geological Survey Scientific Investigations Report, 2010-5168, 36 pp.

Doherty, J.E., Hunt, R.J., Tonkin, M.J., 2011. Approaches to highly parameterized inversion: A guide to using PEST for model-parameter and predictive-uncertainty analysis. U.S. Geological Survey Scientific Investigations Report, 2010-5211, 82 pp.

Doherty, J., 2013. Manual and Addendum for PEST: Model-Independent Parameter Estimation. Watermark Numerical Computing, Brisbane, Australia.

Doherty, J., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. Watermark Numerical Computing, Brisbane, Australia.

Doherty, J., 2016. PEST User Manual, Part I and II. Watermark Numerical Computing, Brisbane, Australia.

Department for Water (DFW), 2012. Science Support for the Musgrave and Southern Basins Prescribed Wells Areas Water Allocation Plan, Technical Report DFW 2012/15, Government of South Australia, Adelaide, 126 pp.

Department of Environment, Water and Natural Resources (DEWNR), 2013. Online groundwater data from Water Connect, as part of DEWNR, Government of South Australia. https://www.waterconnect.sa.gov.au/Systems/GD/Pages/default.aspx. Cited 15 July, 2013.

Dripps, W.R., Bradbury, K.R., 2007. A simple daily soil–water balance model for estimating the spatial and temporal distribution of groundwater recharge in temperate humid areas. Hydrogeol. J. 15(3): 433-444, doi: 10.1007/s10040-007-0160-6.

El Yaouti, F., El Mandour, A., Khattach, D., Kaufmann, O., 2008. Modelling groundwater flow and advective contaminant transport in the Bou-Areg unconfined aquifer (NE Morocco). J. Hydro-Environ. Res. 2: 192-209.

Engl, H.W., Hanke, M., Neubauer, A., 1996. Regularization of inverse problems. Springer, New York, 321 pp.

Erdal, D., Cirpka, O.A., 2016. Joint inference of groundwater-recharge and hydraulic-conductivity fields from head data using the ensemble Kalman filter. Hydrol. Earth Syst, Sci., 20, 555-569, doi: 10.519/hess-20-555-2016.

Essaid, H.I., Cozzarelli, I.M., Eganhouse, R.P., Herkelrath, W.N., Bekins, B.A., Delin, G.N., 2003. Inverse modeling of BTEX dissolution and biodegradation at the Bemidji, MN crude-oil spill site. J. Contam. Hydrol. 67(1): 269-299, doi:10.1016/S0169-7722(03)00034-2.

Evans, S.L., 1997. Estimating long term recharge to thin, unconfined carbonate aquifers using conventional and environmental isotope techniques: Eyre Peninsula, South Australia. Incomplete MSc Thesis, Flinders University, Adelaide, 176 pp.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99: 143-162.

Eyre Peninsula Natural Resources Management Board (EPNRM), 2006. Eyre Peninsula Water Allocation Plans. http://epnrm.sa.gov.au/. Cited 5 March 2012.

Ferguson, I.M., Maxwell, R.M., 2012. Human impacts on terrestrial hydrology: climate change versus pumping and irrigation. Environ. Res. Lett. 7: doi: 10.1088/1748-9326/7/4/044022.

Fetter, C.W., 2001. Applied hydrogeology, 4th Ed. Prentice Hall, New Jersey, p. 598.

Fienen, M.N., Muffels, C.T., Hunt, R.J., 2009. On constraining pilot point calibration with regularization in PEST. Ground Water 47: 835-844, doi: 10.1111/j.1745-6584.2009.00579.x.

Fienen, M.N., Doherty, J.E., Hunt, R.J., Reeves, H.W., 2010. Using predictive uncertainty analysis to design hydrologic monitoring networks: Example applications from the Great Lakes water availability project: U.S. Geological Survey Scientific Investigations Report 2010-5159, 44 p.

Finsterle, S., Kowalsky, M.B., 2011. A truncated Levenberg-Marquardt algorithm for the calibration of highly parameterized nonlinear models. Comput. Geosci. 37(6): 731-738, doi: 10.1016/j.cageo.2010.11.005.

Fitzpatrick, A., Cahill, K., Munday, T., 2009. Informing the hydrogeology of Coffin Bay, South Australia, through the constrained inversion of TEMPEST AEM data. CSIRO, Bently, Western Australia, 159 pp.

Freeze, R.A., Cherry, J.A., 1979. Groundwater. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, USA.

Friedel, M.J., 2005. Coupled inverse modeling of vadose zone water, heat and solute transport: calibration constraints, parameter nonuniqueness, and predictive uncertainty. J. Hydrol. 312: 148-175, doi:10.1016/j.jhydrol.2005.02.013.

Ghose, D.K., Panda, S.S., Swain, P.C., 2010. Prediction of water table depth in western region, Orissa using BPNN and RBFN neural networks. J. Hydrol. 394: 296-304.

Gómez-Hernández, J.J., Gorelick, S.M., 1989. Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakance, and recharge. Water Resour. Res. 25(3): 405-419, doi: 10.1029/WR025i003p00405.

Gómez, A.A., Rodríguez, L.B., Vives, L.S., 2010. The Guarani Aquifer System: estimation of recharge along the Uruguay-Brazil border. Hydrogeol. J., 18: 1667-1684, doi: 10.1007/s10040-010-0630-0.

Gurwin, J., Lubczynski, M., 2005. Modeling of complex multi-aquifer systems for groundwater resources evaluation-Swidnica study case (Poland). Hydrogeol. J. 13: 627-639.

Hamed, K.H., 2008. Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis. J. Hydrol. 249: 350-363.

Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. MODFLOW-2000, the U.S. Geological Survey modular ground-water model – User guide to modularisation concepts and ground-water flow process. U.S. Geological Survey open-file report 00-92, Reston, 130 pp.

Harrington, N., Evans, S., Zulfic, D., 2006. Uley Basin groundwater modelling project. Volume 1: Project overview and conceptual model development. Department of Water, Land and Biodiversity Conservation, Report 2006/01, Government of South Australia, Adelaide, 116 pp.

Hashemi, H., R. Berndtsson, M. Kompani-Zare, and M. Persson (2013), Natural vs. artificial groundwater recharge, quantification through inverse modeling, Hydrol. Earth Syst. Sci., 17, 637-650, doi: 10.5194/hess-17-637-2013.

Hayley, K., Schumacher, J., MacMillan, G.J., Boutin, L.C., 2014. Highly parameterized model calibration with cloud computing: an example of regional flow model calibration in northeast Alberta, Canada. Hydrogeol. J., 22: 729-737, doi: 10.1007/s10040-014-1110-8.

Helsel, D.R., Hirsch, R.M., 2002. Statistical methods in water resources, Book 4, Hydrologic analysis and interpretation. U.S. Geological Survey, Virginia, USA, 510 pp.

Hendricks Franssen, H.-J., Stauffer, F., Kinzelbach, W. 2004. Joint estimation of transmissivities and recharges–application: stochastic characterization of well capture zones. J. Hydrol. 294: 87-102, doi:10.1016/j.jhydrol.2003.10.021.

Hendricks Franssen, H.-J., Brunner, P., Makobo, P., Kinzelbach, W., 2008. Equally likely inverse solutions to a groundwater flow problem including patter information from remote sensing images. Water Resour. Res. 44: W01419, doi: 10.1029/2007WR006097.

Hendricks Franssen, H.-J., Alcolea, A., Riva, M., Bakr, M., van der Wiel, N., Stauffer, F., Guadagnini, A., 2009. A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments. Adv. Water Res. 32(6): 851-872, doi: 10.1016/j.advwatres.2009WR007745.

Herckenrath, D., Langevin, C.D., Doherty, J., 2011. Predictive uncertainty analysis of a saltwater intrusion model using null-space Monte Carlo. Water Resour. Res. 47, W05504, doi: 10.1029/2010WR009342.

Heuvelmans, G., Louwyck, A, Lermytte, J., 2011. Distinguishing between management-induced and climatic trends in phreatic groundwater levels. J. Hydrol. 411: 108-119.

Hill, M.C., 1998. Methods and guidelines for effective model calibration, U. S. Geological Survey, Water-Resour. Investigat. Rep. 98–4005, Colorado, 91 pp.

Hill, M.C., 2006. The practical use of simplicity in developing ground water models. Ground Water, 44(6), 775-781, doi: 10.1111/j.1745-6584.2006.00227.x.

Hill, M.C., Tiedeman, C.R., 2007. Effective groundwater model calibration: With analysis of data, sensitivities, predictions, and uncertainty. John Wiley and Sons Inc., Hoboken, New Jersey.

Hill, M.C., 2010. Comment on ‘‘Two statistics for evaluating parameter identifiability and error reduction” by John Doherty and Randall J. Hunt. J. Hydrol. 380: 481-488, doi: 10.1016/j.jhydrol.2009.10.011.

Hougham, A.L., Moran, S.B., Masterson, J.P., Kelly, R.P, 2008. Seasonal changes in submarine groundwater discharge to coastal salt ponds estimated using ^{226}Ra and ^{228}Ra as tracers. Mar. Chem. 109: 268-278, doi: 10.1016/j.marchem.2007.08.001.

Hunt, R.J., Feinstein, D.T., Pint, C.D., Anderson, M.P. 2006. The importance of diverse data types to calibrate a watershed model of the Trout Lake Basin,

Northern Wisconsin, USA. J. Hydrol. 321: 286-296, doi: 10.1016/j.jhydrol.2005.08.005.

Hunt, R.J., Doherty, J., Tonkin, M.J., 2007. Are models too simple? Arguments for increased parameterisation. Ground Water 45: 254-262.

Hunt, R.J., Luchette, J., Schreuder, W.A., Rumbaugh, J.O., Doherty, J., Tonkin, M.J., Rumbaugh, D.B., 2010. Using a cloud to replenish parched groundwater modelling efforts. Ground Water 48(3): 365-365, doi: 10.1111/j.1745-6584.2010.00699.x.

Hunt, R.J., Doherty, J., 2011. Interesting or important? Resetting the balance of theory and application. Groundwater 49(3), doi: 10.1111/j.1745-6584.2011.00807.x.

Ireson, A.M., Butler, A.P., 2013. A critical assessment of simple recharge models: application to the UK Chalk. Hydrol. Earth Syst. Sci. 17: 2083-2096, doi: 10.5194/hess-17-2083-2013.

Jackson, T.J., 2002. Remote sensing of soil moisture: implications for groundwater recharge. Hydrogeol. J.: 10, 40-51, doi: 10.1007/s10040-001-0168-2.

Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model? Water Resour. Res., 29(8), 2637-2649, doi: 10.1029/93WR00877.

James, S.C., Doherty, J.E., Eddebbarh, A-A., 2009. Practical postcalibration uncertainty analysis: Yucca Mountain, Nevada. Ground Water 47(6): 851-869.

James-Smith, J.M., Brown, K.G., 2002. A groundwater flow model of the Quaternary limestone aquifer in the Uley South lens, Eyre Peninsula, South Australia. Department of Water Resources, Report DWR 2002/001, Government of South Australia, Adelaide, 37 pp.

Jookadi, G., Wahr, J., Swenson, S., 2014. Estimating the human contribution to groundwater depletion in the Middle East, from GRACE data, land surface models and well observations. Water Resour. Res. 50: 2679–2692.

Jyrkama, M.I., Sykes, J.F., Normani, S.D., 2002. Recharge estimation for transient ground water modeling. Ground Water 40: 638-648, doi: 10.1111/j.1745-6584.2002.tb02550.x.

Kannan, N., Santhi, C., Arnold, J.G., 2008. Development of an automated procedure for estimation of the spatial variation of runoff in large river basins. J. Hydrol 359: 1-15, doi: 10.1016/j.jhydrol.2008.06.001.

Keese, K.E., Scanlon, B.R., Reedy, R.C., 2005. Assessing controls on diffuse groundwater recharge using unsaturated flow modeling. Water Resour. Res. 41: W06010, doi: 10.1029/2004WR003841.

Kerrou, J., Renard, P., Cornaton, F., Perrochet, P., 2013. Stochastic forecasts of seawater intrusion towards sustainable groundwater management: application to the Korba aquifer (Tunisia). Hydrogeol. J. 21: 425-440.

Kiraly, L., 1998. Modelling karst aquifers by the combined discrete channel and continuum approach. Bull. Hydrogeol., 16: 77-98.

Knowling, M.J., Werner, A.D., Herckenrath, D., 2015. Quantifying climate and pumping contributions to aquifer depletion using a highly parameterised groundwater model: Uley South Basin (South Australia). J. Hydrol. 523: 515-530, doi: 10.1016/j.jhydrol.2015.01.081.

Knowling, M.J., Werner, A.D., 2016. Estimability of recharge through groundwater model calibration: Insights from a field-scale steady-state example. J. Hydrol., doi:10.1016/j.jhydrol.2016.07.003.

Kroes, J.G., Wesseling, J.G., van Dam, J.C., 2000. Integrated modelling of the soil-water-atmosphere-plant system using the model SWAP 2.0: an overview of theory and an application. Hydrol. Process. 14, 1993-2002.

Langevin, C.D., Zygnerski, M., 2013. Effect of sea-level rise on salt water intrusion near a coastal well field in southeastern Florida. Groundwater 51(5): doi: 10.1111/j.1745-6584.2012.01008.x.

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. v2, p164-168.

Li, L., Zhou, H., Gómez-Hernádez, J.J., 2011. A comparative study of three-dimensional hydraulic conductivity upscaling at the macro-dispersion experiment (MADE) site, Columbus Air Force Base, Mississippi (USA). J. Hydrol. 404: 278-293, doi: doi:10.1016/j.jhydrol.2011.05.001.

Liu, J., Zheng, C., Zheng, L., Lei, Y., 2008. Ground water sustainability: methodology and application to the North China Plain, Ground Water, 46, 897-909, doi: 10.1111.j.1745-6584.2008.00486.x.

Lorenzo-Lacruz, J., Vicente-Serrano, S.M., López-Moreno, J.I., Beguería, S., García-Ruiz, J.M., Cuadrat, J.M., 2010. The impact of droughts and water management on various hydrological systems in the headwaters of the Tagus River (central Spain). J. Hydrol 386: 13-26.

Lubczynski, M.W., Gurwin, J. 2005. Integration of various data sources for transient groundwater modeling with spatio-temporally variable fluxes-Sardon study case, Spain. J. Hydrol. 306: 71-96, doi: 10.1016/j.jhydrol.2004.08.038.

Ma, Z., Kang, S., Zhang, L., Tong, L., Su, X., 2008. Analysis of impacts of climate variability and human activity on streamflow for a river basin in arid region of northwest China. J. Hydrol 352: 239-249.

Mack, T.J., Chornack, M.P., Taherm M.R., 2013. Groundwater-level trends and implications for sustainable water use in the Kabul Basin, Afghanistan. Environ. Syst. Decis. 33: 457-467, doi: 10.1007/s10669-013-9455-4.

Maneta, M.P., Wallender, W.W., 2013. Pilot-point based multi-objective calibration in a surface-subsurface distributed hydrological model. Hydrolog. Sci. J, 58(2): 390-407.

Maréchal, J. C., Dewandel, B., Ahmed, S., Galeazzi, L., Zaidi, F.K., 2006. Combined estimation of specific yield and natural recharge in a semi-arid groundwater basin with irrigated agriculture. J. Hydrol. 329:, 281-293, doi: 10.1016/j.jhydrol.2006.02.022.

McKenna, S.A., Doherty, J., Hart, D.B., 2003. Non-uniqueness of inverse transmissivity field calibration and predictive transport modeling. J. Hydrol. 281: 265-280, doi: 10.1016/S0022-1694(03)00194-X.

Menke, W., 1989. Geophysical data analysis: Discrete inverse theory. Academic, New York, 289 pp.

Moore, C., Doherty, J., 2005. The role of the calibration process in reducing model predictive error. Water Resour. Res. 41(5), W05020.

Moore, C., Doherty, J., 2006. The cost of uniqueness in groundwater model calibration. Adv. Water Resour. 29(4), 605-623, doi: 10.1016/j.advwatres.2005.07.003.

Moore, C., 2006. The use of regularized inversion in groundwater model calibration and prediction uncertainty analysis. PhD thesis, Univ. of Queensland, Brisbane, Australia.

Morgan, L.K., Werner, A.D., Simmons, C.T., 2012. On the interpretation of coastal aquifer water level trends and water balances: A precautionary note. J. Hydrol. 470-471: 280-288, doi: 10.1016/j.jhydrol.2012.09.001.

Morton, W., Steel, T.M., 1968. Eyre Peninsula Groundwater Study Uley South Basin, Progress Report No. 1 – Aquifer Evaluation, Department of Mines, Report 66/45, Government of South Australia, Adelaide, 19 pp.

Naff, R.L., Banta, E.R., 2008. The U.S. Geological Survey modular ground-water model–PCGN: A preconditioned conjugate gradient solver with improved nonlinear control. U.S. Geological Survey Open-File Report, 2008-1331, 44 pp.

Natural Resources Committee (NRC), 2013. Eyre Peninsula water supply: Final report "Under the lens", Parliament of South Australia, Adelaide, 233 pp.

Ordens, C.M., Werner, A.D., Post, V.E.A., Hutson, J.L., Simmons, C.T., Irvine, B.M., 2012. Groundwater recharge to a sedimentary aquifer in the topographically closed Uley South Basin, South Australia. Hydrogeol. J. 20: 61-72, doi: 10.1007/s10040-011-0794-2.

Ordens, C.M., Post, V.E.A., Werner, A.D., Hutson, J.L., 2014. Influence of model conceptualisation on one-dimensional recharge quantification: Uley South, South Australia. Hydrogeol. J. 22: 795-805.

Ordens, C.M., 2014. Recharge to a semi-arid, heterogeneous coastal aquifer: Uley South Basin, South Australia. PhD thesis, Flinders University.

Painter, J.A.C., 1969. Eyre Peninsula groundwater study Uley South basin, progress report no. 2 – aquifer evaluation. Department of Mines, Report 69/77, Government of South Australia, Adelaide.

Palma, H.C., Bentley, L.R., 2007. A regional-scale groundwater flow model for the Leon-Chinandega aquifer, Nicaragua. Hydrogeol. J. 15: 1457-1472, doi: 10.1007/s10040-007-0197-6.

Panagopoulos, G., 2012. Application of MODFLOW for simulating groundwater flow in the Trifilia karst aquifer, Greece. Environ. Earth Sci. 67: 1877-1889.

Panda, D.K., Mishra, A., Jena, S.K., James, B.K., Kumar, A., 2007. The influence of drought and anthropogenic effects on groundwater levels in Orissa, India. J. Hydrol. 343: 140-153.

Paradis, D., Gloaguen, E., Lefebvre, R., Giroux, B., 2015. Resolution analysis of tomographic slug test head data: Two-dimensional radial case. Water Resour. Res. 51: 2356-2376, doi: 10.1002/2013WR014785.

Patil, S.D., Wigington Jr, P.J., Leibowitz, S.G., Sproles, E.A., Comeleo, R.L., 2014. How does spatial variability of climate affect catchment streamflow predictions? J. Hydrol. 517: 135-145.

Peeters, L.J.M., Rassam, D., Lerat, J., 2011. Improving parameter estimation in transient groundwater models through temporal differencing. Perth, WA. In: Chan, F., Marinova, D., Anderssen, R.S. (Eds.), MODSIM2011, 19th International Congress of Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand; Perth, WA; Australia; 12–16 December, 2011, pp. 3959-3965.

Poeter, E.P., Hill, M.C., 1998. Documentation of UCODE, a computer code for universal inverse modeling. U.S. Geological Survey Scientific Investigations Report, 98-4080.

Pool, M., Carrera, J., Alcolea, A., Bocanegra, E.M., 2015. A comparison of deterministic and stochastic approaches for regional scale inverse modelling on the Mar del Plata aquifer. J. Hydrol. 531: 214-229, doi: 10.1016/j.jhydrol.2015.09.064.

Portniaguine, O., Solomon, D.K., 1998. Parameter estimation using groundwater age and head data, Cape Cod, Massachusetts. Water Resour. Res. 34(4): 637-645, doi: 10.1029/97WR03361.

Post, V.E.A., Vandenbohede, A., Werner, A.D., Maimun, Teubner, M.D., 2013. Groundwater ages in coastal aquifers. Adv. Water Resour. 57: 1-11.

Refsgaard, J.C., 1997. Parameterisation, calibration and validation of distributed hydrological models. J. Hydrol. 198: 69-97.

Refsgaard, J.C., van der Sluijs, J.P., Brown, J., van der Keur, P., 2006. A framework for dealing with uncertainty due to model structure error. Adv. Water Resour. 29: 1586-1597, doi: 10.1016/j.advwatres.2005.11.013.

Reynolds, D.A., Marimuthu, S., 2007. Deuterium composition and flow path analysis as additional calibration targets to calibrate groundwater flow simulation in a coastal wetlands system. Hydrogeol. J. 15: 515-535.

Ruud, N., Harter, T., Naugle, A., 2004. Estimation of groundwater pumping as closure to the water balance of a semi-arid, irrigated agricultural basin. J. Hydrol. 297, 51-73.

Sánchez-Vila, X., Guadagini, A., Carrera, J., 2006. Representative hydraulic conductivities in saturated groundwater flow. Rev. Geophys. 44: RG3002, doi: 10.1029/2005RG000169.

Sanford, W., 2002. Recharge and groundwater models: an overview. Hydrogeol. J. 10: 110-120, doi: 10.1007/s10040-001-0173-5.

Sanford, W.E., Plummer, L.N., McAda, D.P., Bexfield, L.M., Anderholm, S.K., 2004. Hydrochemical tracers in the middle Rio Grande Basin, USA: 2. Calibration of a groundwater-flow model Hydrogeol. J. 12(4): 389-407, doi: 10.1007/s10040-004-0326-4.

Scanlon, B.R., Cook, P.G., 2002. Preface theme issue on groundwater recharge, Hydrogeol. J. 10 (1): 3-4, doi: 10.1007/s10040-001-0175-3.

Scanlon, B.R., Healy, R.W., Cook, P.G., 2002. Choosing appropriate techniques for quantifying groundwater recharge. Hydrogeol. J. 10: 18-39, doi: 10.1007/s10040-0010176-2.

Schreüder, W., 2009. Running BeoPEST: In: Preceedings of the 1st PEST Conference, Potomac, MD, November 1-3.

Scibek, J., Allen, D.M., 2006. Modeled impacts of predicted climate change on recharge and groundwater levels. Water Resour. Res. 42: W11405, doi: 10.1029/2005WR004742.

Seibert, J., 2005. HBV Light – user's manual. Department of Physical Geography and Quaternary Geology, Stockholm University, Sweden. Available at: http://people.su.se/jseib/HBV/HBV_manual_2005.pdf.

Shamsudduha, M., Chandler, R.E., Taylor, R.G., Ahmed, K.M., 2009. Recent trends in groundwater levels in a highly seasonal hydrological system: the Ganges-Brahmaputra-Megha Delta. Hydrol. Earth Systm. Sci. 13, 2373–2385.

Shepherd, R.G., 1980. Uley South groundwater basin computer model. Department of Mines and Energy, Report 80/25, Government of South Australia, Adelaide, 24 pp.

Sibenalar, X.P., 1976. Eyre Peninsula groundwater survey – Uley South basin. Department of Mines, Report 76/30, Government of South Australia, Adelaide, 89 pp.

Skøien, J.O., Blöschl, G., Western, A.W., 2003. Characteristic space scales and timescales in hydrology. Water Resour. Res. 39(10), 1304, doi: 10.1029/2002WR001736.

Smerdon, B.D., Allen, D.M., Grasby, S.E., Berg, M.A., 2009. An approach for predicting groundwater recharge in mountainous watersheds. J. Hydrol. 365: 156-172, doi: 10.1016/j.jhydrol.2008.11.023.

Stoertz, M.W., Bradbury, K.R., 1989. Mapping recharge areas using a ground-water flow model – a case study. Ground Water 27 (2): 220-228, doi: 10.1111/j.1745-6584.1989.tb00443.x.

Taylor, N., 2003. Eyre Peninsula water supply master plan. Unpublished report prepared by parson Brinckerhoff, Pty Ltd for SA Water, Adelaide, 278 pp.

Tikhonov, A.N., Arsenin, V.Y., 1977. Solution of ill-posed problems. Winston, Washington D.C., 258 pp.

Tonkin, M.J., Doherty, J., 2005. A hybrid regularized inversion methodology for highly parameterized environmental models. Water Resour. Res. 41, W10412, doi: 10.1029/2005WR003995.

Tonkin, M.J., Doherty, J., Moore, C., 2007, Efficient nonlinear predictive error variance for highly parameterized models. Water Resour. Res. 43, W07429, doi: 10.1029/2006WR005348.

Tonkin, M., Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. Water Resour. Res. 45, W00B10, doi: 10.1029/2007WR006678.

Townley, L.R., 1995. The response of aquifers to periodic forcing. Adv. Water Resour. 18: 125-146, doi: 10.1016/0309-1708(95)00008-7.

van Loon, A.F., van Lanen, H.A.J., 2013. Making the distinction between water scarcity and drought using an observation-modeling framework. Water Resour. Res. 49: 1483-1502.

Vasco, D.W., Datta-Gupta, A., Long, J.C.S., 1997. Resolution and uncertainty in hydrologic characterization. Water Resour. Res. 33(3): 379-397, doi: 10.1029/96WR03301.

von Asmuth, J.R., Maas, K., Bakker, M., Petersen, J., 2008. Modeling time series of ground water head fluctuations subjected to multiple stresses. Ground Water 46: 30-40.

Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W., Sorooshian, S., 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water Resour. Res. 39(5), 1214, doi:10.1029/2002WR001746.

Wada Y., van Beek, L.P.H., van Kempen, C.M., Reckman, J.W.T.M., Vasak, S., Bierkens, M.F.P., 2010. Global depletion of groundwater resources. Geophys. Res. Lett. 37: L20402, doi:10.1029/2010GL044571.

Wallis, I., Moore, C., Post, V., Wolf, L., Martens, E., Prommer, H., 2014. Using predictive uncertainty analysis to optimise tracer test design and data acquisition. J. Hydrol. 515: 191-204, doi: 10.1016/j.jhydrol.2014.04.061.

Wanders, N., Wada, Y., Human and climate impacts on the 21st century hydrological drought. J. Hydrol. (2014), http://dx.doi.org/10.1016/j.jhydrol.2014.10.1047.

Wang, X., 2014. Advances in separating effects of climate variability and human activity on stream discharge: An overview. Adv. Water Resour. 71: 209-218.

Wang, D., Hejazi, M., 2011. Quantifying the relative contribution of the climate and direct human impacts on mean annual streamflow in the contiguous United States. Water Resour. Res. 47, W00J12, doi: 10.1029/2010WR010283.

Wang, B., Jin, M., Nimmo, J.R., Yang, L., Wang, W., 2008. Estimating groundwater recharge in Hebei Plain, China under varying land use practices using tritium and bromide tracers. J. Hydrol. 356(1-2): 209-222, doi: 10.1016/j.jhydrol.2008.04.011.

Werner, A.D., Alcoe, D.W., Ordens, C.M., Hutson, J.L., Ward, J.D., Simmons, C.T., 2011. Current practice and future challenges in coastal aquifer management: flux-

based and trigger-level approaches with application to an Australian case study. Water Resour. Manage. 25: 1831-1853, doi: 10.1007/s11269-011-9777-2.

Werner, A.D., Ward, J.D., Morgan, L.K., Simmons, C.T., Robinson, N.I., Teubner, M.D., 2012. Vulnerability indicators of sea water intrusion. Ground Water 50(1): 48-58.

Werner, A.D., Zhang, Q., Xue, L., Smerdon, B.D., Li, X., Zhu, X., Yu, L., Li, L., 2013. An initial inventory and indexation of groundwater mega-depletion cases. Water Resour. Manage. 27: 507-533.

Werner, A.D., 2014. Karst aquifer recharge: Comments on Somaratne, N. Characteristics of point recharge in karst aquifers. Water 2014, 6, 2782-2807. Water 6: 3727-3738, doi: 10.3390/w6123727.

White, J.T., Doherty, J.E., Hughes, J.D. 2014. Quantifying the predictive consequences of model error with linear subspace analysis. Water Resour. Res. 50(2): 1152-1173, doi: 10.1002/2013WR014767.

Wood, W.W., Sanford, W.E., 1995. Chemical and isotopic methods for quantifying ground-water recharge in a regional, semiarid environment. Ground Water 33: 458-468, doi: 10.1111/j.1745-6584.1995.tb00302.x.

Worthington, S.R.H., 1999. A comprehensive strategy for understanding flow in carbonate aquifer. In: Palmer, A.N., Palmer, M.V, Sasovsky, I.D. (Eds.), Karst Modeling. Karst Waters Institute, Charles Town, West Virginia, USA, pp. 17-29.

Yan, B., Fang, N.F., Zhang, P.C., Shi, Z.H., 2013. Impacts of land use change on watershed streamflow and sediment yield: An assessment using hydrologic modelling and partial least squares regression. J. Hydrol. 484: 26-37.

Ye, B., Yang, D., Kane, D.L., 2003. Changes in Lena River streamflow hydrology: human impacts versus natural variations. Water Resour. Res. 39(7), doi: 10.1029/2003WR001991.

Yidana, S.M., 2011. Groundwater flow modeling and particle tracking for chemical transport in the southern Voltaian aquifers. Environ. Earth Sci. 63: 709-721, doi: 10.1007/s12665-010-0740-y.

Yihdego, Y., Becht, R., 2013. Simulation of lake-aquifer interaction at Lake Naivasha, Kenya using a three-dimensional flow model with the high conductivity technique and a DEM with bathymetry. J. Hydrol. 503: 111-122.

Zhang, D., Liu, X., Liu, C., Bai, P., 2013. Responses of runoff to climatic variation and human activities in the Fenhe River, China. Stoch. Environ. Res. Risk Assess. 27: 1293-1301.

Zhang, Y., 2014. Nonlinear inversion of an unconfined aquifer: Simultaneous estimation of heterogeneous hydraulic conductivities, recharge rates and boundary conditions. Transp. Porous Med. 102: 275-299, doi: 10.1007/s11242-014-0275-x.

Zhou, P., Li, G., Lu, Y., Li, M., 2014. Numerical modeling of the effects of beach slope on water-table fluctuation in the unconfined aquifer of Donghai Island, China. Hydrogeol. J. 22: 383-396, doi: 10.1007/s10040-013-1045-5.

Zulfic, D., Harrington, N., Evans, S., 2007. Uley Basin groundwater modelling project, Vol 2: Groundwater flow model, Report DWLBC 2007/04, Department of Water, Land and Biodiversity Conservation, Government of South Australia, Adelaide, 128 pp.