USING COUPLED MODELING APPROACHES TO QUANTIFY HYDROLOGIC PREDICTION UNCERTAINTY AND TO DESIGN EFFECTIVE MONITORING NETWORKS

by

Joan Brandon Blainey

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF HYDROLOGY AND WATER RESOURCES
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
WITH A MAJOR IN HYDROLOGY
In the Graduate College
THE UNIVERSITY OF ARIZONA

2008
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Joan Brandon Blainey entitled USING COUPLED MODELING APPROACHES TO QUANTIFY HYDROLOGIC PREDICTION UNCERTAINTY AND TO DESIGN EFFECTIVE MONITORING NETWORKS and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date: 02/22/08

Paul A. Ferré

Date: 02/22/08

Jon D. Pelletier

Date: 02/22/08

Victor R. Baker

Date: 02/22/08

Hoshin V. Gupta

Date: 02/22/08

Mary C. Hill

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College. I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: 02/22/08

Dissertation Co-Director: Paul A. Ferré

Date: 02/22/08

Dissertation Co-Director: Jon D. Pelletier
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: ____Joan Brandon Blainey____
ACKNOWLEDGEMENTS

The work presented here could not have been completed without the insight of many individuals. I would like to thank Ty Ferré for his advising. In particular, I am grateful that Ty advises each student as an individual with unique aspirations and contributions. I would also like to thank my U.S. Geological Survey colleagues for their generosity with their time, especially Kyle Blasch, James Callegary, and Mary Hill. Feedback on research ideas from Survey colleagues has greatly enriched this dissertation.

I would also like to thank the following organizations and programs for their financial contributions: Achievement Rewards for College Scientists; the Geological Society of America; the John and Margaret Harshbarger Doctoral Fellow in Hydrogeology; the National Aeronautics and Space Administration Space Grant program; the National Water Research Institute, the University of Arizona, Technology and Research Initiative Fund, Water Sustainability Competitive Grants Program; and the U.S. Geological Survey Arizona Water Science Center. Funding from these organizations was essential for completion of this dissertation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>8</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>13</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>14</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>16</td>
</tr>
<tr>
<td>Research Motivation</td>
<td>22</td>
</tr>
<tr>
<td>Temporal Gravity Change Measurements</td>
<td>23</td>
</tr>
<tr>
<td>Coupled Hydrogeophysical Approach</td>
<td>24</td>
</tr>
<tr>
<td>Dissertation Structure</td>
<td>25</td>
</tr>
<tr>
<td>PRESENT STUDY</td>
<td>27</td>
</tr>
<tr>
<td>Statement of Candidate’s Contribution to Manuscripts</td>
<td>27</td>
</tr>
<tr>
<td>Summary of First Manuscript</td>
<td>28</td>
</tr>
<tr>
<td>Summary of Second Manuscript</td>
<td>30</td>
</tr>
<tr>
<td>Summary of Third Manuscript</td>
<td>32</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>36</td>
</tr>
<tr>
<td>APPENDIX A. ASSESSING THE LIKELY VALUE OF GRAVITY AND DRAWDOWN MEASUREMENTS TO CONSTRAIN ESTIMATES OF HYDRAULIC CONDUCTIVITY AND SPECIFIC YIELD DURING UNCONFINED AQUIFER TESTING</td>
<td>40</td>
</tr>
<tr>
<td>Abstract</td>
<td>40</td>
</tr>
<tr>
<td>Introduction</td>
<td>40</td>
</tr>
<tr>
<td>Numerical Modeling</td>
<td>44</td>
</tr>
<tr>
<td>Parameter Estimation</td>
<td>50</td>
</tr>
<tr>
<td>Measurement Uncertainty</td>
<td>51</td>
</tr>
<tr>
<td>Objective Function</td>
<td>54</td>
</tr>
<tr>
<td>Optimization Algorithm</td>
<td>57</td>
</tr>
<tr>
<td>Results of Parameter Estimation</td>
<td>58</td>
</tr>
<tr>
<td>Concluding Remarks</td>
<td>61</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>63</td>
</tr>
<tr>
<td>References</td>
<td>63</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS – Continued

Tables ........................................................................................................................................ 65
Figures ....................................................................................................................................... 67

APPENDIX B. INFILTRATION ON ALLUVIAL FANS IN ARID ENVIRONMENTS:

INFLUENCE OF FAN MORPHOLOGY ................................................................................ 71

Abstract ................................................................................................................................... 71
Introduction ............................................................................................................................... 72
Conceptual Model of Alluvial Fans ......................................................................................... 74
Numerical Model .................................................................................................................... 80
  Alluvial Fan Geometry ......................................................................................................... 81
  Flow Routing ....................................................................................................................... 84
  Infiltration .......................................................................................................................... 87
Application ............................................................................................................................... 89
  Tiger Wash Case-Study ......................................................................................................... 90
  Numerical Experiments ....................................................................................................... 93
Conclusions .............................................................................................................................. 101
Acknowledgments .................................................................................................................. 105
Notation ................................................................................................................................... 106
References ............................................................................................................................... 107
Tables ....................................................................................................................................... 115
Figures ..................................................................................................................................... 117

APPENDIX C. DESIGNING EFFICIENT HYDROLOGIC MONITORING

NETWORKS USING COST-BENEFIT ANALYSIS ................................................................ 130

Abstract .................................................................................................................................. 130
Introduction ............................................................................................................................. 131
Proposed Network Design Approach .................................................................................... 135
  Consideration of Model Errors ............................................................................................ 136
  Application of a Cost Model ................................................................................................. 136
  Consideration of Measurement Errors ................................................................................. 138
  Consideration of Competing Measurement Sets ............................................................... 139
  Calculating the Performance and Measurement Costs of Each Measurement Set ........... 140
  Static versus Recursive Network Design ............................................................................ 141
Synthetic Case Study ............................................................................................................... 142
  Cost Function and Measurement Costs .............................................................................. 144
  Numerical Water Flow Modeling ......................................................................................... 145
  Numerical Gravity Response Modeling .............................................................................. 146
### TABLE OF CONTENTS – Continued

- **Gravity Measurement Errors** ................................................................. 147
- **Parameter Estimation** ............................................................................. 150
- **Results** ........................................................................................................ 153
  - Initial Uncertainty ....................................................................................... 153
  - Measurement Sensitivity and Parameter Identifiability ......................... 154
  - Average Expected Predictions of Sets with Different Numbers of Measurements .................................................................................................................. 156
  - Expected Predictions of Specific Measurement Sets .............................. 159
  - Expected Predictions and Uncertainty of Specific Measurement Sets ........ 160
  - Expected Performance Cost of Specific Measurement Sets .................... 162
  - Combining the Measurement and Performance Costs to Define the Total Cost .......................................................................................................................... 164
- **Choosing a Monitoring Network** ............................................................... 167
- **Conclusions** ................................................................................................. 169
- **Acknowledgements** .................................................................................... 171
- **References** .................................................................................................... 171
- **Tables** ............................................................................................................ 180
- **Figures** ......................................................................................................... 181
LIST OF ILLUSTRATIONS

APPENDIX A

Figure 1. Error free (lines) and simulated measurements (symbols) of (a) drawdown and (b) gravitational response after 7 days of pumping in an unconfined aquifer with \( b = 50 \text{ m}, K = 10^{-4} \text{ m/s}, Q = 6.309 \times 10^{-2} \text{ m}^3/\text{s}, S = 0.01, S_y = 0.25, \) and \( z_s = 25 \text{ m}. \) The measurement errors are simulated with 1 standard deviation equal to 0.24 m and 0.048 m, and 1 \( \mu \text{Gal} \) for drawdown and gravity measurements, respectively. (c) Drawdown and (d) gravitational responses for \( K = 10^{-3} \text{ m/s} \) and \( S_y = 0.35 \) (dotted curve); \( K = 3.2 \times 10^{-5} \text{ m/s} \) and \( S_y = 0.35 \) (dotted-dashed curve); \( K = 10^{-3} \text{ m/s} \) and \( S_y = 0.15 \) (solid curve); \( K = 3.2 \times 10^{-5} \text{ m/s} \) and \( S_y = 0.15 \) (dashed curve). ............................................. 67

Figure 2. Simulated drawdown in m (top row) and gravitational response in \(-1*\mu \text{Gal}\) (bottom row) after 7 days of pumping in an unconfined aquifer with \( b = 50 \text{ m}, K = 10^{-4} \text{ m/s}, Q = 6.309 \times 10^{-2} \text{ m}^3/\text{s}, S = 0.01, \) and \( z_s = 25 \text{ m} \) for varying values of hydraulic conductivity and specific yield. The left, middle, and right columns show results for observation points located 3 m, 15 m, and 100 m from the pumping well, respectively. ................................................................................................................ 68

Figure 3. The drawdown component (top) and the gravity component (middle) of the objective function (sum of square errors normalized by the average measurement signal), and the combined objective function (bottom) without simulated measurement error for data collected after 7 days of pumping in an unconfined aquifer for the following conditions: \( b = 50 \text{ m}, K = 10^{-4} \text{ m/s}, Q = 6.309 \times 10^{-2} \text{ m}^3/\text{s}, S = 0.01, S_y = 0.25, \) and \( z_s = 25 \text{ m}. \) .......................................................................................................................... 69

Figure 4. Parameter estimates for simultaneous estimation of hydraulic conductivity and specific yield after 7 days of pumping in an unconfined aquifer for the following conditions: \( b = 50 \text{ m}, K = 10^{-4} \text{ m/s}, Q = 6.309 \times 10^{-2} \text{ m}^3/\text{s}, S = 0.01, S_y = 0.25, \) and \( z_s = 25 \text{ m}. \) Parameter estimates are fit to simulated measurements of drawdown, gravity, or both drawdown and gravity at 9 radial distances from the pumping well. (a) Bias (%) and (b) Two standard deviations. For the higher error level, errors are simulated with 1 standard deviation equal to 0.24 m and 1 \( \mu \text{Gal} \) for drawdown and gravity measurements, respectively. For the lower error level, errors are simulated with 1 standard deviation equal to 0.048 m and 1 \( \mu \text{Gal} \) for drawdown and gravity measurements, respectively. Each plotted point is based on 100 random measurement realizations................................................................. 70
APPENDIX B

Figure 1. A. Image of Little Rainbow Valley which is located to the southwest of Phoenix, Arizona. Waterman Wash runs through the center of Little Rainbow Valley (3x vertical exaggeration). Image from NASA World Wind, LandSat7. B. Image of the Maricopa Mountains from Google Earth (2.75x vertical exaggeration). C. Image of the Sierra Estella Mountains from Google Earth (2.75x vertical exaggeration).

Figure 2. Drainage area versus flow depth for 7 gaging stations on piedmonts in Maricopa County (FCDMC, 2006) and paleoflood reconstructions for 5 canyons supplying flows to the Tortolita fan complex. For each gaging station, the three largest peak discharge events during the period of record plus any available historical flood events were plotted. For the paleoflood data, the error bars on the flow depth represents a range in the discharge estimates.

Figure 3. Schematic of alluvial fan model with fan geometry input variables identified ($I$ is the incision depth at the fan apex, $L$ is the relief of the fan, $R$ is the radius of the fan, $x_{ij}$ is the horizontal distance to grid cell $ij$, and $w_0$ is the channel half-width at the fan apex).

Figure 4. Fan geometry for the case-study of Tiger Wash, Maricopa County, Arizona. A. Longitudinal profile down the fan (vertical exaggeration 8x). B. Radial cross-section mid-distance down the fan (vertical exaggeration 1600x).

Figure 5. Plan view of the widening of the active channel with distance along the longitudinal axis as specified by the variable $c$ for a feeder channel half width that is 2.3% of the radial distance of the total fan radius. $c$ values of 7, 20, and 40 correspond to configurations where the active channel area is 15%, 34%, and 49% of the total fan area ($9 \text{ km}^2$).

Figure 6. Examples of $p$ values used to define topographic surfaces in plan view. Black areas correspond to the oldest unincised surface, white areas correspond to topographic islands in the active channel, and gray areas correspond to the lowest topographic surface in the active channel for a single random realization of topography. A. $p = 0.15$ B. $p = 0.5$. C. Surface flow depth for a flow of a 1 hour duration and an input flow depth of 1.5 m ($p = 0.3$). D. Inset of flow depth near the fan apex.
LIST OF ILLUSTRATIONS – Continued

Figure 7. The infiltration rate used in the alluvial fan model that corresponds to the median cumulative infiltration from the explicit Green-Ampt solution for a flow event 8 hours in duration for a sandy soil with a saturated hydraulic conductivity, $K_s$, of 14.6 cm/hr and 100 cm of ponded water. A. Infiltration rate. B. Cumulative infiltration………………………………………………………………………………………………………122

Figure 8. Schematic of a grid cell in the model domain. $d_i$ is the depth of surface water entering the cell, $d_o$ is the depth of surface water exiting the cell, and $(\Delta x)f/v$ is the infiltration loss from the cell…………………………………………………………………………………………123

Figure 9. Streamflow (station ID 5163) (FCDMC, 2006) and precipitation (station ID 5160) (FCDMC, 2006) for Tiger Wash for A. Monsoon storm on 8/25/2003. B. Winter storm on 12/28/2004…………………………………………………………………………………………124

Figure 10. A. Surficial geologic map of Tiger Wash (modified from Klawon and Peartree, 2000) (Qy – Late Holocene active channel, sheetflood, and overbank deposits; Qy1 – Early to late Holocene relict alluvial fan and terrace deposits; Ql – late Pleistocene alluvium; Qm - Middle Pleistocene alluvial fan deposits). The black line denotes the active channel area used in the numerical model. The hatched area denotes inactive channel sediments within the active channel area used in the numerical model. B. Lateral expansion of the active channel area with distance from the fan apex along transect a-a'. The width of the active channel is plotted as a half-width for direction comparison with the parameter $w_0$ used in the numerical model………………………………………………………………………………………………………125

Figure 11. Surface flow depth (m) (top row) and infiltrated water (m) (bottom row) accumulated over the duration of the storm for a 1 hour monsoon flow event (left column) and a 10 hour winter flow event (right column)………………………………………………………………………………………………………126

Figure 12. A. active channel area inundated with surface flow (%) and the infiltration rate into sandy active channel sediments for a ponded water depth of 0.5 m. B. fan-wide infiltration volume ($m^3$) and fan-wide volumetric infiltration rate ($m^3/s$) as a function of event duration………………………………………………………………………………………………………………127

Figure 13. Channel slope versus fan area with an input flow depth of 1.0 m and an incision depth of 2.0 m. A. Fan-wide volumetric infiltration rate divided by the input discharge. B. Longitudinal distance of the center of mass of the surface flow as a percent of the distance downfan, $x_m$………………………………………………………………………………………………………128
LIST OF ILLUSTRATIONS – Continued

Figure 14. Ratio of input flow depth to incision depth versus active channel area (%) with an input flow depth of 1.25 m for A. Fan-wide volumetric infiltration rate \(\text{m}^3/\text{s}\) on the fan divided by the sum of the fan-wide volumetric infiltration rate and discharge off the fan \(\text{m}^3/\text{s}\), \(\left(\Delta x^2\right)f/(Q_e + \left(\Delta x^2\right)f)\), and B. Fraction of the older surface area inundated with flow (%).……………………………………………………………………128

Figure 15. Discharge per drainage area versus the permeability contrast between active channel and unincised surface deposits for a fan where the active channel area is 27% of the total fan area for A. Ratio of infiltration on unincised surfaces to infiltration on active channels, \(I_i/I_a\), and B. Longitudinal distance of the center of mass of the infiltrated water as a percent of the distance downfan, \(x_m'\) ……………129

APPENDIX C

Figure 1. Times series of hydrologic and gravity response to infiltration and drainage. A. Cumulative infiltration. B. Depth to the wetting front (m). C. Gravity change with error bars representing a spread of two standard deviations of measurement error for \(K_s=1.0\text{e-4m/s}\) and \(\varepsilon=3.25\) (solid line), \(K_s=7.92\text{e-5 m/s}\) and \(\varepsilon=3.75\) (dashed line), and \(K_s=5.83\text{e-5 m/s}\varepsilon=4.25\) (dotted line). The vertical solid line represents \(t_d\), the time of drainage……………………………………………………………………181

Figure 2. A. Relative error of the depth to the wetting front at the end of the monitoring period. B. Time elapsed (days) for the wetting front to reach 200 m as a function of the number of measurements in the measurement set..............................................182

Figure 3. Absolute error of the time for the wetting front to reach 200 m for textures A, B, and C for 4 measurements. The symbols identify the best measurement set for each soil texture. The best measurement set for texture C is also has the minimum error across all three soil textures………………………………………………….182

Figure 4. Predictions of the time for the wetting front to reach a depth of 200 m below land surface for alluvium textures A, B, and C with overall best measurement sets: \(g_3-g_8, g_3-g_5-g_7-g_8, g_2-g_4-g_5-g_6-g_7-g_8, g_1-g_2-g_3-g_4-g_5-g_6-g_7-g_8\)…………………183
LIST OF ILLUSTRATIONS – Continued

Figure 5. Cost of the prediction of the time for the wetting front to reach a depth of 200 m below land surface for alluvium texture A as a function of the number of measurements used in parameter estimation for a linear cost function for the best measurement sets across all soil textures. For 8 measurements the measurement cost is 40% of the a priori performance cost. A. Performance cost, measurement cost, and total cost (height of the stacked bars). B. Total cost determined using the most-likely (filled triangles) and probability-weighted (open triangles) approaches…………..184

Figure 6. Total cost for linear cost function cf1 for texture A as a function of the total measurement cost for 8 measurements for prediction based on SCEM with uncertainty................................................................................................................184

Figure 7. The optimal number of measurements for the lowest total cost as a function of the total measurement cost for 8 measurements. The overall best measurement set is the same for each texture.................................................................185
LIST OF TABLES

APPENDIX A

Table 1. Parameter values and parameter ranges used for sensitivity analysis. ........ 65

Table 2. Summary of parameter estimation results for hydraulic conductivity and specific yield based on 9 measurement locations and the reference set of test conditions and measurement errors of either 0.24 m or 0.048 m, and 1 μGal for drawdown and gravity, respectively. ................................................................. 66

APPENDIX B

Table 1. Model inputs and parameters for alluvial fan geometry, flow routing, and infiltration ................................................................. 115

Table 2. Hydraulic properties (saturated hydraulic conductivity, initial water content, saturated water content, and bubbling pressure) for the explicit Green-Ampt solution were assigned on the basis of the geometric mean soil texture (Rawls et al., 1992) ................................................................. 116

APPENDIX C

Table 1. Parameter values and parameter ranges used for synthetic gravity measurements ................................................................. 180
ABSTRACT

Designing monitoring networks that can discriminate among competing conceptual models is a key challenge for hydrologists. This issue is examined by considering the impact of network design on the utility of measurements for constraining hydrologic prediction uncertainty. Specifically, a three-staged approach was developed and is presented as a set of modeling case studies. The first case study presents a sensitivity analysis that examines conditions under which the proposed measurement method is likely to detect observations associated with the hydrologic process and properties of interest. This application is focused on the use of geomorphic information to estimate infiltration on arid alluvial fans.

The second stage is an assessment of the likely utility of the measurement method to determine whether proposed measurements are likely to be useful for identifying hydraulic properties or hydrologic processes. This objective screening approach could reduce the number of unsuccessful uses of geophysical and other indirect measurement methods. A hypothetical site assessment examines whether the measurement method, temporal gravity change, is likely to detect signals associated with drawdown in an unconfined aquifer that occurs in response to pumping. Also, the utility of these measurements for identifying hydraulic conductivity and specific yield was considered.

The third stage, an analysis of optimal network design, compares the projected measurement costs with the expected benefits of constraining hydrologic prediction uncertainty. The final case study presents a network design approach for a feasibility assessment of a proposed artificial recharge site. Predefined sets of proposed measurements of temporal gravity change were considered for various measurement
times. An ensemble approach was used to assess the likely impact of measurement error on prediction error and uncertainty for different combinations of measurement sets. The ensemble of prediction errors was translated to probability-weighted performance costs for each measurement set using a cost function. Total cost was calculated as the sum of the performance and measurement costs. The optimal measurement set, defined as the set with the lowest total cost, depends on the prediction of interest, the per measurement cost, the maximum risk-based cost associated with the hydrologic prediction, and the treatment of uncertainty in defining performance costs.
INTRODUCTION

Hydrology integrates processes that occur over a wide range of spatial and temporal scales. Ultimately, hydrologists seek to describe, understand, and predict the integrated outcomes of these processes. However, we are limited both by computational resources and, moreover, by our ability to acquire adequate quantities of data and to obtain sufficiently accurate data for constraining hydrologic predictions. Computational resource limitations have a primary impact on the spatial and temporal resolution of numerical models. In particular, these limitations force us to simplify highly-detailed, small-scale processes into larger-scale processes with effective properties. Typically, data collection is limited due to the available measurement techniques and, often to a great degree, budgetary and time constraints. As a result, hydrologists have to decide which measurements to make (type, time, location) to constrain models that represent hydrologic processes at multiple scales. Taken together, decisions regarding how to link processes across scales and decisions regarding which measurements to make often control the accuracy, uncertainty, and therefore the value of hydrologic assessments. The overall objective of this dissertation is to explore how data can be utilized to address each of these challenges using coupled numerical models. The dissertation is divided into two main parts: 1) coupling overland flow and infiltration models to predict runoff and infiltration patterns on alluvial fans (Appendix B); and 2) coupling subsurface hydrologic and ground-based gravity instrument response models to infer hydraulic properties (Appendix A) and to design hydrologic monitoring networks for improved hydrologic predictions (Appendix C).
Accurate monitoring of hydrologic processes and subsurface conditions and properties is important in order to discriminate among competing conceptual models of hydrologic investigations. Monitoring is complex due to temporal and spatial variation in hydrologic events, differences in scale between hydrologic processes and measurements, and subsurface heterogeneity. These technical challenges are amplified by the low sampling density that is common in hydrologic investigations due to the limits of the measurement methods, restricted monitoring budgets, and limited project timelines. For example, consider the construction of monitoring wells, a standard method for water level measurements, which is feasible, especially in the shallow subsurface, yet relatively expensive. As point-scale measurements, water level measurements can only provide localized information at discrete locations. Point-scale measurements are more likely to be direct measurements of a hydrologic property or state; they tend to have relatively high accuracy and certainty, often are automated to provide high temporal coverage, but offer low spatial coverage.

Technical challenges to subsurface monitoring arise, in part, because of difficulties with measuring most subsurface properties directly; generally, another property is measured and used to infer the difficult to measure property of interest. Volumetric water content, for example, is inferred by directly measuring the weight and bulk density of a soil sample. Indirect (geophysical) measurements are often intermediate- to large-scale measurements; they have relatively low accuracy and certainty, variable temporal coverage, but offer high spatial coverage. An additional consideration for indirect measurements is the uncertain spatial averaging effects and,
frequently, low sensitivity to small hydrologic events (Hubbard et al., 2002; Kowalsky et al., 2006).

Instrumentation for vadose zone measurements has undergone extensive development in the last few decades. These measurement techniques vary considerably in spatial scale, sample volume, depth of investigation, spatial sensitivity, portability, accuracy, and precision. Most vadose zone instruments (e.g., tensiometers, neutron probes, time domain reflectometry probes, heat dissipation probes) measure subsurface properties on the scale of 100s to 1000s of cubic centimeters (e.g., Ferré et al., 1996). Intermediate- to large-scale measurements, such as low frequency electrical resistivity methods (e.g., electrical resistivity tomography (Daily et al., 1992; Kemna et al., 2002)), are also used for hydrologic investigations. Many other techniques at this measurement scale, such as seismic (e.g., Haines et al., 2007), magnetic anomaly (e.g., Babu et al., 1991), electromagnetic (e.g., Buselli and Lu, 2001), and gravity methods (Montgomery 1971a, 1971b; Pool and Eychaner, 1995; Howle et al., 2003) were originally developed for geologic exploration but are increasingly applied to hydrologic investigations.

Remote sensing techniques, especially airborne and space-borne methods, can monitor soil properties such as water content and provide measurements at a large scale, at resolutions of kilometers (Margulis et al., 2002; Swenson et al., 2006).

When designing, evaluating, or modifying a monitoring plan, hydrologists integrate different types of measurements that can provide distinct information about subsurface properties and processes. A main part of the dissertation, Appendix B, addresses an application in which large-scale geomorphic information and point-scale permeability information is integrated for improved hydrologic predictions of runoff and
recharge on alluvial fans. The purpose of this effort is to predict where and how recharge occurs on alluvial fans as a function of fan morphology (e.g., fan slope, fan area, active channel proportion of fan area, and entrenchment of the active channel). In this effort, a coupled numerical model of steady surface water flow and Green-Ampt-type infiltration was applied to synthetic alluvial fans with connected distributary channel networks. If the model was applied to a real-world fan, the stochastic distributary channel network would likely be replaced with actual topographic information. In this case, with geomorphic site information, the hydrologic-geomorphic portion of the model could be applied to a fan to predict where flows are likely to occur and how deep infiltration is likely to be. These model predictions could be used to guide monitoring efforts to determine the appropriate distance down fan and the depth at which to place monitoring instruments. Careful selection of infiltration monitoring sites on arid alluvial fans is essential when flows are infrequent and of short-duration and both monitoring budgets and monitoring periods are typically limited. Additionally, at the scale of a single fan, the hydrologic-geomorphic portion of the model could be used to interpolate existing hydrologic data. For example, if limited infiltration point-data were collected on a fan, this model could be used to develop a first-order approximation of fan-wide infiltration responses based on fan morphology.

Measurements are used to calibrate hydrologic models by estimating the model parameters that provide the best fit with the observed data. Oftentimes, the parameter estimation problem is ill-posed because no unique solution exists and several combinations of different parameter values may lead to a similar fit with the data. Integrating different types of measurements can have a significant impact on the ability to
assess the hydrologic system by diminishing parameter correlations relative to that with a single measurement type (McKenna and Poeter, 1995; Hill and Østerby, 2003). For example, consider groundwater model calibration in which model parameters, including subsurface hydraulic properties, are determined based only on hydraulic head information and without significant imposed or observed flows. In this case, recharge and hydraulic conductivity parameters are generally correlated (Hill and Østerby, 2003). However, the addition of defined stresses such as significant pumpage, or flow and transport observations can significantly reduce parameter correlation (Anderman et al., 1996; Hill and Østerby, 2003). However, it has proven to be difficult to integrate measurements of different types. As a result, to date, there is no accepted approach to integrating different types of measurements that considers their sample volumes and measurement uncertainties.

While many practicing hydrologists are faced with the question of how to use the available measurements at a specific study site, a larger question is how to design optimal monitoring networks that provide measurements that improve our ability to discriminate among and to constrain models for improved hydrologic predictions. Numerous previous efforts have examined the design of optimal monitoring network for single measurement types, but few of these studies have integrated measurements with a range of support volumes. One of the critical questions is whether large scale measurement methods (those for which the property of interest varies considerably on a scale that is smaller than the measurement volume) can be expected to provide data with sufficient resolution to constrain a hydrologic problem of interest. Given the many differences between point- and large-scale measurements, together with differences in the costs of different
measurement types, it can be difficult to assess the relative value of different types of measurements when planning hydrologic monitoring networks. These questions are examined for two hypothetical case studies that apply a method for assessing the likely utility of an indirect measurement method and, if warranted by this initial assessment, for designing monitoring networks with indirect measurements.

A general approach is presented as a three staged assessment for data collection with a new instrument or for applying an accepted measurement method to new hydrologic conditions. First, a sensitivity analysis should be completed for the proposed observations to ensure that the measurement method is likely to detect signals associated with the hydrologic process and properties of interest. This analysis should consider both the measurement resolution and the expected measurement errors. Second, a first-level assessment of the likely utility of the method should be conducted to determine whether these measurements are likely to be useful for identifying hydraulic properties or hydrologic processes. For this analysis, it is most informative to adopt conditions that are favorable for the successful use of the method. If these analysis suggests that the measurement method is unlikely to be successful under these favorable conditions, it is unlikely that the method will be useful under more challenging conditions. If the analysis suggests that the measurement method is likely to be successful under some conditions, a third, more complete, site-specific analysis of network design should be performed using all available information before collecting field measurements. If warranted by the study objectives and budgetary requirements, the third site-specific analysis of network design compares the projected measurement costs with the expected benefits of constraining hydrologic prediction uncertainty to determine the optimal monitoring network. The
measurement network design approach presented is most appropriate for hydrologic assessments for which there is limited initial study site information and for which it is impractical to modify the design of the measurement network during monitoring. An ensemble approach is used to assess the likely impact of measurement error on prediction error and prediction uncertainty for different combinations of candidate measurements (measurement sets). This three staged approach – conducting a sensitivity analysis of the proposed observations, assessing the value of the proposed observations for parameter estimation, and designing a more efficient monitoring network for constraining hydrologic prediction uncertainty – can provide an objective screening of proposed measurement methods for specific hydrologic applications thereby reducing the number of unsuccessful uses of geophysical and other indirect methods.

Research Motivation

Hydrologic investigations typically begin with an evaluation of existing information to address a particular prediction of interest. These predictions may range from highly specific and applied questions (e.g., quantifying water or solute mass flux at a location through time) to more general, conceptual questions (e.g., testing multiple hypotheses describing a hydrologic system). Based on this initial evaluation of the existing data, if further information is needed, additional measurements are collected. After instrument installation, data is collected over a period of time and used to address the hydrologic prediction of interest. While quantitative tools are typically brought to bear on the evaluation of data and on the creation, analysis, and calibration of numerical models, few quantitative methodologies are applied to the design of hydrologic
monitoring, which frequently occupies the majority of the project budget and timeline. Applying quantitative tools to monitoring may be especially critical for semi-arid and arid zone studies for which the hydrologic events of interest are typically infrequent and of short duration. In practice, most hydrologic monitoring network designs are based on general guidelines or expert experience. This is an unstructured approach to achieving the objective of monitoring network design – quantification of the trade-off between the expected performance and projected costs of competing observation sets for selection of an optimal monitoring network. The primary aim of the present work is development of pre-monitoring numerical assessments so that, if appropriate, the likely type, number, and location of monitoring instruments and the monitoring period can be determined and further refined during data collection. In fact, the hydrologic assessments presented in the dissertation are only preliminary efforts to determine the optimal design of monitoring networks with multiple types of measurements.

Temporal Gravity Change Measurements

Ground-based gravity is the indirect geophysical method examined in this work. Like many non-invasive methods, gravity has low installation costs and relatively high flexibility in choosing a monitoring location. However, it is unclear whether temporal gravity change measurements have sufficient information to constrain a hydrologic analysis due to low signal to noise ratios and the nonunique interpretation of the gravity signal. For example, a gravimeter will respond to changes in the subsurface density distribution due to pumping in an unconfined aquifer. However, the magnitude of the response depends on both the amount of water and its spatial distribution. Under
favorable conditions, ground-based relative gravimeters are sensitive enough to measure changes in gravitational acceleration caused by changes in the subsurface density distribution associated with hydrologic processes (Montgomery 1971a, 1971b; Pool and Eychaner, 1995; Howle et al., 2003; Damiata and Lee, 2006). However, hydrologic applications of temporal gravity surveys are often limited by a relatively low signal to noise ratio. For example, portable relative gravimeters have a measurement resolution of approximately 1 microGals = 1×10^{-8} m/s^2 (Scintrex). This is commonly considered to represent a change of an extensive layer of water, or an infinite slab, that is about 1 inch thick (Telford et al., 1990), when saturation changes occur at depth. In practice, the resolution of the instrument is not the only source of measurement error. Gravity measurements must be adjusted using standard gravity corrections (Telford et al., 1990). Many corrections – such as instrument tilt, temperature, and earth tides – are automated by the instrument. Relative, not absolute, differences in gravity are considered here. Therefore, many of the standard gravity corrections (e.g., Bouguer and elevation corrections) are not needed because the corrections are identical for measurements made before and after a hydrologic event or stress of interest.

Coupled Hydrogeophysical Approach

A key aspect of this work is the use of a hydrogeophysical model that incorporates indirect measurements into hydrologic analysis. In the hydrogeophysical approach presented, the simulated gravity instrument response depends on both the forward hydrologic model, and on model parameters, such as porosity, which affect both the hydrologic and gravity response models. Parameter estimation was conducted using a
coupled hydrogeophysical approach (Rucker and Ferré, 2004; Kowalsky et al., 2005; Ferré et al., 2006). That is, during each step of the inversion process, the hydrologic parameter values were perturbed, the hydrologic forward model was run, and the gravity responses were calculated for the predicted hydrologic response and compared directly to the synthetic gravity observations. This approach avoids many of the errors associated with independent geophysical inversion. This type of numerical investigation can determine whether the effects of spatially distributed sensitivity are likely to control the value of indirect measurements for parameter estimation. It is difficult, if not impossible, to make these types of inferences without performing a coupled hydrogeophysical analysis.

Dissertation Structure

The dissertation consists of three chapters, the first of which summarizes the background, theory, and relevancy of the work. The second chapter provides a brief motivation, overview, and the major conclusions of each manuscript prepared during the course of the dissertation effort. The dissertation is primarily composed of three original research papers that have been published, accepted for publication, or submitted to scientific journals. These manuscripts, Appendices A, B, and C, are presented in chronological order as published by or submitted to the journal with differences only in the presentation style. The title of these manuscripts, along with the journal name and status at the time of the dissertation defense are:


Appendix C: Blainey, J.B. and T.P.A. Ferré, Designing efficient hydrologic monitoring networks using cost-benefit analysis (submitted to *Water Resources Research*).
PRESENT STUDY

The methods, results, and conclusions of this study are presented in three original research papers appended to the dissertation. The following is a summary of the candidate’s contribution to the papers and of the most important findings of the research papers.

Statement of Candidate’s Contribution to Manuscripts

Most of the research of and writing associated with the dissertation was closely guided and reviewed by the candidate’s primary academic advisor. For the first manuscript, the research ideas were primarily initiated by the candidate’s primary academic advisor and about 75% of the research was guided by the candidate. All of the modeling and data analysis were conducted by the candidate. Writing of the first manuscript was the work of the candidate while major reviews and suggestions for changes were contributed by the primary academic advisor. For the second manuscript about 80% of the project idea and design was initiated by the candidate. The modeling work by the candidate involved modifying an existing flow routing code and adding an infiltration component to existing code from Professor Pelletier. The interpretation and writing of the second manuscript was almost exclusively, about 95%, the work of the candidate, with suggestions and corrections supplied by Professor Pelletier. The initial idea of the third manuscript was primarily the idea of the candidate. As the work evolved, the primary academic advisor was the instigator for ideas regarding the use of different cost models. The modeling work, interpretation, and analysis were exclusively the work
of the candidate. The introduction and conclusion sections of the third manuscript were
written in partnership with the primary advisor, but of remainder of the manuscript, about
70% was written by the candidate.

Summary of First Manuscript

The motivation of the first paper was to investigate whether large-scale indirect
measurements, which provide improved spatial coverage but a lower signal to noise ratio
than traditional point-scale direct measurements (Hubbard et al. 2002; Kowalsky et al.
2006), have sufficient resolution to constrain hydrologic predictions of interest.
Specifically, the objective of the first paper is to assess the likely utility of ground-based
gravimetry for constraining effective hydraulic parameters, including hydraulic
conductivity and specific yield, during unconfined aquifer testing. A synthetic case study
was employed in which the signal to noise ratio of the gravity measurements was
adequate to detect drawdown in the unconfined aquifer. For the synthetic case study, the
primary question is – does use of temporal gravity change measurements to constrain
estimates of hydraulic conductivity and specific yield lead to estimates that are
comparable to those based on hydraulic head measurements? Possible differences in the
quality of parameter estimation were hypothesized to arise because gravity measurements
are indirect measures of changes in water distribution and have a large support volume
with a non-uniform spatial sensitivity, whereas pressure head measurements are direct
and local.

A coupled hydrogeophysical approach was used to simulate hydrologic and
gravity responses to seven days of pumping at a constant discharge in a homogenous
unconfined aquifer. Synthetic measurements, drawdown and/or gravity, were produced for a radial transect of nine measurements within 300 m of the pumping well. Synthetic gravity measurements were employed, both independently and in combination with piezometric measurements, to constrain parameter estimates with inverse modeling. The signal to noise ratio of the synthetic gravity measurements was designed to represent a conservative estimate of the maximum achievable repeatability of an absolute gravimeter (FG-5 absolute gravimeter, Micro-g LaCoste, Lafayette, CO).

A major finding of the first paper is that use of only gravity measurements was not likely to lead to accurate simultaneous parameter estimates of hydraulic conductivity and specific yield. However, gravity measurements, used in combination with drawdown measurements, may significantly improve the quality of specific yield estimates. For joint use of drawdown and gravity measurements during optimization, the quality of these parameter estimates was similar for both high- and low-quality drawdown data. This suggests that the addition of gravity data may reduce the required accuracy of drawdown data for aquifer testing. Results also indicate that exact quantification of the measurement error is not essential for effective use the indirect, non-local measurement. When applying a new measurement method to a hydrologic investigation or when applying a measurement method to a new hydrologic application, this type of synthetic numerical study can determine whether the effects of spatially distributed sensitivity are likely to control the value of indirect measurements for parameter estimation. It is difficult to draw these conclusions without performing this coupled hydrogeophysical analysis. This study further refines the areas of opportunity and limitations to the application of gravity methods for aquifer testing.
Summary of Second Manuscript

Mountain-front recharge through alluvial fans is a function of morphology (topography, soil texture, vegetation) as well as channel and flow characteristics (Houston, 2002; Izbicki et al., 2002). The motivation of the second manuscript was to investigate relationships between geomorphology and infiltration on alluvial fans in arid settings with the goal of using geomorphologic information to characterize infiltration over large areas. Previous field studies in arid settings have shown that the spatial variability in infiltration is related to the geomorphic setting as distinguished by interdrainage area, topographic depression, and drainage area (Scanlon et al., 1999) or by distance from the mountain front (Izbicki et al., 2002). The objective of the second study is to develop a first-order prediction of where and how recharge occurs on fans as a function of fan morphology based on simple observables. Previous investigators have developed numerous models of alluvial fans (Price, 1974; De Chant et al., 1999; Coulthard et al., 2002; Nicholas and Quine, 2007); this model, however, focuses on the relationship between fan morphology and hydrologic responses with consideration of the dynamic nature of the distributary channel network. To better understand the geomorphic factors (e.g., fan slope, fan area, active channel proportion of fan area, sediment permeability, geometry and sediment characteristics of fan sequences, and entrenchment of the active channel) that control flow and infiltration on alluvial fans, a coupled numerical model of steady surface flow and Green-Ampt-type infiltration was developed. Using this distributed numerical model of synthetic alluvial fans, an open question in fluvial geomorphology – how fan geometry affects the partitioning of flow between on-
fan infiltration and the valley floor – was addressed. Infiltration responses were compared for fan morphologic end members: fans with narrow, deeply-entrenched channels and fans with shallowly-entrenched active channel areas that rapidly widens with distance down fan. Additionally, the impact of sediment permeability on infiltration responses was investigated by considering fans with different permeability contrasts between the unincised surfaces and the active channel deposits which can be attributed to surface age (McFadden et al., 1987; McFadden et al., 1992; Young et al., 2004) or depositional environment.

A major finding of the second paper is that the amount of infiltration on fans is most sensitive to the geomorphic factors that influence the area of inundation rather than the depth of surface flow. The greatest amount of infiltration occurred on fans with low gradients that resulted in deeper surface flows and more lateral spreading of flow. The ratio of the incision depth to the input flow depth near the fan apex was an important predictor of the amount of infiltration on the fan. This ratio controlled the partitioning of flow between the fan and the valley floor. More specifically, this ratio largely determined whether the surface flow regime consisted of deeper flow confined within the active channel portion of the fan, or whether flow also inundated surfaces outside of the active channel area with shallow sheetflow.

This paper illustrates that fan morphology exerts a first-order control on the recharge potential. Therefore, at the basin scale, this method could be used to compare the recharge potential of different arid and semi-arid watersheds using limited hydrologic information. These results can be used to guide the development of monitoring networks on fans, guide geomorphic mapping for flood hazard prediction, and characterize the
infiltration to runoff ratio at the basin scale which is important for groundwater resource assessment.

Summary of Third Manuscript

Designing monitoring networks that can discriminate among competing conceptual models is a key challenge for hydrologists. In an effort to address this issue, the third paper investigated the impact of network design on the utility of measurements for constraining hydrologic prediction uncertainty. The measurement network design approach presented is most appropriate for hydrologic assessments for which it is impractical to modify the design of the measurement network repeatedly during monitoring. To evaluate the performance of the static network design approach presented, a hypothetical feasibility assessment for a proposed artificial recharge site is employed. Highly permeable sediments and a thick unsaturated zone are advantageous for artificial recharge as these conditions result in high infiltration fluxes and a large in-situ filtering capacity in the vadose zone. The effective hydraulic conductivity is a simple metric to describe the suitability of a site for artificial recharge, however, this value is difficult to determine based on limited core data.

The presented measurement design approach begins by identifying predefined sets of candidate measurements (temporal gravity change) at a single location at a series of measurement times. The model was re-calibrated to each possible combination of candidate measurements. These parameter values, for saturated hydraulic conductivity and the Brooks-Corey exponent, were subsequently used in the forward model to predict the depth of the wetting front at the end of the monitoring period, and to predict the time
for the wetting front to reach a depth of 200 m. This last prediction of interest is
designed as a proxy for site performance that represents a large fraction of the
unsaturated zone but still lies far enough above the water table to avoid interaction with
the saturated zone. An ensemble approach was used to assess the likely impact of
measurement error on prediction error and prediction uncertainty for different
combinations of candidate measurements (measurement sets). The ensemble of prediction
errors was translated to a probability-weighted performance cost for each measurement
set using a cost function. The ensemble approach applied in this study, while more
computationally demanding, focused on the consideration of the effect of measurement
errors on the design process.

The goal of this approach was to develop a method to make informed choices
among the many available models and measurements by comparing the cost of the
measurements with the benefit of improved hydrologic assessment. Specifically, the goal
was to identify which measurement sets are most likely to provide an optimal trade-off
between the reduction of the costs associated with the error and uncertainty of a
hydrologic prediction of interest and the cost of collecting additional measurements.

The monitoring design application of a hypothetical artificial recharge site
presented in the third paper captures some expected design elements. For example, the
method showed that measurements with high signal to noise ratios are included in the
optimal measurement set, defined as the set with the lowest total cost. Additionally, there
is a diminishing return in prediction improvement with the addition of more
measurements in the monitoring network. Finally, as expected, designing a network
based on the most-likely prediction leads to fewer measurements than designing the
network with consideration of prediction uncertainty. Furthermore, some general expectations regarding monitoring network design were confirmed. For example, the size and composition of the best measurement set is a function of the particular prediction of interest and of the subsurface conditions. This approach provides a method to consider these influences quantitatively and to use that analysis to guide network design.

A subtler aspect of the design approach with regard to the trade-off between prediction uncertainty and measurement set cost was illustrated by considering different soil textures. By comparing measurement sets for three different soil textures, or truth models, we saw that the size and composition of the best measurement set also depended on the relationship between the prediction of interest and the properties of the system. For example, finer soils, which had less certain and more biased predictions, support selection of fewer measurements because predictions were so poor that more measurements resulted in little improvement of the prediction of interest. This is an unexpected result that would not be evident with other monitoring design approach approaches that do not allow for the possibility that additional measurements will not be worthwhile.

The network design approach described in the third manuscript can be applied to any hydrologic problem, regardless of its complexity. The general approach presented herein could be modified easily to incorporate additional complexities such as model structural errors, multiple measurement methods, multiple measurement locations, competing hydrologic process models, and more complete measurement costs. However, as the network design becomes more complex it is important to consider whether the exhaustive analysis presented herein is warranted because the appropriate design
approach is always dictated by the complexity of the project goals and budget requirements.
REFERENCES


Scintrex CG-5 brochure, Micro-g LaCoste, Lafayette, Colorado, 2 p.


APPENDIX A. ASSESSING THE LIKELY VALUE OF GRAVITY AND DRAWDOWN MEASUREMENTS TO CONSTRAIN ESTIMATES OF HYDRAULIC CONDUCTIVITY AND SPECIFIC YIELD DURING UNCONFINED AQUIFER TESTING

Joan B. Blainey, Ty P.A. Ferré, and Jeff T. Cordova

Abstract

Pumping of an unconfined aquifer can cause local desaturation detectable with high-resolution gravimetry. A previous study showed that signal to noise ratios could be predicted for gravity measurements based on a hydrologic model. We show that although changes should be detectable with gravimeters, estimations of hydraulic conductivity and specific yield based on gravity data alone are likely to be unacceptably inaccurate and imprecise. In contrast, a transect of low quality drawdown data alone resulted in accurate estimates of hydraulic conductivity and inaccurate and imprecise estimates of specific yield. Combined use of drawdown and gravity data, or use of high quality drawdown data alone, resulted in unbiased and precise estimates of both parameters. This study is an example of the value of a staged assessment regarding the likely significance of a new measurement method or monitoring scenario before collecting field data.

Introduction

As an unconfined aquifer is pumped, the alluvium desaturates, causing a decrease in bulk density within the drawdown cone. Under some conditions, ground-based
Gravimeters are sensitive enough to measure the small changes in gravitational acceleration (referred to hereafter as gravity) caused by this change in subsurface density (e.g., Poeter, 1990; Damiata and Lee, 2006). It has been suggested that high-resolution gravimetry measured at the ground surface may offer a cost-effective method to augment aquifer tests in unconfined aquifers without the expense of installing additional monitoring wells or piezometers (Poeter, 1990; Damiata and Lee, 2006). However, the relationship between drawdown and the associated gravitational response is complex and indirect. Specifically, while the drawdown response depends on the local energy potential of the water at the piezometer, the gravity response depends on a spatially-weighted average of the subsurface density change integrated over a large measurement volume. That is, drawdown is a local measurement and gravity is a non-local measurement. Additionally, whereas drawdown is a direct measurement of the change of a state variable in the flow equation (energy potential), change in the gravity response is an indirect measurement that is used to infer a change in a state variable (water storage). As a result, it can be difficult to predict whether gravity measurements, used either alone or together with piezometric measurements, will be useful for hydraulic parameter estimation through aquifer tests.

In a recent paper, Damiata and Lee (2006) presented a general expression for the gravitational attraction of a drawdown cone based on the solution of Neuman (1972; 1973). They proposed that successful use of gravity to monitor an unconfined aquifer test depended on the presence of a sufficiently large gravity signal relative to the resolution of gravity measurements (sufficient signal to noise ratio). Specifically, Damiata and Lee (2006) examined the suitability of gravity measurements for aquifer testing in a shallow
unconfined aquifer based on the detectability of the gravity signal given present-day gravimeter resolutions of approximately 1 microGal ($1 \mu\text{Gal} = 1 \times 10^{-8} \text{m/s}^2$) with portable gravimeters and 0.01 to 0.1 $\mu\text{Gal}$ with fixed-station (absolute) gravimeters. They found that, in many cases, current gravimeters have sufficient resolution to detect drawdown; they also identified those conditions that limit the signal to noise ratio and, therefore, the likely utility of gravimetry.

Damiata and Lee (2006) demonstrated that modern gravimeters have an adequate signal to noise ratio to detect drawdown, which is a necessary condition for the use of gravity for aquifer testing. We extend the analysis of Damiata and Lee (2006) to examine whether the signal to noise ratio is sufficient to ensure successful hydraulic parameter estimation. Specifically, we examine whether the use of gravity measurements to constrain the analysis of an aquifer test leads to estimates of hydraulic conductivity and specific yield that are comparable to estimates based on measurements of hydraulic head. We hypothesize that differences may arise because gravity is an indirect measurement of water distribution and because gravity has a large support volume with a non-uniform spatial sensitivity, whereas pressure head measurements are direct and local. We expect that these characteristics of gravimetry may give rise to different parameter sensitivities and parameter interactions, which may affect the quality of parameter estimates. We examine the use of gravity measurements for constraining parameter estimations both independently and in combination with piezometric measurements. To facilitate comparison of our assessments with those of Damiata and Lee (2006), we adopted many of the simplifications of their hydrologic model (homogeneous, isotropic aquifer, no
delayed drainage), their reference set of hydrologic conditions, and their ranges of hydraulic parameters.

Although this paper focuses on the assessment of the likely utility of gravity for aquifer testing, we suggest that a similar analysis could and should be applied before collecting measurements with any new instrument or when applying an accepted method to new hydrologic conditions. Specifically, we recommend an assessment performed in three stages. First, an analysis such as that presented by Damiata and Lee (2006) should be completed to ensure that the measurement method is likely to detect signals associated with the hydrologic process of interest. This analysis should consider both measurement resolution and expected measurement errors. Second, a first-level assessment of the likely utility of the method, such as that presented here, should be conducted to determine whether the measurements are likely to be useful for identifying hydraulic properties or hydrologic processes. For this analysis, it is most informative to adopt conditions that are favorable for the successful use of the method. If the analysis suggests that the measurement method is unlikely to be successful under these favorable conditions, it is unlikely that the method will be useful under more challenging conditions. If the analysis suggests that the measurement method is likely to be successful under some conditions, a third, more complete, site-specific analysis should be performed using all available information before collecting field measurements. This approach can provide an objective screening of proposed measurement methods for specific hydrologic applications, reducing the number of unsuccessful uses of geophysical and other indirect methods.
Numerical Modeling

Modeling the gravity response to pumping in an unconfined aquifer requires that a model of aquifer response be coupled with a model of gravimeter response. Damiata and Lee (2006) simulated the water table response to pumping in an unconfined aquifer using Neuman’s (1972; 1973) radially symmetric solution for drawdown assuming an instantaneous release of groundwater above the moving water table. The gravitational response to the calculated changes in water table elevation was simulated with the integral expression derived by Damiata and Lee (Equation (1), p. 350). The upper limit of integration, defined as the maximum radius to the edge of the drawdown cone, was determined with a root-finding algorithm (Press et al., 1992, pp. 359-360). We employed a conceptually similar but different approach to simulating drawdown and the associated gravitational response. To simulate drawdown in an unconfined aquifer, a well-documented, widely-used numerical model (Moench, 1996; Barlow and Moench, 1999) was used which will be referred to as the Moench model. For direct comparison with Damiata and Lee (2006), instantaneous drainage was simulated. That is, by employing the Moench model with the option of an instantaneous release of water from storage, the model is identical to the Neuman (1972; 1973) solution. Because the Moench model allows for consideration of delayed drainage, this effect could be added for a site-specific assessment, if warranted.

Numerical modeling was performed to simulate the gravitational response to the reduction in subsurface water mass caused by pumping in an unconfined aquifer. Assuming that the compressibilities of the aquifer and of the pore water are negligible with regard to the gravitational response, the change in bulk density depends only on the
change in volumetric water content throughout the cone of depression. That is, as in the work of Damiata and Lee (2006), the gravitational response to pumping is attributed to a reduction in subsurface density due to instantaneous and complete drainage above the water table. The aquifer is further assumed to be homogenous and isotropic with physical dimensions and hydraulic parameters identical to the reference set of test conditions presented by Damiata and Lee (2006, Table 1). For this reference set, the depth to the static water level, $z_s$, is 25 m below ground surface (bgs), the initial saturated thickness of the aquifer, $b$, is 50 m, and the storativity, $S_i$, is 0.001 (Table 1). The hydraulic conductivity, $K$, is 0.0001 m/s, and the specific yield, $S_y$, is 0.25, which is representative of a sand. The pumping rate, $Q$, is 0.06309 m$^3$/s (1000 gallons per minute (gpm)) for a fully penetrating, fully screened pumping well; this pumping rate is typical of that used for a municipal supply well. With the Moench model, the well can be treated as a line sink (infinitesimal well diameter) in the manner of Neuman (1972; 1973), or as a finite diameter pumping well; we used the latter option for a pumping well with a radius of 0.1 m. The piezometric response after seven days of pumping was simulated in nine fully penetrating observations wells located at radial distances ranging from 5 to 300 m from the pumping well.

In this application, the gravity response is a relative gravity response defined as the change in the vertical component of the gravitational acceleration at a point on the ground surface compared to the vertical gravitational acceleration at that location before pumping began. Analytical expressions exist to describe the gravity response to simple bodies, which lead to greater computational efficiency than the numerical integration approach presented by Damiata and Lee (2006) wherein the gravitational attraction is that
of a solid of revolution about the pumping well. This computational efficiency is especially important for inverse analyses. We chose to calculate the gravity response using a solution based on the superposition of semi-infinite cylindrical shells, which makes use of the relatively simple solution for the gravitational responses of semi-infinite cylindrical shells (Telford et al., 1990). For this approach, the gravity response is calculated as the sum of the responses to a series of concentric circular cylindrical shells centered about the pumping well, which closely approximates the shape of the drawdown cone. The bottom of each shell extends to infinite depth. For each radial increment, there is a shell with a density equal to negative one multiplied by the specific yield that extends upward to the initial horizontal water table elevation; there is a coincident shell with a density equal to the specific yield that extends upward to the average water table elevation across that radial increment at the time of observation. The superposition of these two concentric cylinders results in a finite shell, extending from the initial water table elevation to the current water table elevation, with a density of negative one multiplied by the specific yield. The shells are adjacent, such that the resultant finite cylinders fill the drawdown cone entirely. The contribution of each cylinder to the gravity response is:

\[
g(r, \theta) = 2\pi \rho R \left[ (R/2r) - (R/2r)^3 P_2(\mu) + 2(R/2r)^5 P_4(\mu) - \ldots \right] \quad r > z > R \text{ or } r > R > z \quad (1)
\]

\[
g(r, \theta) = 2\pi \rho R \left[ 1 - 2(r/2R)P_1(\mu) + 2(r/2R)^2 P_2(\mu) - 2(r/2R)^4 + \ldots \right] \quad R \geq r \geq z \quad (2)
\]
where \( z \) is the vertical distance from the ground surface to the top of the cylinder; \( R \) is the radius of the cylinder; \( r \) is the direct distance between the gravimeter location and the center of the top of the cylinder; \( \cos \theta \) is defined as \( z/r \); \( \gamma \) is the universal gravitational constant, which has a value of \( 6.672 \times 10^{-11} \) N m\(^2\)/kg\(^2\); and \( \rho \) is the density of the cylinder. \( P_n(\mu) = \cos \theta \) is a series of Legendre polynomials with coefficients determined from binomial expansion (Telford et al., 1990). A comparison of our solutions to those employed by Damiata and Lee (2006) for the base case conditions showed differences no greater than 1.098 \( \mu \)Gal in the predicted gravitational response after seven days of pumping (Damiata and Lee, 2006 Figure 2 and Figure 1, this paper).

Like Damiata and Lee (2006), we examined the sensitivity of the gravity response associated with pumping in an unconfined aquifer to changes in the specific yield and hydraulic conductivity. Specific yield values were varied between 0.15 and 0.35, which represents a textural variation from silt to gravelly sand (Fetter, 1994). In the work of Damiata and Lee (2006), the hydraulic conductivity was varied over two orders of magnitude to represent this same textural range (Freeze and Cherry, 1979). However, in this work the lower bound of the hydraulic conductivity was limited so that drawdown in the pumping well was never greater than the initial saturated thickness of the aquifer (Table 1). There is some question whether a drawdown solution, such as the Moench model or the Neuman (1972; 1973) solution, can be applied if the change in the saturated thickness of the aquifer due to pumping is not small compared with the initial saturated thickness of the aquifer. However, we chose to retain the aquifer conditions examined by Damiata and Lee (2006) for comparison. Furthermore, because we generate synthetic
data with the same model that we use to interpret the synthetic data, any limitations of the model would not result in model structural errors that would impact parameter estimates.

Drawdown and gravity transects are plotted (Figure 1) after seven days of pumping for the maximum and minimum hydraulic conductivity and specific yield values considered. For the largest value of hydraulic conductivity considered (0.001 m/s), the water level gradient near the well is shallow and drawdown is less than 1 m at a distance of 5 m from the pumping well. Because the drawdown is small, the gravity response is also small for high hydraulic conductivity, regardless of the specific yield value. For a small hydraulic conductivity (0.000032 m/s), the water level gradient is steep near the well. This leads to a larger gravity response that is strongly dependent on the value of the specific yield, ranging from -39 μGal to -68 μGal for specific yield values of 0.15 and 0.35, respectively at a distance of 5 m from the pumping well.

The dependence of the gravity response on the hydraulic conductivity and specific yield can be understood based on the physical basis of the gravity method. Specifically, decreasing the specific yield resulted in increased drawdown at every location, but it also led to a smaller decrease in the local change in bulk density due to drainage. Together, these effects caused dewatering to occur at greater depths, increasing the distance to the center of mass of the density decrease, and thereby decreasing the magnitude of the gravity response. In contrast, an increase in the hydraulic conductivity led to decreased drawdown with no difference in the local density change, which resulted in decreased gravitational response with increased hydraulic conductivity.
The drawdown and gravity responses can be shown more completely through response surfaces (Figure 2). For each response surface, parameter values for specific yield and hydraulic conductivity were varied simultaneously over the range of parameter values reported in Table 1. The drawdown (top row) and gravity change (bottom row) were plotted for each parameter pair and contoured (Figure 2). Response surfaces are shown for measurements made 3, 15, and 100 m from the pumping well.

For each measurement location, decreasing the specific yield or decreasing the hydraulic conductivity resulted in enhanced drawdown. For the specified parameter ranges, hydraulic conductivity had a stronger influence than specific yield on drawdown responses, as indicated by the sub-parallel contours on the sensitivity map for all but the most distant monitoring location (Figure 2). As the hydraulic conductivity decreased, the gradient near the pumping well increased, but the density change and volume of the drawdown cone were unaffected. As a result, as the hydraulic conductivity decreased, drawdown near the well increased and monitoring points located farther from the well (not shown) experienced decreased drawdown. The drawdown at every observation location increased as the specific yield decreased.

The contours on the gravity response surfaces are not parallel to the specific yield axis for any of the observation locations, indicating that specific yield had a much stronger influence on the gravity response than on the drawdown response (Figure 2). In general, the response patterns among measurement locations were more similar for the gravity response than for the drawdown response; this may be due to the distributed spatial sensitivity of the gravity response compared to the point scale of the water level measurements.
Specific yield had a stronger influence on the gravity response surfaces than on the drawdown response surfaces, especially near the pumping well, suggesting that gravity measurements may be useful for estimation of specific yield. However, specific yield and hydraulic conductivity show stronger correlation in the gravity response (Figure 2), suggesting that gravity measurements alone may not be capable of resolving both parameters with certainty. To examine how the drawdown and gravity measurements differ in terms of constraining parameter estimates, the two data types were used both independently and jointly to estimate hydraulic conductivity and specific yield. When using the two measurement types jointly, both simultaneous and sequential parameter estimation were considered. For sequential estimation, drawdown measurements were used to estimate hydraulic conductivity and then gravity measurements were used to estimate specific yield with the value of hydraulic conductivity fixed at the previously estimated value.

Due to the availability of automated piezometric logging, it is most common to collect drawdown data at few locations for many times during an aquifer test. Gravity could be collected as time series data but most gravimeters are not amenable to automated logging. In addition, gravity measurements cannot be made as rapidly as drawdown measurements; it would be difficult to make accurate gravity measurements at the beginning of an aquifer test when water levels change rapidly. Therefore, to examine gravity under advantageous conditions for this first-level assessment, we chose to consider data collected at a single time at multiple locations. For the reference set of conditions, the detection distance is 300 m for a portable gravimeter with a resolution of
Therefore, nine ground surface observation points for drawdown and gravity measurements were located at radial distances of 5.0, 8.3, 13.9, 23.2, 38.7, 64.6, 107.8, 179.8, and 300.0 m from the pumping well. Parameter estimation was conducted using a coupled hydrogeophysical approach (Ferré et al., 2006); that is, during each step of the inversion process, the gravity response was calculated for the predicted hydrologic response and compared directly to the (synthetic) gravity observations.

**Measurement Uncertainty**

The utility of measurements for parameter estimation depends on measurement error (Poeter and Hill, 1997). We propose that the value of the measurement also depends on the distributed spatial sensitivity of the measurement method, likely decreasing with an increase in the measurement volume. Both of these sources of estimation uncertainty were examined independently using synthetic measurements. Specifically, synthetic measurements were produced based on the noise-free drawdown and gravity responses calculated for the Damiata and Lee (2006) base case parameter set after seven days of pumping. Measurement errors were generated for drawdown and gravity data (one standard deviation equal to 1μGal) and were added to the noise-free responses to produce synthetic data. Note that the measurement errors (repeatability) presented here are significantly different than the instrument resolution for signal detectability presented by Damiata and Lee (2006), as summarized in the Introduction.

In practice, gravity measurements must be adjusted using standard gravity corrections (Telford et al., 1990). In this application, we are interested in relative (not
absolute) differences in gravity due to water level changes. Many of the standard
ground corrections (e.g., Bouguer and elevation corrections) are not required because the
correction would be identical for measurements made at each location before and after
pumping. We assumed that corrections that would affect a relative gravity survey (e.g.
earth-tide and instrument drift) have been applied correctly. With this assumption, we
used the manufacturer-reported measurement repeatability to describe the measurement
error.

For this first-level assessment of measurement utility, we intentionally separated
measurement error and model structural error. Specifically, aquifer responses were
generated numerically and measurement error was added; we then interpreted the aquifer
responses using the same model that was used to generate the synthetic data. Following
this approach and assuming that the measurements were collected correctly and the
appropriate gravity corrections were applied, we defined measurement errors as
independent with normally distributed errors and zero bias. In contrast, if a different
numerical model were used to generate the aquifer response than was used to interpret it,
or if an incorrect or incomplete model were used to interpret field data, correlations could
appear among measurement errors. For example, if measurements indicated a delayed
drainage response of the aquifer, and the numerical model used to generate aquifer
responses did not account for delayed drainage, the density change in the vadose zone
would not be equal to the specific yield until drainage was complete. This could
introduce a common error to all gravity measurements, which would lead to correlations
among measurement errors. For this first-level assessment, we chose to ignore delayed
drainage. But, this issue could be considered during a site-specific analysis. Similarly, a
more complete and detailed site-specific model could address the possible impacts of heterogeneity, anisotropy, well-bore storage effects, and uncertainties in flow boundary conditions.

Measurement errors were generated using a standard normal distribution and a specified measurement uncertainty with the same units as the units of the measurement. For each simulation, 100 random realizations of synthetic measurement data were computed to ensure that conclusions from parameter estimation were not dependent on a single measurement realization. A sequence of nine normally distributed random numbers with a mean of 0 and a standard deviation of 1 was generated for each realization, one for each measurement location. Then, the measurement uncertainty for each measurement type, specified as a standard deviation, was multiplied by this series to give an identical, but scaled, sequence of measurement errors for drawdown and gravity. The gravity errors were lagged by one realization relative to the drawdown error so that an identical set of errors was applied for each measurement type across all realizations, but the drawdown and gravity errors were uncorrelated for each realization.

The gravity measurement uncertainty (one standard deviation) was set to 1 μGal, which is a conservative estimate of the maximum achievable repeatability of an absolute gravimeter (FG-5 absolute gravimeter, Micro-g LaCoste, Lafayette, CO, mention of brand names does not constitute an endorsement by the U.S. Geological Survey) reported as 1 μGal (two standard deviations). We assumed the gravity measurements were uncorrelated with a zero mean. (Note that any consistent bias in gravity responses would have resulted in unbiased relative gravity measurements because only changes from background values are considered.) To isolate the effects of spatial averaging from the
effects of different measurement errors among the methods, the measurement error for
drawdown was defined such that it resulted in a signal to noise ratio that was comparable
to that of the gravity measurements. That is, the drawdown measurement uncertainty, $\sigma_d$,
was calculated as:

$$\frac{\sigma_g}{g_{\text{meas}}} = \frac{\sigma_d}{d_{\text{meas}}},$$

(3)

where $\sigma_g$ represents the gravity measurement uncertainty. The model outputs $g_{\text{meas}}$ and
$d_{\text{meas}}$ are the average error-free gravity and drawdown responses over all of the
observation locations after seven days of pumping. Based on a gravity measurement
uncertainty of 1 $\mu$Gal, a drawdown measurement uncertainty of 0.24 m was calculated.
An example of a measurement transect is shown with these errors for gravity and
drawdown (Figure 1). To examine the combined impact of the higher measurement error
and spatially distributed sensitivity of gravity compared to drawdown, a second
drawdown measurement error level of 0.048 m was considered as well. This drawdown
error level is greater than the measurement resolution of a standard water level tape, but it
is a conservative estimate of the combined effects of well interference, operator error, and
well completion uncertainties on the quality of typical hydraulic head data.

Objective Function

The closeness of the fit between model predictions and the drawdown and gravity
data was calculated with a single objective function. In the objective function,
measurements were normalized by the average signal to account for the differing units of drawdown and gravity, and assigned weights. Because normalizing the measurements accounts for the differing units of drawdown and gravity, the primary function of the objective function weighting is to reduce the influence of measurements that are less informative and increase the influence of measurements that are more informative. Ideally, an individual weight would be applied to each measurement based on prior information about the value of that measurement. However, in general, little a priori information is available to determine the quality of each measurement. Lacking this specific information, it can be assumed that the precision of each measurement depends primarily on the measurement type, allowing for the definition of a single weighting factor for each measurement type. With this simplification, the objective function was:

\[
\text{ObjF} = w_g \left[ \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \left( \frac{g_{\text{meas}} - g_{\text{sim}}}{g_{\text{meas}}} \right)^2 \right] + w_d \left[ \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \left( \frac{d_{\text{meas}} - d_{\text{sim}}}{d_{\text{meas}}} \right)^2 \right]
\] (4)

where \( w_g \) is the weight on the squared residuals of the normalized gravity data, \( w_d \) is the weight on the squared residuals of the normalized drawdown data. The integers \( n_r \) and \( n_t \) are the numbers of observation well locations and measurement times, respectively. For this study, the observation locations and measurement times were identical for drawdown and gravity with \( n_r = 9 \) and \( n_t = 1 \).

To demonstrate the contributions of drawdown and gravity to parameter estimation, we show an error surface for each component of the objective function (Fig 3, top and middle panels). These plots show the square of the residuals normalized by the
average measurement signal calculated based on error-free measurements. The results demonstrate that drawdown measurements and gravity measurements provide different information about the parameters (Figure 3). The drawdown portion of the objective function (top panel, Figure 3) shows a broad, sub-horizontal minimum that indicates low sensitivity to the specific yield parameter. There is also higher sensitivity for small values of hydraulic conductivity and specific yield. The gravity component of the objective function (middle panel, Figure 3) also has a broad, poorly defined minimum. The gravity measurements are more sensitive to the specific yield parameter than are the drawdown measurements as indicated by the slope of the contours near the minimum of the objective function. The gravity component of the objective function surface indicates high sensitivity for small values of hydraulic conductivity and large values of specific yield. Finally, there is strong correlation between the hydraulic conductivity and the specific yield. The bottom panel (Figure 3) shows the combined objective function with equal weighting applied to the drawdown and gravity components of the objective function. The closed contours of the combined error surface near the minimum of the objective function indicate an improved objective function surface, demonstrating the value of combining gravity and drawdown observations.

The assignment of weights on each measurement type in the objective function is used to illustrate the relative value of each measurement type for constraining parameter estimation. One common approach to combining different measurement types in a single sum of squares objective function is to weight each measurement type by the inverse of the expected variance of the measurement error for that measurement type (Poeter and Hill, 1997). This weighting scheme essentially equates the value of each measurement
type with the observation uncertainty, which may be a combination of the measurement error and other processes that contribute to the observation uncertainty. In this synthetic study, the observation uncertainty is well defined and equal to the measurement error. However, when considering field data there is considerable subjectivity in the estimation of observation uncertainty.

For ease of comparison among observation sets, we calculate relative weighting factors, $W_g$ and $W_d$, which we define as:

$$W_d = \frac{w_d}{w_d + w_g} \quad \text{and} \quad (5)$$

$$W_g = \frac{w_g}{w_d + w_g}. \quad (6)$$

The relative weighting factors sum to one. If drawdown data are used independently, $W_d = 1$ and $W_g = 0$. If gravity data are used independently, $W_d = 0$ and $W_g = 1$. For the low quality drawdown data, the signal to noise ratio of the drawdown data was chosen to be equal to that of the gravity data (Equation 3), giving $W_d = W_g = 0.5$. For the high quality drawdown data, $W_d = 25W_g$, or equivalently, $W_d = 0.96$ and $W_g = 0.04$.

**Optimization Algorithm**

The common logarithm of hydraulic conductivity and the specific yield parameter were varied either simultaneously or sequentially to find the optimal fit between the observed and measured data as defined by the objective function. The Nelder-Mead simplex algorithm (Nelder and Mead, 1965; Lagarias et al., 1998) was used to find the minimum of the objective function. The Nelder-Mead simplex algorithm is a direct local
search method for multidimensional unconstrained minimization; neither numerical nor analytic derivatives are employed to approximate the function of interest for minimization (Nelder and Mead, 1965; Lagarias et al., 1998). In $n$-dimensional space, this algorithm is characterized by $n+1$ vectors that are the vertices of the simplex. For each step of the search, a new point is generated in or near the simplex. The objective function value of the new point is compared to the objective function value of the vertices in the existing simplex and through a sequence of logical steps, the new point may replace an existing vertex of the simplex. The search is repeated until the diameter of the simplex is less than the specified tolerance, which was set to $1 \times 10^{-4}$.

Results of Parameter Estimation

The utility of measurements for constraining parameter estimates can be described based on the average bias and variance among the 100 realizations examined. Parameter estimation converged for each realization based on only hydraulic head measurements, combined gravity and hydraulic head measurements, or sequential estimation. However, when only gravity information was used, unconstrained inversion led to unacceptably poor convergence. Therefore, we constrained the inversion by implementing a penalty function on high estimates of specific yield ($\geq 0.40$) for these cases. The penalty function increased the value of the objective function calculated with Equation (4) by 1.25%.

Only parameter estimates based on gravity data alone showed significant absolute biases ($\geq 10\%$) in both parameter estimates (Table 2). Use of only drawdown measurements with high error ($\sigma_d = 0.24$), led to 3% bias in the specific yield estimate (Table 2). For the simple hydrologic conditions considered (homogeneous and isotropic
porous medium, constant pumping with no boundary effects), all of the data sets that included drawdown measurements led to bias of less than 1% in the hydraulic conductivity estimates (Table 2), which is consistent with the drawdown response surfaces that showed greater sensitivity to hydraulic conductivity than specific yield.

The absence of bias in the estimates for data sets that included drawdown measurements suggests that good quality drawdown data ($\sigma_d < 0.048$) or combined drawdown (either low or high quality) and gravity data can be used to estimate the hydraulic conductivity and specific yield with an acceptably small bias (< 0.3%), on average. However, aquifer tests are rarely repeated in practice. Hence, we are interested in whether a data set, based on any single aquifer test, is likely to be informative. Therefore, it is more meaningful to examine the variability of these parameter estimates among the realizations analyzed. The 95% confidence interval (mean plus and minus two standard deviations) was selected to describe the most likely range of estimates that would be found for any given aquifer test. However, it should be stressed that this can only be used to assess the likely performance of an aquifer test; the results do not describe the actual parameter bias or uncertainty that would be found using any single measured data set.

Uncertainties of the hydraulic conductivity estimates based on drawdown data alone were insignificant for most hydrologic applications (95% confidence interval bounds were 10% and 2% of the mean for lower and higher quality data, respectively). Use of gravity data alone led to unacceptably large uncertainties in the hydraulic conductivity estimate (95% confidence interval bounds were 130% of the mean) (Table 2, Figure 4). The addition of gravity data to the drawdown data resulted in a slight
decrease in the uncertainty of the hydraulic conductivity and specific yield estimates estimate. The effect of the drawdown error level when used together with gravity data was minor (95% confidence interval bounds ranged between 5% and 1% of the mean for the low and high quality drawdown data, respectively) (Figure 4). This raises the interesting possibility that adding gravity data may reduce the required accuracy of water level measurements, allowing for the use of existing infrastructure wells for pumping tests.

The uncertainties in the specific yield estimates based on the low quality drawdown data were large (95% confidence interval bounds were 0.15 to 0.36) (Table 2, Figure 4). Uncertainties in the specific yield estimates based on gravity data alone were unacceptably large (67% confidence interval bounds were 0.07 to 0.37) (Table 2, Figure 4). The addition of gravity data to the drawdown data led to lower uncertainties in the specific yield estimate than those using either data type independently (95% confidence interval bounds were 0.232 to 0.269 and 0.238 to 0.263 for low and high quality drawdown data, respectively) (Table 2, Figure 4).

The definition of the relative weights for each measurement type in the objective function based on the variance of the measurement error led to $W_g$ values of 0.5 and 0.04 for the low and high quality drawdown data, respectively. Use of these weights in the objective function led to acceptably low biases of both parameters for both drawdown error levels (Figure 4). However, there is very little sensitivity of the parameter uncertainty or parameter bias to the value of $W_g$ in the range of 0.04 to 0.5 (results not shown). This suggests that accurate assessment of the value of the observation
uncertainties under field conditions, and thereby the weights in the objective function, is not critical to effective use of both measurement types for the conditions examined.

Sequential parameter estimation retained the accurate and precise estimates of hydraulic conductivity based on drawdown data alone. For the low quality drawdown data, the 95% confidence interval bounds on specific yield were similar for both sequential estimation and simultaneous estimation (0.238 to 0.266 and 0.232 to 0.269, respectively). For the high quality drawdown data, the 95% confidence interval bounds on specific yield were nearly identical for both sequential estimation and simultaneous estimation (0.236 to 0.264 and 0.238 to 0.263, respectively). Sequential estimation resulted in slightly less certain estimates of hydraulic conductivity with little impact on the uncertainty of the specific yield estimate compared with simultaneous parameter estimation. Overall, for the case study presented, the differences between sequential and simultaneous estimation were trivial.

Concluding Remarks

Through a synthetic example, we demonstrated that the capacity of an instrument, in this case a gravimeter, to detect a signal above the noise in the data is not a sufficient condition to guarantee utility of the measurement method to constrain hydrologic parameter estimation. We suggested that the utility of a measurement also depends on parameter correlations that arise due to the relationship of the measured property to the hydrologic process of interest. Furthermore, we showed that these correlations can be determined through a hydrogeophysical assessment.
Our specific findings were that a transect of gravity measurements collected after seven days of pumping, used independently, were not likely to lead to accurate parameter estimates. However, gravity measurements (non-local, indirect measures of changes in storage), when used with drawdown measurements (local, direct measures of changes in energy potential), may significantly improve the quality of specific yield estimates from unconfined aquifer tests. The quality of the estimates was similar for high- and low-quality drawdown data if they were interpreted jointly with gravity data, which suggests that the addition of gravity data could reduce the required accuracy of drawdown measurements for aquifer testing.

For the synthetic case studies presented, the observation error is identical to the measurement error. This is not true for field applications; however, our results indicate that exact quantification of the observation errors is not essential for effective use an indirect, non-local measurement. Our results indicate that it is important to conduct a synthetic case study analysis before applying measurement methods to novel applications. This numerical investigation can determine whether the effects of spatially distributed sensitivity are likely to control the value of indirect measurements for parameter estimation. It is difficult, if not impossible, to draw these conclusions without performing a coupled hydrogeophysical analysis. This study further refines the areas of opportunity and limits to the application of gravity methods for aquifer testing. We suggest that virtual experiments like this should be performed routinely when assessing the likely utility of indirect measurements for hydrologic analysis.
Acknowledgements

Financial assistance provided to the first author by the Arizona Water Science Center of the U.S. Geological Survey is gratefully acknowledged. We also acknowledge support from the Technology and Research Initiative Fund Water Sustainability Program. The authors wish to thank the editor and anonymous reviewers for their comments and suggestions which improved this manuscript. Constructive comments on the original manuscript were also provided by Kyle Blasch. We thank Mary Hill and Eileen Poeter for their comments and suggestions during revision of this manuscript.

References


### Tables

<table>
<thead>
<tr>
<th>Property or parameter</th>
<th>Property or Parameter value</th>
<th>Parameter range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thickness of Aquifer (m)</td>
<td>50.0</td>
<td>-</td>
</tr>
<tr>
<td>Depth to Static Water Level (m bgs)</td>
<td>25.0</td>
<td>-</td>
</tr>
<tr>
<td>Hydraulic Conductivity (m/s)</td>
<td>0.0001</td>
<td>0.000032 – 0.001</td>
</tr>
<tr>
<td>Storativity</td>
<td>0.001</td>
<td>-</td>
</tr>
<tr>
<td>Specific Yield</td>
<td>0.25</td>
<td>0.15 – 0.35</td>
</tr>
<tr>
<td>Radius of Borehole (m)</td>
<td>0.1</td>
<td>-</td>
</tr>
<tr>
<td>Density of Groundwater (kg/m³)</td>
<td>1000.0</td>
<td>-</td>
</tr>
<tr>
<td>Flow Rate (m³/s)</td>
<td>0.06309^a</td>
<td>-</td>
</tr>
<tr>
<td>Duration of Pumping (days)</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>Radial observation locations (m) from</td>
<td>5.0, 8.3, 13.9, 23.2, 38.7,</td>
<td>-</td>
</tr>
<tr>
<td>pumping well for parameter estimation</td>
<td>64.6, 107.8, 179.8, 300.0</td>
<td>6.309 × 10⁻² m³/s = 1000 gpm</td>
</tr>
</tbody>
</table>

^a 6.309 × 10⁻² m³/s = 1000 gpm.

Table 1. Parameter values and parameter ranges used for sensitivity analysis.
<table>
<thead>
<tr>
<th>Data Type</th>
<th>Data</th>
<th>Specific Yield (-)</th>
<th>Hydraulic Conductivity (m/s × 10^{-4})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(mean ± 2 st. dev.)</td>
<td>(mean ± 2 st. dev.)</td>
</tr>
<tr>
<td>True parameter values</td>
<td>0.25</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Single data type</td>
<td>Drawdown (σ_d = 0.24 m)</td>
<td>0.2574 ± 0.1029</td>
<td>1.000 ± 0.1029</td>
</tr>
<tr>
<td></td>
<td>Drawdown (σ_d = 0.048 m)</td>
<td>0.2508 ± 0.02038</td>
<td>0.9996 ± 0.02125</td>
</tr>
<tr>
<td></td>
<td>Gravity</td>
<td>0.2228 ± 0.2971</td>
<td>0.8846 ± 1.298</td>
</tr>
<tr>
<td>High drawdown error</td>
<td>Drawdown and gravity (W_g = 0.5)</td>
<td>0.2506 ± 0.01817</td>
<td>1.001 ± 0.04738</td>
</tr>
<tr>
<td></td>
<td>Sequential estimation</td>
<td>0.2521 ± 0.01418</td>
<td>0.9991 ± 0.1056</td>
</tr>
<tr>
<td>Low drawdown error</td>
<td>Drawdown and gravity (W_g = 0.04)</td>
<td>0.2504 ± 0.01226</td>
<td>0.9999 ± 0.01420</td>
</tr>
<tr>
<td></td>
<td>Sequential estimation</td>
<td>0.2503 ± 0.01408</td>
<td>0.9996 ± 0.02126</td>
</tr>
</tbody>
</table>

Table 2. Summary of parameter estimation results for hydraulic conductivity and specific yield based on 9 measurement locations and the reference set of test conditions and measurement errors of either 0.24 m or 0.048 m, and 1 μGal for drawdown and gravity, respectively.
Figures

Figure 1. Error free (lines) and simulated measurements (symbols) of (a) drawdown and (b) gravitational response after 7 days of pumping in an unconfined aquifer with $b = 50$ m, $K = 10^{-4}$ m/s, $Q = 6.309 \times 10^{-2}$ m$^3$/s, $S = 0.01$, $S_y = 0.25$, and $z_s = 25$ m. The measurement errors are simulated with 1 standard deviation equal to 0.24 m and 0.048 m, and 1 $\mu$Gal for drawdown and gravity measurements, respectively. (c) Drawdown and (d) gravitational responses for $K = 10^{-3}$ m/s and $S_y = 0.35$ (dotted curve); $K = 3.2 \times 10^{-5}$ m/s and $S_y = 0.35$ (dotted-dashed curve); $K = 10^{-3}$ m/s and $S_y = 0.15$ (solid curve); $K = 3.2 \times 10^{-5}$ m/s and $S_y = 0.15$ (dashed curve).
Figure 2. Simulated drawdown in m (top row) and gravitational response in $-1\mu$Gal (bottom row) after 7 days of pumping in an unconfined aquifer with $b = 50$ m, $K = 10^{-4}$ m/s, $Q = 6.309 \times 10^{-2}$ m$^3$/s, $S = 0.01$, and $z_s = 25$ m for varying values of hydraulic conductivity and specific yield. The left, middle, and right columns show results for observation points located 3 m, 15 m, and 100 m from the pumping well, respectively.
Figure 3. The drawdown component (top) and the gravity component (middle) of the objective function (sum of square errors normalized by the average measurement signal), and the combined objective function (bottom) without simulated measurement error for data collected after 7 days of pumping in an unconfined aquifer for the following conditions: $b = 50$ m, $K = 10^{-4}$ m/s, $Q = 6.309 \times 10^{-2}$ m$^3$/s, $S = 0.01$, $S_y = 0.25$, and $z_s = 25$ m.
Figure 4. Parameter estimates for simultaneous estimation of hydraulic conductivity and specific yield after 7 days of pumping in an unconfined aquifer for the following conditions: $b = 50$ m, $K = 10^{-4}$ m/s, $Q = 6.309 \times 10^{-2}$ m$^3$/s, $S = 0.01$, $S_y = 0.25$, and $z_s = 25$ m. Parameter estimates are fit to simulated measurements of drawdown, gravity, or both drawdown and gravity at 9 radial distances from the pumping well. (a) Bias (%) and (b) Two standard deviations. For the higher error level, errors are simulated with 1 standard deviation equal to 0.24 m and 1 \( \mu \)Gal for drawdown and gravity measurements, respectively. For the lower error level, errors are simulated with 1 standard deviation equal to 0.048 m and 1 \( \mu \)Gal for drawdown and gravity measurements, respectively. Each plotted point is based on 100 random measurement realizations.
APPENDIX B. INFILTRATION ON ALLUVAL FANS IN ARID ENVIRONMENTS: INFLUENCE OF FAN MORPHOLOGY

Joan B. Blainey and Jon D. Pelletier

Abstract

Mountain-front recharge through highly permeable alluvial fans can be an important source of groundwater recharge in arid climates. To better understand the geomorphic factors (e.g., fan slope, fan area, active channel proportion of fan area, sediment permeability, and entrenchment of the active channel) that control flow and infiltration on alluvial fans, we developed a coupled numerical model of steady surface water flow and Green-Ampt-type infiltration. The model was applied to synthetic alluvial fans using random walkers to create connected distributary networks. The purpose of this approach is to predict where and how recharge occurs on fans as a function of fan morphology. Using the numerical model, we examined how the fan shape and the sequence of fan surfaces influenced where infiltration occurred on the fan. We also investigated how fan morphology influenced the partitioning of infiltration between the fan and the valley floor. Finally, we examined how infiltration influenced the spatial distribution of flooding. The greatest amount of infiltration occurred on low gradient fans where water spread laterally with shallower ponded water depths, although the large inundation area often included less permeable sediments outside of the active channel. The ratio of the incision depth to the input flow depth was an important predictor of the
amount of infiltration. The greatest amount of infiltration occurred on fans with incision
depths slightly smaller than the input flow depth. These results have implications for
groundwater resource assessment and for development of monitoring networks on fans in
arid environments.

Introduction

Flooding on alluvial fans, where highly permeable fan deposits surround a basin,
can be an important source of groundwater recharge in arid environments (Bull, 1977;
Hendrickx et al., 1991; Houston, 2002). Surface flows from upland catchments are
focused at the fan apex where channels distribute flow towards the valley floor over
gently sloping fan deposits. Prediction of the amount and location of infiltration on fans
is difficult due to complex flood behavior and the heterogeneous and poorly-constrained
permeability structure of fan sediments. Flooding on fans is typically characterized by
deep channel flow in several large channels with a complex distributary channel network
(National Research Council, 1996). Infiltration is controlled by the spatial distribution of
paleosols (Weissmann and Fogg, 1999; Bennett et al., 2006) that can lower permeability
values locally by up to two orders of magnitude relative to the parent material (McFadden
et al., 1987; McFadden et al., 1992; McDonald et al., 1996). Often only limited
hydrologic information is available to constrain predictions of distributed infiltration on
fans. Typically, only sparse streamflow data or flood stage information have been
collected due to infrequent flows of short duration coupled with the uncertainty of where
these flows will occur.
Mountain-front recharge through alluvial fans, an intermittent recharge mechanism, is not easily quantified for groundwater resource evaluation (Houston, 2002; Weissmann et al., 2002b) due to complex responses to climate change, climate variability, and changes in land use. Quantifying infiltration over large areas is complex because of the large spatial and temporal variability of infiltration observed whether using point-scale measurements (direct physical methods, indirect physical methods, environmental tracers) alone or in combination with larger-scale geophysical methods used to interpolate and extrapolate between point measurements (Liu et al., 1995; Scanlon et al., 1999; Massuel et al., 2006). On highly permeable alluvial fans, a combination of water level, soil texture, and hydrochemical data have been used in event-based distributed models to estimate water fluxes and travel times to the water table in arid environments (Houston, 2002; Massuel et al., 2006). For example, for a flash flood event that occurred during January 2000 on a fan in arid northern Chile, about 70% of the flow percolated to the underlying aquifer producing groundwater rises at the distal portion of the fan that persisted for 5 to 9 months (Houston, 2002).

Mountain-front recharge through alluvial fans is a function of morphology (topography, soil texture, vegetation) as well as channel and flow characteristics (Houston, 2002; Izbicki et al., 2002). Geomorphology provides an alternative approach to point-scale field measurements of infiltration to characterize infiltration over large areas in arid environments (Scanlon et al., 1999). Previous field studies in arid settings have shown that the spatial variability in infiltration is related to the geomorphic setting as distinguished by interdrainage area, topographic depression, and drainage area (Scanlon
et al., 1999) or by distance from the mountain front (Izbicki et al., 2002). The objective of this study is to examine the influence of fan morphology on infiltration on fans using simple observables. Specifically, we considered how the shape of the fan and the sequence of fan surfaces influence where infiltration occurs using a two dimensional (2-D) distributed numerical model of synthetic fans with coupled surface flow and infiltration. Another characteristic of fan morphology with significance for infiltration is the permeability of fan sediments, which is related to the depositional environment, fan age, soil development, and geology of the source basin (Weissmann et al., 2002a; Young et al., 2004; Winfield et al., 2006). Using the numerical model, we also examined how fan morphology, including sediment permeability, influences the partitioning of flow between on-fan flows and those that extend downslope of the fan environment. Finally, we examined how infiltration affected the spatial distribution of flooding during flow events. The purpose of this approach is to predict where and how recharge occurs on fans as a function of fan morphology. These results can be used to guide the development of monitoring networks on fans, guide geomorphic mapping for flood hazard prediction, and characterize the infiltration to runoff ratio at the basin scale which is important for groundwater resource assessment.

Conceptual Model of Alluvial Fans

Alluvial fans are sedimentary landforms that form at the base of mountain fronts where confined feeder streams emerge onto valley floors and deposit the sediment load of the streams. Alluvial fans develop where there is sufficient sediment supply, sediment
accumulation, and adequate relief for vertical fan growth. Alluvial fans comprise greater than 30% of the landscape of the southwestern United States (Antsey, 1965). The planform morphology of an alluvial fan is generally semi-circular whereas the three-dimensional morphology is conic; the fan apex is located at the point where the feeder stream emerges from the base of the mountain front. Deposition occurs as a feeder stream emerges from a steep channel near the mountain front due to a decrease in stream competence associated with the unconfined nature of flow on fans and, secondarily, due to a decrease in channel slope (Bull, 1977). If a fan is unrestricted by adjacent fans, distributary channels form a radial pattern emanating from the fan apex. Overall, radial topographic profiles are concave-upward or almost constant and cross-fan profiles are concave-downwards (Bull, 1977; National Research Council, 1996). Typical fan radii range from 5 to 15 km and vary in slope between 0.5 and 10 degrees (Rust and Koster, 1984).

In the Basin and Range physiographic province of the western United States, many alluvial fans are composed of a suite of terraces that rise from the active channel as a sequence of steps. Fans in Death Valley, California exemplify this classic form which is created by temporal changes in the sediment composition and sediment concentration in the channels draining the mountain front (Bull, 1977, 1991). During aggradational periods, the ratio of sediment to water flux is high and sediment is deposited across the fan, or in the available area adjacent to an older abandoned terrace, to form a low-relief deposit via repeated episodes of channel avulsion (Bull, 1979). In contrast, entrenchment occurs during periods with a lower ratio of sediment to water flux. Conditions that result
in a reduced ratio of sediment to water flux are numerous and include cooler and wetter climate conditions that re-vegetate hillslopes and anchor the regolith (McDonald et al., 2003).

On stream-dominated alluvial fans, incision at the fan apex produces a fan-head trench in which the longitudinal stream profile is at a lower elevation and lower gradient than the surrounding fan surface. The trench is deepest at the apex and becomes shallower with distance downfan. In general, the greatest difference in elevation between the active channel and older surface deposits occurs near the fan apex and may range over tens of meters. Near the toe of the fan, however, the difference between the oldest and youngest deposits may be less than 1 m. Shifts in the location of the intersection point, both upfan and downfan, over time create sequences on the alluvial fan.

Fan sediments are complex and heterogeneous (Weissmann and Fogg, 1999; Weissmann et al., 2002a) because deposition occurs by numerous surges of sediment-laden water and backfilling of channels. For example, fan sediments may be debris flow dominated at the apex while the distal portion of the fan may be dominated by fluvial sands. Many fans have near-surface sediments that progressively fine with distance away from the fan apex (Bull, 1964; Waters and Field, 1986) while other fans are poorly sorted and have planar bedding (Blair and McPherson, 1994). Previous studies of fans of the Mojave Desert found significant correlation between surface age and surface sediment texture (McFadden et al., 1987, 1992). For example, Young et al. (2004) found significant correlation between the surface age and texture of Quaternary soils from five sampling sites located on a broad alluvial fan in the Mojave National Preserve in
California. More than 90% of the variability in the saturated hydraulic conductivity of these samples was explained by the age of the soil (Young et al., 2004). Younger soils had higher sand contents and high hydraulic conductivities. As the surface age increased there was a reduction in the sand content and a corresponding increase in the content of silt and clay (Young et al., 2004). Young et al. (2004) concluded that a single-soil modeling approach for soil water retention was appropriate for both younger soils, which had no significant soil horizon development, and for older soils in which the wetting front remained in a fine-grained, gravel-poor vesicular A (Av) horizon.

The Av horizon, or any horizon that produces a larger permeability contrast with surrounding deposits, has important implications for infiltration and ecological processes at both intradrainage and interdrainage scales. The fine-grained Av horizon is a widespread surficial horizon of desert soils that almost consistently occurs below a desert stone pavement (e.g., McFadden et al., 1987, 1992) and is associated with soils forming in hot and arid climates. Fines entrained by the wind and dispersed over the landscape accumulate over long time scales to form this horizon. Entrapment of fine material is augmented by desert pavements due to the surface roughness which creates local air-flow turbulence and results in deposition of eolian material (Wells et al., 1985). The rough surface of the desert pavement also protects the dust from raindrop impact (McFadden et al., 1998). The source of the dust is often playas both in modern times (Reheis and Kihl, 1995) and during the Pleistocene-to-Holocene transition which may have resulted in an increase in dust accumulation and soil development as the climate become more arid with
a reduction in vegetation and the playas expanded (e.g., Wells et al., 1985; McFadden et al., 1992; Reheis et al., 1995).

Beneath the Av horizon of Late-Pleistocene soils, at depths of 3 to 4 cm, there is often a Bt argillic horizon that has developed from the downward movement of soluble or suspended material of high clay content from the A horizon into the B horizon. In such soils the Bt horizon extends to a depth of about 50 cm. And below the Bt horizon there is frequently a Bk or calcic horizon that develops from the accumulation of calcium carbonate in the soils which extends to a depth of 75 or 150 cm (McFadden et al., 1998).

More recently, Winfield et al. (2006) examined alluvial fan sediments from Oro Grande Wash and Sheep Creek Wash in the western Mojave Desert and noted the impact of textural and structural characteristics on the relation of water content to water potential and unsaturated hydraulic conductivity. Structure was defined as the arrangement of soils resulting from aggregate formation, depositional sorting, shrink-swell processes, and macropores from animal burrows and root channels. Texture was determined to have a greater influence than structure on the water retention properties of the water-laid and debris-laid sediments (Winfield et al., 2006).

In addition to sediment texture, infiltration on fans is also strongly dependent on the morphology of the fan drainage network which can be grossly characterized by the intersection point. Upfan of the intersection point the stream is incised into the fan surface and is part of a tributary drainage network while downfan of the intersection point the drainage network spreads across the fan and deposition can occur (Hooke, 1967). The drainage network on alluvial fans is not static on human time scales (e.g., National
Research Council, 1996; Klawon and Pearthree, 2000; Field, 2001). When floods deposit large quantities of sediment the conveyance capacity of the channel is reduced which can result in forcing flows overbank into adjacent channels. Conversely, the drainage network can also be modified by channel erosion from overbank flooding.

Fans with deeply entrenched main stem channels and a narrow active channel area in plan form are generally associated, along with other factors, with low sediment supplies from upland drainages. Examples of such fans are the fans of the Tortolita piedmont (Pearthree et al., 1992; Demsey et al., 1993; Vincent et al., 2004) near Tucson, Arizona and the Harquahala piedmont near Phoenix, Arizona (Klawon and Pearthree, 2000; Pearthree et al., 2004). Fans of the opposite end-member type have shallow entrenchment (less than 1 m) at the apex and active channels with rapid lateral expansion with downfan distance. These types of fans are typical of depositional environments with sufficient sediment supplies to maintain the backfilling of channels. In the Little Rainbow Valley near Phoenix, Arizona, both types of fan morphologies occur juxtaposed on either side of a valley (Figure 1). To the northeast of Little Rainbow Valley are the Sierra Estrella Mountains (Figure 1) which exemplify fans with shallow entrenchment and a wide active channel area. The Maricopa Mountains, located to the southwest of Rainbow Valley (Figure 1), exemplify the deeply entrenched and narrow active channel area morphology.

Numerous numerical models of alluvial fans (e.g., Price, 1974; De Chant et al., 1999) have been developed. Many of these simulate the evolution of alluvial fans and morphological feedbacks of both external forcings (changes in base level, changes in
flood magnitude and frequency associated with climate change, and changes in water and sediment inputs to the fan) and autogenic mechanisms (e.g., Coulthard et al., 2002; Nicholas and Quine, 2007). In this alluvial fan model we focus on the relationship between fan morphology and the hydrologic response with consideration of the dynamic nature of the drainage network.

Numerical Model

The numerical model consists of the equations used to define the geometry of a fan and its distributary channel system, routing of surface flow, and infiltration of surface flow. A random network of distributary channels was created using a random walk technique as described below in section 3.1. Flow events were characterized by the duration of the flow event and a time-invariant discharge in the feeder channel applied at the fan apex. Steady-surface flow was routed based on the continuity equation and Manning’s equation (Chow, 1959) using an iterative finite difference approach (e.g., Pelletier et al., 2005) as described in the Flow Routing section. Green-Ampt infiltration was applied as a sink for surface flows as described in the Infiltration section.

The driving input to the model is water depth in the active channel at the fan apex. In the application portion of the paper, we considered a range of flow depths between 0.5 and 1.0 m. Based on streamgage data from piedmonts in Maricopa County (Flood Control District of Maricopa County (FCDMC), 2006) and on paleoflood reconstructions for five canyons supplying flows to the Tortolita fan complex (House, 1991), we noted small variations in flow depth for a two order of magnitude range in drainage area (4.7
km$^2$ to 390 km$^2$) (Figure 2). For each gaging station, the three largest peak flows on record were plotted as was any available historical flood information. The period of record for these stations ranged from 4 to 16 years. The error bars on flood depths determined from paleoflood reconstructions represent variations due to assumptions about subcritical or supercritical flows, and the range in elevations of the slack water deposits or other paleostage indicators (House, 1991) at the study site. Most flows, 28 out of 33, have depths between 0.5 and 1.75 m (Figure 2). This small variation in flow depth is a result of channel adjustment to maintain critical flow which is highly efficient for routing water through channels (Bull, 1979; Grant, 1997). This range of flow depth for Maricopa County is similar to alluvial fan bankfull depths of 0.5 to 1.5 m observed at fan heads in the western U.S. (Stock and Schmidt, 2005).

Alluvial Fan Geometry

The radially symmetric fan has a constant slope along the longitudinal axis (Figure 3) and the cross-fan topographic profile is concave-downwards (Figure 4). In our study, the suite of fan terraces frequently found in the Basin and Range was simplified in the numerical model and only two surfaces were simulated. While most fans have a sequence of several older terraces, we represented this sequence as a single integrated unit in this model. The entrenchment depth diminishes from a maximum at the fan apex to a minimum value of zero at the toe of the fan. The model domain is a square that encompasses the fan radius at the toe of the fan. The topographic elevation at each node of the inactive surface is calculated as:
where \( L \) is the total relief of the fan, \( x_{ij} \) is the distance between the fan apex and location \((i,j)\) in a horizontal plane, and \( R \) is the fan radius.

The most prominent topographic feature is the entrenchment of the active portion of the fan. The alluvial fan consists of an entrenched or active channel area, and an overbank flow area, or inactive area which is the highest topographic surface and represents the oldest geomorphic surface (Figure 3). The term inactive surface is a bit of a misnomer as this surface is inundated in the model and is hence more appropriate thought of as a less-active surface. For the active portion of the fan the maximum incision depth at the fan apex, \( I \), is used to calculate the topographic elevation at each node as:

\[
    z_{ij} = L \left( \frac{R - x_{ij}}{R} \right) 
\]

(1)

In plan view the width of the active channel area increases as a geometric expansion with distance down the fan (Figure 5) as specified by the variable \( c \), an expansion factor. The variable \( c \) is used to specify the widening of the active channel in the downfan direction as:

\[
    w = w_0 \left( 1 + \frac{x_i}{R} \right)^{\frac{\ln(c)}{\ln 2}}
\]

(3)

where \( x_i \) is distance down fan in the \( i \) direction, \( \ln \) is the natural logarithm, \( w_0 \) is the half-width of the active channel at the apex, and \( R \) is the fan radius, and the variable \( w \) is the
half-width of the active channel at a downfan distance of $i$. For a fan with a channel half-width at the apex equal to 2.3% of the fan radius (3.0 km), a $c$ value of 5 results in downfan expansion of the active channel such that the radius of the active channel is 23% of the total fan radius at the toe of the fan, a downfan distance of 100% (Figure 5).

In the numerical model the elevation of the two fan surfaces above the active channel is determined by the fan gradient and the entrenchment depth of the active channel. Consequently, the value of these parameters affects the geomorphic interpretation of these surfaces as does the sediment permeability. Real-world fans have dissected early to mid-Pleistocene surfaces that will not be inundated even during the largest floods except perhaps at the distal portion of the fan where the depth of dissection is small. Mid-Holocene deposits, however, are expected to be inundated on occasion near the proximal portion of the fan and to have undergone some soil development processes. Consequently, the highest topographic surfaces in the model represent modern to mid-Holocene deposits while the active channel deposits are conceived of as modern.

Within the entrenched portion of the fan a network of distributary channels were created using a random walk method (Price, 1974). Random numbers were used to determine the location of channels in the active portion of the fan based on specified probabilities. A set of random walkers (one for every pixel in the input channel cross section) was initiated at the top of the fan. If a grid node is located in the entrenched channel at the top of the fan, and a randomly drawn number is less than 0.5, then a set of $N$ nodes located one node downfan and towards the right bank is specified as a channel if it is also located within the entrenched portion of the fan. If another independent random
number is less than 0.5 and another independent random number is less than \( p \), then a set of \( N \) nodes located one node downfan and towards the left bank is specified as a channel if it is also located in the entrenched portion of the fan. The value of \( N \) increased downfan following Equation 3. In this way, a connected distributed network was created that expands geometrically downfan (Figure 6D). This application of the random walk method resulted in three distinct topographic surfaces which represent the smoother and higher topographic surface of the unentrenched portion of the fan, a lower topographic surface that represents the active portion of the entrenched channel, and a topographic surface consisting of islands of higher topography within the entrenched channel (Figure 6) which have the same height as the unentrenched portion of the fan. Consequently, differences in altitude between the islands and the surrounding terrain diminish downfan such that near the toe of the fan, several shallow distributary channels may merge into a broad sheet of flow.

*Flow Routing*

The continuity equation is the basis for all flow routing models and states that the imbalance in the rate of mass into the system and in the rate of mass leaving the system equals the rate of change in mass in the system. With the assumption of steady flow the rate of change of mass in the system is zero. Furthermore, if the fluid is assumed to be incompressible (of constant density), the continuity equation reduces to

\[
\nabla \cdot q = 0
\]  
(4)
where \( q \) is the vector of discharge per unit. Expressed over a control volume in finite-difference form in two-dimensions, Equation 4 yields:

\[
\frac{Q_{x,i-1,j} - Q_{x,i+1,j} + Q_{x,i,j+1} - Q_{x,i,j-1}}{\Delta x \Delta y} = 0
\]  

(5)

where \( Q_{x,i,j} \) and \( Q_{y,i,j} \) are the volumetric discharge rates in the \( x \) and \( y \) direction and \( \Delta x \) and \( \Delta y \) are the pixel dimensions.

Discharge in the entrenched channel at the fan apex, specified as a flow depth which depends on the fan slope and roughness, was the input forcing to the model. With the exception of this constant-head boundary at the apex of the fan, the model boundary conditions are head-dependent flux (or a Cauchy-type boundary condition) thus the flux at the model boundary is determined by the depth of surface water. A bifurcation method was used to route flow to multiple downslope directions, with a maximum of 8 possible directions (4 cardinal and 4 intercardinal), weighted by bed slope (Freeman, 1991). The routing algorithm was initiated by ranking all grid nodes in the model domain from high to low elevation. Beginning with the highest elevation node, surface flow was distributed to all of the neighboring downslope nodes, weighted by slope. Specifically, the outflow from a cell is shared between all of the adjacent cells of lower elevation. The fraction of flow passed on to the adjacent cell \( i \) is the maximum of 0 and the surface water slope of cell \( i \) divided by the sum over each of the 8 adjacent cells of the maximum of 0 and the slope of each adjacent cell. Next, routing was performed for the second-highest elevation node in the basin, and then proceeding by rank order to the lowest elevation node. This method ensures that incoming flows have been accounted for prior to the distribution of
flows downstream. The bifurcation method with multiple downslope directions affords an advantage for modeling distributed flow across low-relief surfaces (Clevis et al., 2003). Steady-state flow depths were determined by applying the Manning’s equation to the routed discharge at each node. The bed roughness coefficient, or Manning’s $n$, was set to 0.035 (Table 1).

To route flow down the fan, the downstream ranked order must be determined a priori. Because the downstream rank order of the bed topography and the down stream rank order of the water surface profile may differ, the flow routing algorithm was implemented iteratively. For each of $n$ iterations, flow in the feeder channel was routed downfan based with bifurcation routing determined by the rank order of the water or ground-surface elevation. For the first iteration, ranking was based on the ground-surface elevation or bed topography. For each subsequent iteration, a small fraction of the calculated water depth was added to the topography of the bed to create a new topographic surface at each node. This new topographic surface was used to rank and route flow downstream to determine a new surface water profile. This process was repeated until convergence was achieved when the total mass of water on the fan did not increase with additional iterations.

For this paper, the assumption of steady-flow (depth and velocity do not vary with time at a given location) oversimplifies the hydrologic system considering flow events in arid climates are often of short duration and have hydrographs with steeply sloped rising and falling limbs. However, steady-flow was an appropriate level of complexity for examining a variety of synthetic fans geometries with inputs based on arguments about
critical flow depth. In other applications such as an investigation of a particular real-
world fan, steady flow is not likely an appropriate assumption if flood attenuation is
important particularly if hydrograph data are available.

*Infiltration*

At each node surface water was lost to infiltration based primarily on the
sediment permeability, the depth of surface water, and the flow event duration. We used
the explicit Green-Ampt model to simulate infiltration because vertical infiltration
models, and more specifically, infiltration approximated by the Green-Ampt solution
have been successfully fit to measured profiles of infiltration in alluvial deposits
including alluvial deposits with significant lateral flow (Martin et al., 1993). The Green-
Ampt model assumes piston flow through an unsaturated soil driven by the difference in
the water surface potential and a constant wetting front capillary potential; water is
assumed to infiltrate into a homogeneous soil as a saturated rectangular pulse at a
uniform volumetric water content. For the Green-Ampt solution, the infiltration rate
diminishes exponentially with time until approaching the saturated vertical hydraulic
conductivity (Figure 7). The expression for the infiltration rate of the explicit Green
Ampt solution consists of four terms that approximate an infinite series which results in
less than 2% error when compared to the implicit Green-Ampt solution (Salvucci and
Entekhabi, 1994). The four term explicit solution for the infiltration rate is (Salvucci and
Entekhabi, 1994)
where $K_s$ is the saturated hydraulic conductivity, $t$ is time, and $\chi$, which has units of time, is defined as

$$\chi \equiv K_s^{-1} (\theta_s - \theta_i) (-\psi_{cr} + \psi_s)$$

where $\theta_s$ is the saturated volumetric water content (cm$^3$/cm$^3$), $\theta_i$ is the initial volumetric water content (cm$^3$/cm$^3$), $\psi_{cr}$ is the potential at the wetting front (cm), and $\psi_s$ is the potential at the soil surface (cm). The potential at the wetting front is approximated by the bubbling pressure head, the largest negative value of the pressure head for which the soil remains saturated. These soil properties were assigned on the basis of soil texture (Table 2) (Rawls et al., 1992) with the assumption of an initially dry soil.

Within the active portion of the fan, frequent flows deposit sand and gravel and consequently, the active channel was conservatively characterized by sediments of coarse sand. Islands of higher topography within the active channel area are inundated less frequently and some eolian accumulation of finer material may occur resulting in an increased proportion of silt and clay in the surface sediments. The undissected portions of the fan, which represents the oldest geomorphic surface, have the largest proportion of silt and clay. The active channel sediments, portions of the fan consisting of islands in the active channel, and the undissected portion of the fan are characterized by homogenous deposits of coarse sand, loamy sand, and sandy loam respectively.

At each node with a surface flow depth greater than zero, the explicit Green-Ampt solution, a time dependent solution, was applied to the steady-state numerical model by
calculating cumulative infiltration as a function of time. The applied infiltration rate corresponds to the median cumulative infiltration (Figure 7) for each soil texture. Thus, cumulative infiltration determined from the explicit Green-Ampt solution at the duration of the flow event was applied as a constant infiltration rate in the steady-state model. To implement this solution in the numerical model with minimal computation time, second order polynomials were fit to the time-dependent infiltration rate as a function of the surface water flow depth for each soil texture. Infiltration, a sink in the mass balance, was incorporated in the mass balance as:

$$d_o = d_i - \Delta xf / v,$$

where $d_i$ is the depth of surface water into a grid cell, $d_o$ is the depth of water out of the cell, and $v$ is the absolute magnitude of the velocity of the surface flow at the cell node, $\Delta x^2$ is the area of a square grid cell, and $f$ is the infiltration rate from equation 6. Since $d_i$ and $d_o$ were nearly identical, a single value of velocity was used for flow in and out of the grid cell (Figure 8).

Application

As an illustration of the model for a real-world example, we applied the numerical model to Tiger Wash near Phoenix, Arizona using existing geologic maps, streamflow data, a channel cross-section, and precipitation gage data. We compared surface flow depths and infiltration resulting from a low-recurrence interval monsoon event and a low-recurrence interval winter event. Monsoon storms which are primarily generated during
July and August in the Sonoran Desert produce small-scale convective thunderstorms of short duration and high intensity. A second wet season occurs during the winter. During winter larger-scale frontal systems can persist for days creating storm events of much longer duration but typically of lower intensity than monsoon storms. Based on streamflow and precipitation gage data at Tiger Wash (FCDMC, 2006), we simulated storm events on December 29, 2004 and August 25, 2003 (Figure 9). The storm on December 29, 2004 was a typical winter flow event. Several precipitation events of increasing magnitude were recorded at the gage before any streamflow was detected due to the large drainage area of the watershed (221 km$^2$). This storm produced 10 hours of flow at the gage and a peak discharge of 105 m$^3$/s. The monsoon event on August 25, 2003 produced flow at the gage for about 1 hour, peaking at a discharge of 23 m$^3$/s. Both flows are estimated to have recurrence intervals on the order of 2 to 5 years based on the gaged precipitation (FCDMC, 2006) and on the National Oceanic and Atmospheric Administration’s point precipitation frequency estimates (Bonnin et al., 2006).

**Tiger Wash Case-Study**

We applied water volumes in the observed hydrographs as a steady input discharge for the duration of the storm by integrating the hydrograph over the duration of the flow event and calculating time-invariant flow depth inputs (0.38 m and 0.23 m for the winter and monsoon flows, respectively) to the numerical model (Figure 9). Additional input variables were determined from a cross-section at the stream gage (channel entrenchment = 3 m; channel width = 37 m) and from digital terrain model data.
(channel slope = 0.0085) (FCDMC, 2000) with a reported positional accuracy of ±1.5 m. A surficial geologic map of the Tiger Wash distributary system (Klawon and Pearethree, 2000) was used to define the geometry of the active portion of the fan for the numerical model. Late Holocene active channel, sheetflood, and overbank deposits (Qy) and early to late Holocene relict alluvial fan and terrace deposits (Qy1) were classified as active channel, whereas, the older mapped units (late Pleistocene alluvium (Ql), middle Pleistocene alluvial fan deposits (Qm), and early to middle Pleistocene alluvial fan deposits) were regarded as inactive surfaces in the numerical model (Figure 10). The lateral expansion of the active portion of the fan was determined by measuring the width of the units classified as active surfaces on the geologic map as a function of distance from the fan apex (Figure 10). In the numerical model the fan radius was set to 7,900 m, the maximum distance for which the active channel area could be delineated from the map with confidence. Based on this characterization, the active channel deposits comprised 19% of the fan area. The hydraulic parameters required for Green-Ampt infiltration were assigned to the three surfaces (active channel, islands in the active channel, and inactive surfaces) on the basis of soil texture (Table 2). These surfaces correspond to coarse sand, loamy sand, and sandy loam with saturated hydraulic conductivities of 20.0 cm/hr, 5.98 cm/hr, and 2.18 cm/hr, respectively. These are the base case saturated hydraulic conductivity values (Table 1) used throughout the modeling experiments unless otherwise characterized.

The winter flow event persisted long enough to approach the saturated hydraulic conductivity which resulted in significantly lower infiltration rates relative to the
monsoon event. For example, for a flow depth of 0.3 m, the infiltration rate for the winter event was 61% that of the monsoon event (0.2357 m/hr and 0.3851 m/hr, respectively) for the sandy active channel sediments. Based on the output of the numerical model, both the winter and summer flows were confined to the active portion of the fan (Figure 11). Surface flow depths were at a maximum near the fan apex and rapidly decreased with distance down fan. In contrast, infiltration was less variable across the fan.

It was difficult to separate the effects of storm duration from discharge magnitude on the infiltration responses because the winter and monsoon discharges were not equal. To this end, flow duration was varied with a constant input discharge (flow depth = 0.5 m) and the infiltration responses were examined. Because the event-average infiltration rate applied in the numerical model is much lower when events last longer, the infiltration rate of the active channel sediments decreased rapidly with increasing event duration. For example, for a ponded water depth of 0.5 m applied to sandy active channel sediments, infiltration rates decreased from 43.2 cm/hr for a flow event 1 hour in duration to 22.4 cm/hr for a flow event 24 hours in duration (Figure 12). For these conditions, infiltration rates did not decrease much for flow durations greater than 6 hours due to the dominance of the large, early time infiltration rates (Figure 12). The infiltration rate had a strong effect on the proportion of the active channel area inundated by surface flows. As the applied infiltration rates decreased the area of inundation of the active channel sediments increased because modeled infiltration was a sink for surface flows. For example, the inundation area approximately doubled, from 4% to 8% of the active channel area, as the infiltration rate varied from a maximum of 43.2 cm/hr for a 1 hour
flow event to a minimum of 22.4 cm/hr for a 24 hour flow event (Figure 12).

The applied infiltration rate and the area of inundation had a complex effect on the fan-wide infiltration volume (infiltration accumulated during the flow event in each grid cell times the area of the grid cell summed over all grid cells (m$^3$)). To compare fan-wide infiltration volumes for events of varying duration, the fan-wide volumetric infiltration rate was also examined as a function of the event duration. For 1 to 4 hour events, an increase in the event duration resulted in a slight decrease in the fan-wide volumetric infiltration rate (m$^3$/s) which reflected the large decrease in the applied infiltration rate which dominated the infiltration responses (Figure 12). However, for events 8 to 24 hours in duration, the fan-wide volumetric infiltration rate (m$^3$/s) increased reflecting the dominance of the inundation area. The 24-hour flow duration resulted in the largest fan-wide volumetric infiltration rate, however, this value was only 10% larger than that of the 1 hour flow event. The decrease in fan-wide volumetric infiltration rate which occurs at the 8 hour duration reflects a change in the balance of the impact of the applied infiltration rate and the inundation area.

**Numerical Experiments**

A series of numerical experiments was carried out on synthetic fans to analyze the relationships between fan morphology (fan area, depth of active channel entrenchment, active channel area, sediment permeability, and fan slope) and the associated runoff and infiltration. The numerical experiments were performed in a Monte Carlo framework and averaged over different realizations of channel topography created by the stochastic
distributary channel network. Output metrics for ten independent realizations of the random-walk process were averaged in order to isolate the controlling effects of the deterministic fan geometry. For each realization a constant input flow depth of 1 m was distributed across the fan during a 1 hour flow event with the assumption of an initially fully drained soil profile.

Key model outputs of greatest interest to fluvial geomorphologists and hydrologists are how much infiltration occurs and where it occurs. To determine how much infiltration occurs we employed an infiltration to runoff ratio metric. The infiltration to runoff ratio was calculated as the fan-wide volumetric infiltration rate (m$^3$/s) divided by the sum of the fan-wide volumetric infiltration rate and discharge off the fan at the model boundaries or \( \frac{(\Delta x^2) f}{(Q_e + (\Delta x^2) f)} \). The infiltration to runoff ratio is a unitless metric that scales between 0 and 1. To determine where on the fan infiltration occurs, we used two additional metrics: the ratio of infiltration on unincised surfaces to infiltration on active surfaces, \( \frac{I_i}{I_a} \), and the longitudinal distance to the center of mass of the infiltration as a percent of the distance down fan, \( x_m' \), which was calculated as:

\[
x_m' = \left( \frac{100 \sum m_i x_i}{R \sum m_i} \right),
\]

where \( m_i \) is the mass of infiltrated water in grid cell \( i \) and \( x_i \) is the longitudinal distance to the node of grid cell \( i \). For example, if the mass of infiltrated water was uniformly distributed along the longitudinal axis \( x_m' \) would be equal to 50 %. And as the center of mass of the infiltrating water shifts towards the apex \( x_m' \) approaches 100 %.
An open question in fluvial geomorphology is how fan geometry affects the partitioning of flow between on-fan infiltration and surface flow leaving the fan system. Flows leaving the fan system may eventually reach the valley floor depending on evapotranspiration rates and transmission losses. With sufficient connection to the groundwater system these flows can become an axial river which either exits the watershed or drains internally to a playa.

We investigated the ratio of fan-wide volumetric infiltration rate to input discharge to determine what fraction of the input flow infiltrated on the fan and what fraction of the flow exited the fan system (Figure 13). In the first experiment we examined the partitioning of flow as a function of fan size and fan slope. We independently varied fan area and slope to consider a full spectrum of fan morphologies; however, correlations have been noted, based primarily on fans of the southwestern U.S., between fan size and drainage area (Denny, 1965; Blair and McPherson, 1994), and between fan slope and drainage area (Denny, 1965; Harvey, 1987). Fan area was varied between 1 km\(^2\) and 100 km\(^2\) and, simultaneously, fan slope was varied between 0.0175 and 0.0875 at the base case parameter values (Table 1) with the exception of the active channel expansion factor, \(c\), which was varied with fan area to maintain a constant active channel proportion (25 %) of the total fan area.

Ponded surface flow depths were greater on the more gently sloped fans due to reduced surface water velocities. Consequently, the greatest amount of infiltration occurred on fans with shallower slopes where water also tended to spread out laterally (Figure 13). In contrast, shallower, more laterally confined flow occurred on the steeper
fans. Most of the water ran off the smaller fans, and at a minimum, 4% of the input flow infiltrated on the fan. For the largest and lowest gradient fans, no water ran off and, at a maximum, 100% of the flow infiltrated on the fan. As fan area increased there was a corresponding linear response in the proportion of infiltrated water for fan areas up to 20 km$^2$; this response was largely independent of slope. About 20% of the incoming flow exited to the valley floor at the threshold of 20 km$^2$. Fan slope was a minor influence on the partitioning of flow between the fan and the valley floor for fans greater than 30 km$^2$. For fans greater than 30 km$^2$, the variability in the portioning of flow between the fan and the valley floor was small (0% to 30% of the incoming flow exited the fan system). Infiltration volume per input discharge depended only on channel slope and was independent of fan area for the largest fans (greater than 50 km$^2$).

The deepest surface flows occurred on fans with the shallowest slopes as indicated by the center of mass of the surface flow ($x_{m}^s$) which shifted towards the fan apex with decreasing fan slope (Figure 13). In general, the center of mass of the surface flow was affected more by fan area than by channel slope. For the largest fans $x_{m}^s$ approached 10% of the fan radius. As fan area approached zero, $x_{m}^s$ approached 45% of the fan radius and was independent of fan slope.

The magnitude and frequency of flooding on alluvial fans as a percentage of the fan surface is controlled by the ratio of the flow depth to the entrenchment depth at the fan apex. This ratio largely determines whether flows are confined within the active channel, or whether flows fill the active channel and also inundate older surfaces. In this case, the surface water flow regime consists of deeper flow in the active channel and
shallow sheetflow on the unincised surfaces.

In the second experiment, we examined the influence of the ratio of the input flow depth to the entrenchment depth on infiltration responses as a function of the active channel area. To examine the threshold between these two flow regimes we considered the morphologic end members: fans with narrow, deeply-entrenched channels and fans with shallowly-entrenched active channel areas that rapidly widens with distance down fan. In the second experiment, we simultaneously varied the depth of active channel entrenchment (from 0 to 4 m) at the apex and the active channel area (from 15% to 55% of the total fan area) for an input flow depth of 1.25 m. The remaining parameters used to specify the fan geometry were set at the base case parameter values (Table 1).

Surface flow was radial and unconfined for fans with small incision depths. In addition, overbank flooding occurred near the apex and a maximum of 94% of the unincised surface area was inundated by surface flow (Figure 14). As the active channel area increased, the area of inundation slightly decreased indicating that the point of outburst onto the unincised surface moved towards the toe of the fan. Flow was contained within the entrenched portion of the fan and almost no inundation of the unincised surface occurred on the most deeply entrenched fans.

The active channel proportion of the fan strongly influenced the infiltration regime. In the second experiment, the amount of infiltration was as quantified as an infiltration to runoff ratio. The largest infiltration to runoff ratio occurred on fans with incision depths slightly smaller than the input flow depth, or at an incision depth of 1 m (Figure 14); a maximum of 93% of outflow exited the fan system as infiltration. Flow
was largely confined to the active channel for incision depths much greater than the input flow depth and, consequently, almost no infiltration occurred on the unincised surfaces. The least amount of infiltration, only 40% of the outflow exited the system as infiltration, occurred on fans with the largest incision depth and the smallest active channel area. For incision depths much smaller than the input flow depth, large portions of the unincised surface were inundated with sheetflow resulting in infiltration on the unincised surfaces; however, shallower incision depths resulted in shallower ponded water depths in the active channel area. Because the active channel had the highest permeability sediments, this resulted in reduced infiltration volume. Overall, the results of the second experiment showed that infiltration volume was primarily dependent on the fan area inundated with surface flow and, secondarily, on the depth of surface flow. It is important to consider that for fans with a deep fanhead trench it may not be appropriate to compute area inundation metrics based on total fan area but rather only on the area distal from the intersection point. For these deeply dissected fans, even the largest floods will not inundate the inactive surface and thus referencing the total fan area is ambiguous when comparing inundation areas among different fan morphologies.

Although the geomorphic factors considered thus far (fan area, fan slope, active area proportion of fan, incision depth) significantly affect infiltration responses, the hydraulic conductivity is the primary parameter used to determine the infiltration rate. In the third and final experiment we considered the impact of sediment permeability on distributed infiltration responses. As discussed in the conceptual model section, sediment permeability varies between basins due to differences in geology, depositional
environment, slope, history of the sediment to water flux, tectonics, and soil development. The permeability of the sediments, and the permeability contrast between different geomorphic surfaces, has a primary impact on the infiltration rate and a secondary influence on the surface water flow regime by influencing the area of inundation and the depth of surface flows. We varied the permeability contrast between two surfaces and examined where on the fan infiltration occurred. In particular, we investigated permeability contrasts between the unincised and active channel deposits which can be attributed to surface age (Young et al., 2004) or depositional environment. For example, older surfaces on fans adjacent to playas may accumulate eolian material resulting in reduced permeability given sufficient time and prevailing winds from the playa towards the fan. For fans adjacent to playas, the older surfaces (the undissected surface and islands in the active channel) can be characterized by a texture of clay loam (Young et al., 2004). In the active portion of the fan, frequent flow events transport fine particles as suspended load and no significant eolian accumulation of fine particles would occur. Consequently, the active channel sediments were assigned hydraulic properties based on a texture of coarse sand. In the third experiment, the input discharge divided by drainage area (from 2 to 16 (m$^3$/s)/km$^2$) and the permeability contrast between the sediments of the active channel and the older surfaces were simultaneously varied for a 1 hour flow event. The permeability contrast, $\Delta K$, was calculated as:

$$\Delta K = K_i / K_a,$$  

(10)
where $K_i$ and $K_a$ are the hydraulic conductivities of the unincised channel sediments and the active channel sediments, respectively. The permeability contrast was varied by more than 2 orders of magnitude, from 0.0095 to 0.14, by changing only the permeability of the unincised sediments ($K_i$) when $K_a$ was set equal to 20.0 cm/hr. Values of $\Delta K$ of about 0.01 are representative of fans with an influx of eolian sediment (Young et al., 2004). The remaining parameters used to specify fan geometry were set at the base parameter values (Table 1) with the exception of the active channel expansion factor, $c$, which was set to 15, or equivalently, the active channel area comprised 27% of the total fan area.

We also employed the metric $I_i/I_a$ (the ratio of infiltration on unincised surfaces to infiltration on active surfaces) to determine where on the fan infiltration occurred as the input discharge and the permeability contrast were varied. In general, most of the infiltration occurred in the active channel. More infiltration occurred on older surfaces due to increased overbank flow as the input discharge was increased. For fans with a significant influx of eolian material, the metric $I_i/I_a$, ranged from 0.1 to 0.2 m$^3$/m$^3$ indicating that, at most, infiltration on the unincised surfaces was one-fifth that of the active surfaces. In contrast, for fans without a significant influx of eolian material, the metric $I_i/I_a$ ranged from 0.2 to 0.65 m$^3$/m$^3$ indicating that, at most, infiltration on the unincised surfaces was one-half that of the active surfaces although the unincised surfaces represented 73% of the total fan area. The partitioning of infiltration between the unincised and active surfaces became more sensitive to the value of the permeability contrast than to the input discharge for the unincised surfaces with low permeabilities.
($\Delta K = 0.03$). This behavior represents a threshold value of hydraulic conductivity at which little infiltration occurs on the highest topographic surfaces.

Because most of the infiltration occurred in the active channel, altering the permeability contrast of the sediments had no influence on where infiltration occurred on the fan as quantified by the center of mass of the infiltration ($x_m'$. The center of mass of the infiltration increased in a nearly linear manner as the input discharge per drainage area increased from 0 to 4 $(m^3/s)/km^2$. The center of mass of the infiltration approached a maximum value of 51% as the input discharge increased and was essentially constant varying only between 49.5 and 51% for values of input discharge per drainage area between 4 and 16 $(m^3/s)/km^2$ (Figure 15). The center of mass of the infiltrated water was much greater and less variable than the center of mass of the surface water which varied between 26.5 and 36.5% (not shown) due to much greater water depths near the fan apex than at the toe of the fan for the range of parameter values considered.

Conclusions

We simulated a variety of fan morphologies and quantified the amount and location of infiltration on fans using numerical experiments of synthetic alluvial fans. Based on the model results most of the infiltration occurred in the active channel which had the deepest surface flow and highly permeable sediments. We also found that the amount of infiltration on fans was most sensitive to the geomorphic factors that influenced the area of inundation rather than the depth of surface flow. The greatest amount of infiltration occurred on fans with low gradients which resulted in greater
surface flow depths and more lateral spreading of flow across the fan. The dominance of
the area of inundation over the depth of ponded surface water was also illustrated with
the Tiger Wash application in which the fan-wide volumetric infiltration rate was shown
to be 6% larger for the 24 hour flow event than for the 1 hour flow event due to an
increased area of inundation although the infiltration rate of the 24 hour flow event was
considerable smaller than the infiltration rate of the 1 hour flow event. The magnitude
and frequency of flooding on the inactive fan surfaces is largely controlled by the ratio of
the flow depth to the entrenchment depth of the active channel at the fan apex. This ratio
partitions flow between the active channel and the unincised surfaces. As such, this ratio
largely determines whether the surface flow regime consists of deeper flow confined
within the active channel, or whether flow also inundates unincised surfaces with shallow
sheetflow. The greatest amount of infiltration occurred on fans with incision depths
slightly smaller than the input flow depth. Investigation of the ratio of the flow depth to
the entrenchment depth at the fan apex also illustrated that the fan-wide volumetric
infiltration rate ($L^3/T$) was found to be primarily dependent on the fan area inundated
with surface flow and, secondarily, on the depth of surface flow.

With regard to where infiltration occurs on fans, significant differences were
found between the mass distribution of surface flows and the mass distribution of
infiltrated water. The location of the center of mass of the surface flows in the down fan
direction was closer to the fan apex and also more variable than the location of the center
of mass of the infiltrated water. The water depth, generally deeper at the fan than the toe,
has a stronger influence on the center of mass of the surface flow than on the center of
mass of the infiltrated water which tends to be controlled by highly permeable sediments in the active channel rather than the depth of ponded surface water. Although the geomorphic factors considered (fan area, fan slope, active area proportion of fan area, incision depth) significantly affect infiltration responses, the hydraulic conductivity is the primary parameter used to determine the infiltration rate. We considered the impact of contrasts in permeability between the active channel and the unincised surfaces on how much infiltration occurred on the unincised surfaces. The permeability contrast between different geomorphic surfaces has a primary impact on the infiltration rate and a secondary influence on the surface water flow regime by influencing the inundation area and the depth of surface flows. For fans with a one order of magnitude permeability contrast, infiltration on the unincised surfaces was, at most, one-half that of the active surfaces although the unincised surfaces comprised 73% of the fan area. In contrast, for fans with a two order of magnitude contrast between the permeability of active and unincised sediments, infiltration on the unincised surfaces was, at most, one-fifth that of the active surfaces.

We have shown that fan morphology exerts a first-order control on the recharge potential. Therefore, at the basin scale, this method could be used to compare the recharge potential of different arid and semi-arid watersheds using limited hydrologic information. Frequently, when an estimate of recharge potential is needed at the basin scale, simple precipitation-recharge or precipitation-elevation-recharge empirical relations such as the Maxey-Eakin method (Maxey and Eakin, 1949; Eakin, 1960) are employed. Using a modification of this fan model, discharge-drainage area and
discharge-frequency relations could be used to estimate recharge-runoff ratios for a collection of alluvial fans and their adjacent drainage basins. Given detailed recharge measurements for one basin, the model results could be used to scale the recharge estimates of an instrumented fan to adjacent fans based on fan morphology.

We also illustrated that fan morphology strongly influenced the area of flow inundation which has implications for flood hazard assessment. If the model was applied to a particular fan, the stochastic distributary channel network would likely be replaced with real topography. In this case, if geomorphic site information was available the hydrologic-geomorphic portion of the model could be applied to a particular fan with topographic information to predict where flows are likely to occur and how deep infiltration is likely to be. These model predictions could be used to guide monitoring efforts to determine the appropriate distance down fan and the depth at which to place monitoring instruments. Careful selection of monitoring sites is essential when flows are infrequent and of short-duration and both monitoring budgets and monitoring periods are typically limited. Later on, these data typically would be used to calibrate a flow model that incorporated the dynamic momentum equation. However, the modified version of the fan model could be used to explore the feasible parameter space efficiently before moving to the more computationally intensive dynamic momentum equation approach common for flood hazard assessment.

Additionally, at the scale of a single fan, the hydrologic-geomorphic portion of the model could be used to spatially interpolate existing hydrologic data. For example, if limited infiltration point-data were collected on a fan, this model could be used to
develop a first-order approximation of fan-wide infiltration responses based on fan morphology. Similarly, if flow was gaged or estimated from high-water marks at a single cross-section, this model could be used to estimate the fan-wide hydrologic response. Moreover, the model could be used to develop scaling relationships between flow magnitude and inundation area for different magnitude storm events.

We have shown that the surface permeability of the fan sediments has a primary influence on the hydrologic response. This underscores the importance of understanding the history of sediment deposition on alluvial fans particularly how climate variation has affected the prograding and retrograding of fan sediments over time. These processes produce spatial variability in sediment permeability in the vertical direction and a temporal variability in recharge rates that are important to characterize for assessing water supplies and for predicting the migration of chemical constituents.

Acknowledgments

Adrian Harvey, Tom Coulthard, and Gregory Tucker provided comments and suggestions that significantly improved the manuscript. We gratefully acknowledge support from the University of Arizona, Technology and Research Initiative Fund, Water Sustainability Competitive Grants Program.
Notation

c, active channel expansion factor
d₀, depth of surface water in to a grid cell, m³/s
dᵢ, depth of surface water out of a grid cell, m³/s
f, infiltration rate, m/s
q, discharge per unit width, m²/s
I, incision depth at the fan apex, m
Iₐ, total infiltration on active surface, m³
Iᵢ, total infiltration on unincised surfaces, m³
Kₐ, saturated vertical hydraulic conductivity of the active surface sediments, m/s
Kᵢ, saturated vertical hydraulic conductivity of the unincised surface sediments, m/s
Kₛ, saturated hydraulic conductivity, m/hr
L, relief of fan, m
mᵢ, mass of water in grid cell i, kg/m³
n, Manning’s channel bed roughness coefficient
N, number of iterations
p, random walk probability parameter
Qₑ, total discharge off the fan at the model boundaries, m³/s
Qᵢ, net volumetric flow rate in to a grid cell, m³/s
Qₒ, net volumetric flow rate out of a grid cell, m³/s
R, radius of the fan, m
v, velocity of surface flow, m/s
w, channel half-width, m
w₀, channel half-width at the fan apex, m
xᵢⱼ, horizontal distance from the fan apex to grid cell i,j, m
xᵢᵢ, center of mass of the infiltrated water along the longitudinal axis of the fan, percent
xᵢᵢ, center of mass of the surface flow along the longitudinal axis of the fan, percent
zᵢⱼ, relative topographic elevation at grid cell i,j, m
Δx, grid cell length, m
θᵢ, initial water content, m³/m³
θₛ, saturated water content, m³/m³
ρ, density of water, kg/m³
ψₑᵣ, potential at the soil surface, m
ψₛ, potential at the wetting front, m
References


Weissmann, G.S., J.F. Mount, and G.E. Fogg. 2002a. Glacially driven cycles in accumulation space and sequence stratigraphy of a stream-dominated alluvial fan,


<table>
<thead>
<tr>
<th>Parameter description (units)</th>
<th>Variable</th>
<th>Range of values</th>
<th>Base case values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model inputs</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Flow depth in feeder channel (m)</td>
<td>-</td>
<td>0.5 – 1.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Flow event duration (hr)</td>
<td>-</td>
<td>1 – 10</td>
<td>1</td>
</tr>
<tr>
<td><strong>Fan geometry</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active channel expansion factor</td>
<td>c</td>
<td>7 – 65</td>
<td>15</td>
</tr>
<tr>
<td>Channel half-width at apex of fan (m)</td>
<td>$w_0$</td>
<td>20.0 – 70.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Incision depth at apex of fan (m)</td>
<td>I</td>
<td>1.0 – 4.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Radius of the fan (m)</td>
<td>$R$</td>
<td>1,000 – 10,000</td>
<td>3,000</td>
</tr>
<tr>
<td>Random walk probability parameter</td>
<td>$p$</td>
<td>0.30-0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Relief of fan</td>
<td>$L$</td>
<td>1° – 5°</td>
<td>2.3°</td>
</tr>
<tr>
<td><strong>Flow routing and infiltration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grid cell length (m)</td>
<td>$\Delta x$</td>
<td>5.0 – 10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Manning’s $n$ bed roughness coefficient</td>
<td>-</td>
<td>0.035</td>
<td>0.035</td>
</tr>
<tr>
<td>Saturated hydraulic conductivity</td>
<td>$K_s$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(cm/hr)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive surface (texture)</td>
<td></td>
<td>0.1852 – 2.5000</td>
<td>2.180 (sandy loam)</td>
</tr>
<tr>
<td>Active channel islands (texture)</td>
<td></td>
<td>0.2000 – 5.9800</td>
<td>5.980 (loamy sand)</td>
</tr>
<tr>
<td>Active channel (texture)</td>
<td></td>
<td>9.0910 – 58.1820</td>
<td>20.000 (coarse sand)</td>
</tr>
<tr>
<td>Initial water content ($m^3/m^3$)</td>
<td>$\theta_i$</td>
<td>0.020 – 0.075</td>
<td>0.020 – 0.041$^a$</td>
</tr>
<tr>
<td>Saturated water content ($m^3/m^3$)</td>
<td>$\theta_s$</td>
<td>0.390 – 0.501</td>
<td>0.401 – 0.417$^b$</td>
</tr>
<tr>
<td>Bubbling pressure (cm)</td>
<td></td>
<td>7.26 – 32.56</td>
<td>7.26 – 14.66$^c$</td>
</tr>
<tr>
<td>Number of iterations</td>
<td>$N$</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>

$^a$ 0.041 (sandy loam), 0.035 (loamy sand), 0.02 (sand).
$^b$ 0.412 (sandy loam), 0.401 (loamy sand), 0.417 (sand).
$^c$ 14.66 cm (sandy loam), 8.69 cm (loamy sand), 7.26 cm (sand).

Table 1. Model inputs and parameters for alluvial fan geometry, flow routing, and infiltration.
Soil texture & $K_s$ (cm/hr) & $\theta_i$ (cm$^3$/cm$^3$) & $\theta_s$ (cm$^3$/cm$^3$) & Bubbling pressure (cm) \\
--- & --- & --- & --- & --- \\
Sand & 20.00 & 0.020 & 0.417 & -7.26 \\
Loamy sand & 5.98 & 0.035 & 0.401 & -8.69 \\
Sandy loam & 2.18 & 0.041 & 0.412 & -14.66 \\
Loam & 1.32 & 0.027 & 0.463 & -11.15 \\
Silt loam & 0.68 & 0.015 & 0.501 & -20.76 \\
Clay loam & 0.23 & 0.075 & 0.390 & -25.89 \\
Silty clay loam & 0.15 & 0.040 & 0.471 & -32.56 \\

Table 2. Hydraulic properties (saturated hydraulic conductivity, initial water content, saturated water content, and bubbling pressure) for the explicit Green-Ampt solution were assigned on the basis of the geometric mean soil texture (Rawls et al., 1992).
Figure 1. A. Image of Little Rainbow Valley which is located to the southwest of Phoenix, Arizona. Waterman Wash runs through the center of Little Rainbow Valley (3x vertical exaggeration). Image from NASA World Wind, LandSat7. B. Image of the Maricopa Mountains from Google Earth (2.75x vertical exaggeration). C. Image of the Sierra Estella Mountains from Google Earth (2.75x vertical exaggeration).
Figure 2. Drainage area versus flow depth for 7 gaging stations on piedmonts in Maricopa County (FCDMC, 2006) and paleoflood reconstructions for 5 canyons supplying flows to the Tortolita fan complex. For each gaging station, the three largest peak discharge events during the period of record plus any available historical flood events were plotted. For the paleoflood data, the error bars on the flow depth represent a range in the discharge estimates.

Figure 3. Schematic of alluvial fan model with fan geometry input variables identified ($I$ is the incision depth at the fan apex, $L$ is the relief of the fan, $R$ is the radius of the fan, $x_{ij}$ is the horizontal distance to grid cell $ij$, and $w_0$ is the channel half-width at the fan apex.)
Figure 4. Fan geometry for the case-study of Tiger Wash, Maricopa County, Arizona. A. Longitudinal profile down the fan (vertical exaggeration 8x). B. Radial cross-section mid-distance down the fan (vertical exaggeration 1600x).
Figure 5. Plan view of the widening of the active channel with distance along the longitudinal axis as specified by the variable $c$ for a feeder channel half width that is 2.3% of the radial distance of the total fan radius. $c$ values of 7, 20, and 40 correspond to configurations where the active channel area is 15%, 34%, and 49% of the total fan area ($9 \text{ km}^2$).
Figure 6. Examples of $p$ values used to define topographic surfaces in plan view. Black areas correspond to the oldest unincised surface, white areas correspond to topographic islands in the active channel, and gray areas correspond to the lowest topographic surface in the active channel for a single random realization of topography. A. $p = 0.15$ B. $p = 0.5$. C. Surface flow depth for a flow of a 1 hour duration and an input flow depth of 1.5 m ($p = 0.3$). D. Inset of flow depth near the fan apex.
Figure 7. The infiltration rate used in the alluvial fan model that corresponds to the median cumulative infiltration from the explicit Green-Ampt solution for a flow event 8 hours in duration for a sandy soil with a saturated hydraulic conductivity, $K_s$, of 14.6 cm/hr and 100 cm of ponded water. A. Infiltration rate. B. Cumulative infiltration.
Figure 8. Schematic of a grid cell in the model domain. \( d_i \) is the depth of surface water entering the cell, \( d_o \) is the depth of surface water exiting the cell, and \( (\Delta x)f/v \) is the infiltration loss from the cell.
Figure 10. A. Surficial geologic map of Tiger Wash (modified from Klawon and Pearthree, 2000) (Qy – Late Holocene active channel, sheetflood, and overbank deposits; Qy1 – Early to late Holocene relict alluvial fan and terrace deposits; Ql – late Pleistocene alluvium; Qm - Middle Pleistocene alluvial fan deposits). The black line denotes the active channel area used in the numerical model. The hatched area denotes inactive channel sediments within the active channel area used in the numerical model. B. Lateral expansion of the active channel area with distance from the fan apex along transect a-a'. The width of the active channel is plotted as a half-width for direction comparison with the parameter \( w_0 \) used in the numerical model.
Figure 11. Surface flow depth (m) (top row) and infiltrated water (m) (bottom row) accumulated over the duration of the storm for a 1 hour monsoon flow event (left column) and a 10 hour winter flow event (right column).
Figure 12. A. active channel area inundated with surface flow (%) and the infiltration rate into sandy active channel sediments for a ponded water depth of 0.5 m. B. fan-wide infiltration volume (m$^3$) and fan-wide volumetric infiltration rate (m$^3$/s) as a function of event duration.
Figure 13. Channel slope versus fan area with an input flow depth of 1.0 m and an incision depth of 2.0 m. A. Fan-wide volumetric infiltration rate divided by the input discharge. B. Longitudinal distance of the center of mass of the surface flow as a percent of the distance downfan, $x_{m}^{s}$.

Figure 14. Ratio of input flow depth to incision depth versus active channel area (%) with an input flow depth of 1.25 m for A. Fan-wide volumetric infiltration rate (m$^3$/s) on the fan divided by the sum of the fan-wide volumetric infiltration rate and discharge off the fan (m$^3$/s), $\frac{\Delta v^2 f}{Q_e + (\Delta v^2 f)}$, and B. Fraction of the older surface area inundated with flow (%).
Figure 15. Discharge per drainage area versus the permeability contrast between active channel and unincised surface deposits for a fan where the active channel area is 27% of the total fan area for A. Ratio of infiltration on unincised surfaces to infiltration on active channels, $I_i/I_a$, and B. Longitudinal distance of the center of mass of the infiltrated water as a percent of the distance downfan, $x_{mi}$. 
APPENDIX C. DESIGNING EFFICIENT HYDROLOGIC MONITORING NETWORKS USING COST-BENEFIT ANALYSIS

Joan B. Blainey and Ty P.A. Ferré

Abstract

Designing monitoring networks that can discriminate among competing conceptual models is a key challenge for hydrologists. We investigated this issue by examining the impact of network design on the utility of measurements for constraining hydrologic prediction uncertainty. The goal of the investigation was to determine whether an effective monitoring network could be designed objectively based on limited initial data by comparing projected measurement costs with the expected benefits of improved hydrologic assessment. The measurement network design approach we present is most appropriate for hydrologic assessments for which it is impractical to modify the design of the measurement network repeatedly during monitoring. To evaluate this approach, we examine a hypothetical feasibility assessment for a proposed artificial recharge site. Specifically, predefined sets of candidate measurements (temporal gravity change) at a single location at a series of measurement times are considered. We use an ensemble approach to assess the likely impact of measurement error on prediction error and uncertainty for different combinations of candidate measurements (measurement sets). The ensemble of prediction errors is translated to a probability-weighted performance cost for each measurement set using a cost function. The total cost is calculated as the
sum of the performance and measurement costs. We show that the optimal measurement set, defined as set with the lowest total cost, depends on the prediction of interest, the per measurement cost, the maximum risk-based cost associated with the hydrologic prediction, and the treatment of uncertainty in defining performance costs.

Introduction

We investigate the issue of monitoring network design from the following perspective. The immediate use of a hydrologic model is to produce acceptably accurate and certain predictions of interest. These predictions may range from highly specific and applied questions (e.g., quantifying water or solute mass flux at a location through time) to more general, conceptual questions (e.g., testing multiple hypotheses describing a hydrologic system). Both the accuracy and the certainty of the predictions of interest impact the value of the hydrologic analysis. The accuracy and certainty of hydrologic predictions are affected by all sources of error and uncertainty, often described as model error and measurement error. The purpose of hydrologic data collection is to improve our ability to discriminate among and to constrain models to improve our hydrologic predictions of interest. The objective of monitoring network design is to quantify the trade-off between the expected performance and projected cost of competing observation sets so that an optimal network can be chosen.

Prediction uncertainty is a result of complex interactions among unknown model errors and measurement errors. The precise distribution and magnitude of measurement errors can be difficult to predict. The standard statistical approach is to assume that errors
are normally distributed with a zero mean and a variance that is defined by the performance of the measurement device in the environment of interest (e.g., Tobias and Trindade, 1995). While there is little data to support these assumptions in the field, especially for large-scale, indirect measurement methods, this assumption is made commonly in many scientific and engineering disciplines (e.g., Ding, 1999; Vrugt et al., 2005; Kumar and Patel, 2007; Zohar and Geiger, 2007). In addition to these measurement errors, there are less obvious interpretation errors that affect the value of hydrologic measurements, especially larger-scale, indirect measurements. These errors are due in part to the nonunique relationship between the instrument response and the spatial distribution of the property of interest within the sample volume of the instrument (e.g., Ferre et al., 1998; Margulis et al., 2002; Singha and Moysey, 2006). As a result, the error and uncertainty associated with the interpretation of an instrument response depends on one’s underlying conceptualization of the distribution of the property of interest and on the effect of that distribution on the instrument response. These interpretation errors are rarely quantified but there is no reason to expect that they are normally distributed or zero mean.

Model errors and measurements errors both affect model performance via uncertainty in model predictions. Like measurement errors, some aspects of model performance can be described reasonably well, such as solution accuracy and, to some degree, parameterization effects (e.g., Tiedeman et al., 2003, 2004). Often, these effects are quantified by calculating the expected reduction in uncertainty of predictions of interest with improved parameter estimation. This expected improvement can be assessed
through a sensitivity analysis of a single calibrated model. This approach is efficient because it reduces the number of model calibrations. But, it is limited in its ability to assess the combined effects of unknown model and measurement errors (Moore and Doherty, 2005). These unknown model errors have been referred to as model structural errors or conceptual model errors (Sorooshian and Dracup, 1980; Carrera and Neuman, 1986; Neuman, 2003; Huard and Mailhot, 2006). Like the interpretation errors described above, these errors are more difficult to define because they stem directly from one’s conceptualization of the hydrologic system. Specifically, conceptual model errors reflect the inappropriateness and incompleteness of the model’s representation of the dominant physical processes. This more subtle model error can result in large uncertainties if an incomplete conceptual model underlies the model structure or if the model does not fully capture or resolve the spatial and temporal variability of the system (Neuman and Wierenga, 2003; Renard, 2007). Several recent efforts have been developed to consider these errors and their effects on predictions more explicitly (Binley and Beven, 2003; Neuman, 2003; Neuman and Wierenga, 2003; Vrugt et al., 2005; Poeter and Hill, 2007).

Some hydrologic analyses do not warrant extensive design efforts because the system is relatively simple or because the available budget is too small to allow for many monitoring options. Other investigations are not amenable to measurement network design because they are purely exploratory; measurements are collected to observe the system with no intention of answering specific questions. We focus on designing data collection schemes for more challenging and/or expensive hydrologic studies that aim to address specific questions and/or testable hypotheses. Designing networks for these
studies is a complex problem due to temporal and spatial variation in hydrologic events, differences in scale between hydrologic processes and measurements, and the heterogeneity of the subsurface. These challenges are amplified by the low sampling density that is common in hydrologic investigations due to the limits of the measurement methods and budgetary restrictions. In practice, hydrologists usually need to trade-off the accuracy and uncertainty of point measurements with the spatial coverage obtained from larger-scale measurements. Point scale measurements are more likely to be direct measurements of the hydrologic property or state; they tend to have relatively high accuracy and certainty, can be automated to provide high temporal coverage, but offer low spatial coverage (Hubbard et al., 2002; Kowalsky et al., 2006). Large-scale measurements are typically indirect (geophysical); they have relatively low accuracy and certainty, variable temporal coverage, but offer high spatial coverage (Hubbard et al., 2002; Kowalsky et al. 2006). Given the many differences between point and larger-scale measurements, together with differences in the costs of different measurement types, it can be difficult to assess the relative value of measurements of different types when planning a hydrologic monitoring network.

We propose a network design approach that assesses the value of hydrologic measurements in the context of the model(s) used for interpretation and the specific question(s) posed. That is, the value of the data is assessed based on their likely ability to constrain specific hydrologic predictions as part of an integrated monitoring network. One advantage of our approach is that it can consider all sources of model error and measurement error simultaneously. In addition, the approach has been designed to
consider both measurement performance and measurement cost explicitly. Finally, the approach has been designed to consider both direct and indirect measurements as part of an integrated monitoring network. The objective of this investigation is to test the ability of our design approach to identify optimal monitoring networks for a specific, hypothetical example. By examining a simple, well known system, we can test whether our approach recovers known optimal design elements for simple conditions and explore the insights that the approach offers for more challenging design questions.

Proposed Network Design Approach

From our perspective, described above, monitoring network design can be achieved in six steps. First, we define the range of predictions of interest based on all existing data and using all proposed models that honor the existing data. Second, we apply a cost function to translate these predictions to associated a priori performance costs; these performance costs represent the user-defined risk associated with the uncertainty of the hydrologic predictions of interest. Third, we propose a range of candidate measurements (types, locations, and times) to narrow the range of the predictions of interest; the cost of any individual measurement may not exceed the a priori performance cost. Fourth, we seek combinations of the candidate measurements that are most likely to reduce the expected performance cost; the cost of any measurement set may not exceed the a priori performance cost. Fifth, we compare the expected reduction in the performance cost with the cost of collecting and analyzing each proposed measurement set. Finally, we identify an optimal monitoring network based on
the trade-off between the expected reduction in performance cost and the projected measurement cost. In the following sections, we describe the components of our monitoring network design approach in general, as it could be applied to any monitoring design problem. Then, we present an application of our approach to a simple, hypothetical case study of monitoring network design for a feasibility assessment of a proposed artificial recharge site.

Consideration of Model Errors

We use an ensemble approach to capture the impacts of multiple, competing conceptualizations rather than relying on a single calibrated model. Therefore, the first step is to propose multiple models. We then calibrate each model to all existing data to produce an ensemble of acceptable parameter sets for each model. The goodness of fit of each parameter set to the existing data is used to define its relatively likelihood. If the model cannot be fitted to the data acceptably well, it is rejected. Finally, the predictions of interest are made with each model using each parameter set in the ensemble. Predictions made with each parameter set are weighted by the likelihood to form an ensemble of a priori predictions of interest; this is repeated for each model.

Application of a Cost Model

Cost functions are used frequently to quantify the risk tolerance of individuals or organizations in engineering design; however, this approach is not commonly applied in hydrologic investigations (Orr and Meystel, 2005). We use a cost model to describe the
cost associated with an error of a hydrologic prediction. If the model prediction was correct, the performance cost would be zero; costs increase with increasing over- or under-prediction. Cost functions for hydrologic applications can be complex and asymmetric (Renard, 2007). For example, underestimation of contaminant mass flux across a regulatory boundary may lead to uncontrolled releases, which may have higher costs than overestimation of mass flux, which could lead to overdesign of remediation efforts. For some applications, there may be a maximum cost associated with total failure of the hydrologic assessment.

If the correct values of the predictions of interest were known, the difference between the correct values and each model prediction would define the prediction error. The most likely prediction error, identified from the distribution of predictions of interest, could be used with the cost function to define the most likely performance cost. Alternatively, the cost of each prediction in the ensemble could be weighted by its probability and summed over the distribution of predictions to define the probability-weighted performance cost. In practice, the correct value of the prediction of interest is not known a priori. To address this uncertainty, we assume that several of the calibrated models are correct. We then calculate the maximum likelihood and risk-weighted average performance cost for each of these “truth” models. Finally, we average these costs across the truth models. If the truth models are chosen to represent a range of predictions of interest, the average performance cost tends to be more heavily weighted by the truth model that is associated with more costly outcomes. That is, the selected monitoring network is conservative, focusing on the more costly possible conditions.
Consideration of Measurement Errors

In contrast to previous network design approaches (e.g., James and Gorelick, 1994; Zhang et al., 2005), we do not consider observations to be error-free measurements of the property of interest. Rather, we assume that the measurements have unknown errors, but that the error characteristics (magnitude, distribution, range) can be approximated for each measurement method considered. The error distribution need not be normal or symmetric. Measurement errors can be homoscedastic or heteroscedastic, correlated or uncorrelated. To allow for this flexibility in handling measurement errors, we use a Monte Carlo approach to consider measurement error. Specifically, we predict each of the error-free candidate observations using each calibrated model, just as we calculated the distributions of values of the predictions of interest. We then add random errors to each candidate observation to form multiple measurement realizations; the added errors satisfy the defined measurement error characteristics. We then repeat our analyses for different measurement realizations.

To minimize the impacts of measurement interpretation errors, we use a coupled instrument model and hydrologic model approach (Margulis et al., 2002; Rucker and Ferré, 2004; Kowalsky et al., 2005; Ferré et al., 2006). That is, we use instrument forward models to predict instrument responses based on hydrologic model outputs. This ensures that the hydrologic and instrument response models consider the same underlying property distributions.
Consideration of Competing Measurement Sets

Substantial research has been conducted on effective measurement network design (e.g., Loaiciga, 1988; Massmann and Freeze, 1987a, 1987b; Knopman and Voss, 1989; Tucciarelli and Pinder, 1991; Ritzel et al., 1994). The goal of most of these assessments has been to reduce measurement redundancy to minimize future monitoring costs (Orr and Meystel, 2005; Major, 2007). For these applications, extensive existing data sets are considered with the objective of producing an acceptably accurate hydrologic assessment of interest with a subset of these measurements (e.g., Wagner, 1995; Vrugt et al., 2002; Michael et al., 2005). The ASCE Task Committee (2003) provides an extensive review of the state of the art of this post audit approach. With an extensive pre-existing data set in hand, all possible observation combinations can be assessed to identify near-optimal observation sets. However, as the number of observations increases, sampling techniques like genetic algorithms are used to reduce the computational demand (ASCE Task Committee, 2003; Reed et al., 2003; Reed and Minsker, 2004). We refer to approaches that consider complete, pre-existing measurement sets simultaneously as batch methods. Other studies have attempted to select future measurements, one at a time, by identifying the next measurement that is most likely to reduce the performance cost (e.g., Wagner and Gorelick, 1989; James and Gorelick, 1994). Very few studies have attempted to select combinations of future measurements (e.g., Zhang et al., 2005). The post audit analyses typically consider combinations of measurements explicitly because, in general, selecting one measurement at a time cannot adequately assess the combined value of multiple measurements. This
limitation is amplified if there is limited pre-existing data. We explicitly consider the likely value of combinations of future measurements. To do this, we form many sets of the candidate observations. Each candidate observation has multiple measurement realizations to account for measurement errors, as described above. We then consider the ensemble of predictions across all measurement realizations for each measurement set. For a small number of candidate measurements, all or most of the possible measurement sets can be examined. For a large number of candidate measurements, an optimization tool, such as genetic algorithms, can be used to explore the possible measurement sets formed from the candidate measurements. Or, rather than seeking an optimal measurement network, our approach can be used to compare the likely performance of specific measurement sets proposed by the user.

*Calculating the Performance and Measurement Costs of Each Measurement Set*

We calibrate each model to each proposed measurement set for each measurement realization. We then calculate the performance cost for each calibrated model for each truth model to produce distributions of the predictions of interest. We weight the predicted performance costs by their likelihoods and sum them to calculate the expected performance cost associated with each measurement set. Because each measurement set may have a different combination of types of measurements and different numbers of measurements, each could have a different measurement cost. Therefore, the measurement cost could be calculated as the sum of the installation, collection, and analysis costs of each measurement set (e.g., Lee and Kitanidis, 1996). The performance
cost can be plotted against the measurement cost. The minimum performance cost for each measurement cost value defines the Pareto front of optimal monitoring network designs. To make a final selection among the optimal network designs, the preferred trade off of the performance and measurement costs must be defined by the user.

*Static versus Recursive Network Design*

Approaches to using hydrologic models to assimilate existing data and predict the value of future data can be broadly classified as either static or recursive. Static methods design the network once, whereas recursive methods continually alter the design as data are collected. Post-audit methods are typically static, considering a large amount of preexisting data during design. Many practical hydrologic investigations require static designs because there is little or no opportunity to change the design after it has been installed. Our approach, as described above, can be applied directly to provide a static network design. To use our approach for recursive design would require a method to apportion measurement resources through time. Specifically, at the onset of monitoring, the Pareto front of optimal network designs would be produced based on the a priori data. The user would then decide how to balance the value of performance and measurement costs for the initial time step. Based in part on this decision, they would commit some portion of the total monitoring budget and the monitoring network with the lowest performance cost that does not exceed this measurement cost would be selected and installed. At some later time, the process would be repeated, including the a priori data
and the added measurements collected to that time. This continues until the performance
cost is acceptably low or until the monitoring budget is entirely committed.

Synthetic Case Study

The network design approach described above can be applied to any hydrologic
problem, regardless of its complexity. Its ensemble treatment of model uncertainty and
measurement error allows for consideration of the interactions of model and
measurement errors simultaneously. But, this approach can be computationally intensive,
especially when considering large measurement sets comprised of many candidate
observations. For these more challenging problems, it will be critical to propose efficient
hydrologic and instrument response models, to find efficient methods to solve these
problems, and to use efficient inverse methods. Here, we present a simple synthetic case
study to conduct an exhaustive examination of all measurement combinations. By
applying our approach to a simple case study we can test not only whether the approach
identifies some of the well-understood aspects of optimal monitoring design (e.g., the
inverse relationship between the measurement cost and the number of measurements in
the network), but also whether the approach identifies differences in performance among
possible monitoring networks that would be difficult or impossible to identify a priori.

The case study that we chose involves the assessment of a site as a potential
location for an artificial recharge facility. Advantageous conditions for artificial recharge
include highly permeable sediments and a thick unsaturated zone. This results in high
infiltration fluxes and a large in-situ filtering capacity in the vadose zone. The effective
hydraulic conductivity is the simplest metric to describe the suitability of a site for artificial recharge. However, it is very difficult to determine this value at the field scale based on limited core data. For this case study, we propose the time of propagation of a wetting front to a specified depth during a trial infiltration as a more representative proxy for site performance. Specifically, we define the prediction of interest as the time for the wetting front to reach a depth of 200 m below land surface. This depth could represent a large fraction of the unsaturated zone, but still lie far enough above the water table to avoid interaction with the saturated zone. We consider monitoring of a short duration trial infiltration event at a proposed artificial recharge site. Limited soil texture data are available for the site. We consider the use of up to eight ground-based gravity measurements measured at the edge of the flooded area. Measurements are made approximately daily during a five-day flooding event and every three days during two weeks of redistribution. These measurements have a known per-measurement cost.

We chose ground-based gravity as an example of an indirect geophysical method. Like many non-invasive methods, gravity has low installation costs and relatively high flexibility in choosing a monitoring location. However, it is unclear whether these indirect measurements have sufficient information to constrain a hydrologic analysis due to low signal to noise and due to nonunique interpretation of the gravity signal. For example, a gravimeter will respond to changes in the subsurface density distribution during infiltration and subsequent redistribution. However, the magnitude of the response depends on both the amount of water and its spatial distribution. We use our design approach to determine whether all or some of the candidate gravity measurements are
likely to improve the estimation of the time required for the wetting front to reach 200 m. As part of our analysis, we determine the expected change in prediction accuracy and uncertainty as a function of the number of measurements. For each optimal measurement set, we determine which measurement times are preferred. A similar analysis could be performed for other geophysical methods (e.g., time domain electromagnetics, surface nuclear magnetic resonance, electrical resistivity tomography) or direct measurements (e.g., time domain reflectometry, tensiometers, piezometers).

To further simplify the analysis, we assume that the hydrologic and instrument response models are correct (model structural error is not considered). We also assume that the subsurface can be considered to be homogeneous at the scale of the problem. Finally, we assume that the gravity measurement errors have a Gaussian distribution and a known variance.

**Cost Function and Measurement Costs**

We define measurement costs as the product of a fixed per-measurement cost and the number of measurements collected. We use a simple cost function, with zero cost associated with a perfect prediction, and a symmetric, linear cost associated with overprediction or underprediction. The maximum performance cost is fixed at the total cost of the project and is associated with a prediction over or under a threshold model prediction error. Finally, we consider both the maximum likelihood and the probability-weighted performance costs. We assume that the performance and measurement costs are directly comparable. Therefore, we seek to minimize the total cost, defined as the sum of
the performance cost and the measurement cost. We identify the optimal subset of the
candidate measurements and show how the optimal measurement network design
changes as a function of the per-measurement cost.

**Numerical Water Flow Modeling**

In practice, artificial recharge basins are cyclically flooded for periods of days to
weeks. The dry period between recharge cycles serves to minimize the reduction in
infiltration rates over time due to physical and biological clogging (Okubo and
Matsumoto, 1979; Bouwer, 2002). For this synthetic study, we consider a single cycle of
infiltration and drainage with 5 days of infiltration followed by 14 days of drainage in a
square recharge basin approximately 275 meters on a side. The sandy alluvial aquifer
beneath the recharge basin is assumed to be isotropic and homogeneous with a residual
water content ($\theta_r$) of 0.05 cm$^3$/cm$^3$ and a saturated water content ($\theta_s$) of 0.40 cm$^3$/cm$^3$. At
the beginning of the infiltration period, we assumed that the sediments have drained to
the residual water content, reflecting pre-operational hydrologic conditions. During
infiltration, we assume a zero ponding height with flux occurring at a rate equal to the
saturated hydraulic conductivity. During drainage, the flux at the ground surface is zero
and a rectangular drainage model (Jury and Horton, 2004) is used. This simple model
produces constant water content with depth during drainage. This is not a physically
reasonable result; water content should decrease toward the ground surface. Therefore,
using this model to interpret field data could introduce a significant model structural
cost. However, for this synthetic case study, we can use this simple model to reduce our
computational effort because we generate the synthetic data and then interpret it using the same model.

We use the Brooks-Corey (1964) function to relate the unsaturated hydraulic conductivity, $K$, [L/T] to the volumetric water content, $\theta$, [L$^3$/L$^3$]

$$K(\theta) = K_s \left( \frac{\theta - \theta_i}{\theta_s - \theta_i} \right)$$

where $\varepsilon$ is the unitless Brooks-Corey exponent (Brooks-Corey 1964) and $\theta_i$ is the initial water content. During drainage, the water content as a function of time is

$$\theta(t) = \left( \frac{\varepsilon K_s}{L} (\theta_s - \theta_i)^{\varepsilon+1} (t - t_d) + (\theta_s - \theta_i)^{-\varepsilon} \right)^{-1/\varepsilon} + \theta_i \quad \text{for } t > t_d$$

where $L$ [L] is the depth of the wetting front, and $t_d$ [T] is the time when drainage begins. Note that the Brook-Corey exponent is only employed to describe drainage; infiltration occurs at full saturation. Three different textures of sandy alluvium are considered as “truth” models for our analysis. The soils are textures A, B, and C. The Brooks-Corey exponent and the saturated hydraulic conductivity of each soil type are reported in Table 1.

**Numerical Gravity Response Modeling**

During infiltration, the saturation of the near-surface sediments increases. This causes a local increase in bulk density. When infiltration ceases, the alluvium drains under the force of gravity resulting in a redistribution of the recharged water; we assume zero evapotranspirative loss. After the end of infiltration there is a decrease in the bulk
density near the ground surface and an increase in the bulk density at greater depth. As a result, the gravity change signal decreases because the distance between the gravimeter and the center of mass of the infiltrated water increases. The gravitational response was simulated at a distance of 2 m away from the edge of the recharge basin at eight measurement times, four each during infiltration (1.25 day interval) and drainage (3.25 day interval). In this application, the gravity response is a relative gravity response defined as the change in the vertical component of the gravitational acceleration at a point on the ground surface compared to the vertical gravitational acceleration at that location before infiltration began. Changes in the gravitational response over time were calculated assuming that the compressibilities of the aquifer and of the pore water are negligible with regard to the gravitational response. As a result, the change in bulk density depends only on the change in volumetric water content of the sediments. The change in the gravity change response with time was simulated using a right rectangular prism model (e.g., Nagy, 1966; Cady, 1980; Montana et al., 1992). For our hydrologic models, the prism is a region of uniform water content change (from preinfiltration conditions) that extends from the ground surface to the wetting front. The lateral extent of the prism is identical to that of the detention basin.

*Gravity Measurement Errors*

When proposing an indirect measurement method for a hydrologic investigation, the first step is to determine whether the method is likely to have sufficient sensitivity to detect the expected hydrologic changes (Damiata and Lee, 2006; Blainey et al., 2007).
Under favorable conditions, ground-based relative gravimeters are sensitive enough to measure changes in gravitational acceleration caused by changes in the subsurface density distribution associated with hydrologic processes. However, hydrologic applications of temporal gravity surveys are often limited by a relatively low signal to noise ratio (Poeter, 1990). For example, portable relative gravimeters have a measurement resolution of approximately 1 microGal = $1 \times 10^{-8}$ m/s$^2$ (Scintrex). This is commonly considered to represent a change of 2.5 cm in the thickness of an extensive layer of water at a significant depth (Telford et al., 1990). For our application, far more water is added and the water content change occurs at shallow depths. Therefore, it is likely that relative gravity will be sufficiently sensitive.

Once it is determined that a measurement is likely to be sensitive enough to respond to a hydrologic change, the next step is to determine whether the measured value is likely to provide useful information for hydrologic assessment (Blainey et al., 2007). The utility of measurements for parameter estimation depends in part on the measurement error (Poeter and Hill, 1997). We assumed that gravity measurement errors are normally-distributed and homoscedastic. In practice, gravity measurements must be adjusted using standard gravity corrections (Telford et al., 1990). Many corrections – such as instrument tilt, temperature, and earth tides – are automated by the instrument. Relative, not absolute, differences in gravity are considered with this case study. Therefore, some standard gravity corrections (e.g., Bouguer and elevation corrections) are not needed because the corrections are identical for measurements made before and after the artificial recharge event. We assumed that corrections affecting relative gravity surveys, e.g.,
earth-tide and instrument drift, have been applied correctly. That is, errors in these corrections will contribute to the magnitude of the variance of the measurement error, but they will have zero mean. With this assumption, the simulated measurement errors were based on the manufacturer-reported values of measurement repeatability with a zero mean.

Measurement errors for the gravity response were generated with one standard deviation equal to 20 milliGals ($20 \times 10^{-5} \text{ m/s}^2$) and added to the noise-free responses to produce synthetic measurements. These measurement errors are a conservative representation of the standard field repeatability value of less than 5 microGals for the CG-5 portable relative gravimeter (Scintrex, Concord, Ontario). We consider that this increased error accounts for incorrect gravity corrections as well as other typical sources of measurement uncertainty (e.g., operator inconsistency).

For each error-free candidate measurement, we generated 100 random measurement error realizations. The error is normally distributed with a constant variance and zero mean at any candidate measurement time across the 100 error realizations. In fact, an identical set of error realizations is used for each of the eight candidate measurement times; but, the sequence of the errors was shuffled for each candidate time. With this approach, the specific set of errors did not unduly affect any single candidate measurement time, permitting a fair comparison of the likely value of the candidate measurement times. However, errors for a single measurement realization, across the candidate measurement times, are not correlated and they are not ensured to be normally distributed or to have a zero mean.
Parameter Estimation

Parameter estimation was conducted using a coupled hydrogeophysical approach (Rucker and Ferré, 2004; Kowalsky et al., 2005; Ferré et al., 2006). That is, during each step of the inversion process, the hydrologic parameter values were perturbed, the hydrologic forward model was run, and the gravity responses were calculated for the predicted hydrologic response. The hydrologic parameters that provided the best fit between the synthetic gravity measurements (with error) and the simulated gravity response (without error) were identified. Two approaches were used: single- and multiple-fit inversion. For each approach, the optimal parameters were identified for each error realization for the candidate measurements included in the measurement set. For the synthetic case study, we could examine every combination of the 8 candidate measurements. The specific objective function minimized was:

\[
ObjF(x) = \sum_{i=1}^{n_t} (g_{\text{meas}}(x) - g_{\text{sim}}(x))^2
\]  

(3)

where the integer \( n_t \) is the number of measurement times \( (n_t = 8) \). By using the coupled hydrogeophysical inversion approach, measurements of different types can be assigned weights in the objective function based on the inverse of their measurement variance (Hill 1997). However, in this case, only a single measurement type is considered, so all measurements received the same weight.

Two different optimization algorithms were used to estimate the saturated hydraulic conductivity and the Brooks-Corey exponent parameters, simultaneously. First,
we considered all possible combinations of measurements for measurements sets of size $n$ where $n$ varies from 1 to 8. This results in 8, 28, 56, 70, 56, 28, 8, and 1 unique combinations of measurement sets of size 1 through 8 measurements, respectively, leading to 255 unique measurement sets. A set of observations, with errors, was formed for each of these 255 sets for each of the 100 measurement error realizations, leading to 25,500 observation sets. We used the Nelder-Mead simplex algorithm (Nelder and Mead, 1965; Lagarias et al., 1998) to find the minimum of the objective function for each of these observation sets. The Nelder-Mead simplex algorithm is a direct local search method for multidimensional unconstrained minimization; neither numerical nor analytic derivatives are employed to approximate the function of interest for minimization (Nelder and Mead, 1965; Lagarias et al., 1998). In $n$-dimensional space, this algorithm is characterized by $n+1$ vectors that are the vertices of the simplex. For each step of the search, a new point is generated in or near the simplex. The objective function value of the new point is compared to the objective function value of the vertices in the existing simplex and through a sequence of logical steps. The new point may replace an existing vertex of the simplex. The search is repeated until the diameter of the simplex is less than the specified tolerance, which was set to $1 \times 10^{-9}$ milliGal$^2$ for the objective function and $1 \times 10^{-8}$ for the parameter estimates. The result of this analysis was 100 paired values of the hydraulic conductivity and the Brooks-Corey exponent for each of the 255 measurement sets.

Based on this exhaustive study of the best fit parameter sets for all 255 measurement sets, we identified the optimal measurement set for each number of
observations. Specifically, the optimal measurement set was identified as the set that produced the smallest absolute error for the prediction of interest when averaged over 100 measurement error realizations. For this smaller number of measurement sets, we examined the interaction between prediction uncertainty and the number of measurements in a measurement set. To better define parameter estimation uncertainty, we used the Shuffled Complex Evolution Metropolis (SCEM) algorithm (Vrugt et al., 2003). The SCEM algorithm is a global optimization algorithm that infers the best parameter set and its underlying posterior distribution given an initial feasible parameter space. That is, for a single measurement set, the SCEM algorithm produces a population of parameter sets that fits the observations and the likelihood of each of these parameters sets. In contrast, the Nelder-Mead simplex algorithm produces a single best-fit parameter set. The SCEM algorithm merges the strengths of several different types of algorithms including the Downhill Simplex procedure (Nelder and Mead, 1965) with random search, the systematic evolution of points in the direction of global improvement, and complex shuffling. The SCEM algorithm is a Markov Chain Monte Carlo sampler that uses the Metropolis-Hastings strategy (Metropolis et al., 1953) to evolve the population. Once the algorithm terminates, the Gelman and Rubin convergence diagnostic was used to determine if model convergence was achieved (Gelman and Rubin, 1992). We expect that such a hybrid approach that identifies optimal (or near optimal) measurement sets with a faster, less robust technique and then performs a more complete analysis of fewer measurement sets will be necessary for more complex field applications of our approach, as well.
Results

Results are presented in the order that they were obtained during the analysis. That is, in presenting the results, we have attempted to mimic the process that a user would follow in applying our measurement network design approach.

*Initial Uncertainty*

Network design analysis can be complicated and computationally expensive; many hydrologic assessments will not warrant the effort. The first step to determining whether such an analysis is worthwhile is to quantify the initial prediction uncertainty. This can be achieved as follows. First, sample parameters from a uniform distribution with ranges set by a priori limits on the parameter values. Clearly, the more restricted these ranges are, due to more extensive existing site information, the smaller the initial uncertainty. Then use the parameters to drive the forward hydrologic model to produce the predictions of interest. In our case, we calculated the time for the wetting front to reach 200 m for 100 randomly sampled parameter sets, with $K_s$ ranging from an order of magnitude lower to an order of magnitude higher than that of alluvium texture B and the Brooks-Corey exponent ranging from 3.0 and 4.5. Alternatively, and preferably, an ensemble inverse method could be used. This approach is described in more detail, below. If the range of predictions is acceptably small, no further data are necessary. If there is significant uncertainty in the predictions, the analysis can move to the next step: estimation of parameter identifiability.
Measurement Sensitivity and Parameter Identifiability

We recommend an assessment performed in three stages before collecting measurements with any new instrument or when applying an accepted method to new hydrologic conditions (Blainey et al., 2007). First, an analysis should be completed to determine whether the measurement method is likely to detect signals associated with the hydrologic process of interest. This analysis should consider both measurement resolution and expected measurement errors. Second, a first-level assessment of the likely utility of the method should be conducted to determine whether the measurements are likely to be useful for identifying hydraulic properties or hydrologic processes. For this analysis, it is most informative to adopt conditions that are favorable for the successful use of the method. If the analysis suggests that the measurement method is unlikely to be successful under these favorable conditions, it is unlikely that the method will be useful under more challenging conditions. If the analysis suggests that the measurement method is likely to be successful under some conditions, a third, more complete, site-specific analysis should be performed using all available information before collecting field measurements. This approach can provide an objective screening of proposed measurement methods for specific hydrologic applications thereby reducing the number of unsuccessful uses of geophysical and other indirect methods.

Based on the preexisting information, we proposed three “truth” models of the soil hydraulic properties (identified as textures A, B, and C in Table 1). Based on these parameters sets, we predict the expected values of the hydrologic and instrument responses. For our application, we show the cumulative infiltration, wetting front depth,
and gravity time series (Figure 1). The eight gravity measurements in the time series are referred to as g1, g2, g3, g4, g5, g6, g7, and g8; g1 is the first measurement, g4 occurs at the end of infiltration, and g8 is the last measurement collected after 18 days have elapsed since the start of infiltration. The vertical line on each panel of Figure 1 shows the end of infiltration (time g4). For our imposed boundary conditions, cumulative infiltration (the length of water added) increases linearly during infiltration and remains constant during drainage. In response, the depth to the wetting front increases linearly during infiltration then slows during drainage. The advance of the wetting front during drainage is influenced by the sediment texture. For example, 50% of the total advance of the wetting front occurs after the end of infiltration with texture A. In contrast, 44% of the advance of the wetting front occurs after the end of infiltration for the finer textured soil C. One significant difference among the truth models is that the wetting front reaches the depth of interest (200 m) during the monitoring period for texture A, but it does not reach the target depth during monitoring for textures B or C. The gravity change is only slightly nonlinear during infiltration. This nonlinearity arises because the depth at which the water content is changing increases with increasing time. The gravitational response is inversely dependent on the square of the distance of the center of mass of the mass change and the gravity instrument, leading to a decrease in the gravity change with time. During drainage, the gravity change response decrease is clearly non-linear. This response occurs because there is a rapid decrease in the water content at early time, leading to a strong decrease in the gravity signal. Essentially, water is released from storage throughout the previously wetted portion of the profile, accompanied by further
advance of the wetting front at greater depth. Again, due to the inverse distance sensitivity of gravity, this leads to a decrease in the gravity signal through time with no associated change in the total amount of water stored. At later time, the rate of the wetting front advance and water content change are slower, leading to a slower rate of change in the gravity response. Because the gravity response during drainage depends on the water content change with time, the response varies among the three soils examined. The coarsest soil has a 24% decrease in the gravity signal during drainage, whereas the finest soil has a 12% change. One realization of the gravity measurement error is shown for each soil type. Note that the expected gravity changes during infiltration and drainage are larger than the measurement error, indicating that the proposed method has sufficient sensitivity to identify the hydrologic change. However, it may be difficult to assess the nonlinearity of the response given the measurement error; therefore, it is unclear from examining the predicted response whether the data will be useful for parameter estimation.

Average Expected Predictions of Sets with Different Numbers of Measurements

As an initial test of the utility of the gravity data to constrain the hydrologic model, we examined whether it was possible to recover the depth of the wetting front at the end of the monitoring period. In addition, we examined whether there was a general relationship between the number of measurements and the error in the estimated depth of the wetting front at the end of the monitoring period. To do this, each pair of estimates for $K_s$ and $\varepsilon$ from the 255 unique measurement sets was applied in the forward hydrologic
model to predict the depth of the wetting front \( (z_{wf}) \) at the end of the monitoring period for each of the three soils considered. We combined the inversion results of all 100 error realizations for all measurement sets with the same number of measurements. Then, we calculated the relative error (average prediction error divided by the correct wetting front depth) for each measurement set size and for each soil type. The results show that the estimation of the water table depth, which is a hydrologic condition that occurs during the monitoring period, is likely to be very successful (maximum relative error is 4% for a measurement set with two measurements). The results also show the expected nonlinear decrease in likely prediction error with an increasing number of measurements (Feyen and Gorelick, 2004) (Figure 2 A). That is, there is a diminishing return as more measurements are added to the measurement set. This diminishing return is a key characteristic of a monitoring network that may benefit from the application of a design analysis because it demonstrates that the cost of a measurement may exceed its value depending upon the composition of the monitoring network. For this example, it appears that prediction of the wetting front depth at the end of the monitoring period is not improved significantly when the number of measurements increases from 5 to 8. Notice also that the performance depends strongly on the soil texture for a small number of measurements; the performance is approximately equal across textures for larger measurement sets. These general results (diminishing return and reduced dependence on specific subsurface conditions with increasing number of measurements) could have been anticipated without conducting our network design analysis. In this case, we include them to demonstrate that our approach does recover these anticipated results. In addition, we
include them to show that our analysis can give more specific insights into the relative value of different numbers of measurements. Measurement sets with only one measurement contained insufficient information for parameter estimates to converge for all realizations of measurement error and consequently were not evaluated.

The previous analysis demonstrated that the monitoring network was likely to benefit from the application of a design methodology and that the measurements were likely to constrain an assessment of a hydrologic state during the monitoring period. Next, we examined the likely performance of the proposed monitoring plan for the prediction of interest, which occurred after the monitoring period for two of the three soils considered. The relative error in the time of the wetting front to reach 200 m depth, $t_{wf}$, also decreases as the number of measurements increases. But, unlike the assessment of the wetting front depth at the end of the monitoring period, very large estimation errors are expected for the finer soils textures considered, even with all eight measurements. This difference in performance occurs because the prediction of interest occurs during the monitoring period for texture A, but not for textures B and C. Based on the previous analysis, it is not clear that a monitoring network comprised of the candidate measurements is likely to constrain the prediction of interest for all plausible conditions. Therefore, the next step was to examine the likely performance of individual measurement sets for the three “truth” soils. Finally, the results of these analyses will be combined to determine whether the proposed measurements are likely to be worthwhile. Specifically, in our approach, if the proposed measurements are not likely to be worth the cost of collecting them, the optimal monitoring set will have zero measurements.
Expected Predictions of Specific Measurement Sets

The goal of this stage of the analysis is to determine whether specific measurement sets are likely to perform well for the soils and targets considered. Specifically, we examine the variability in performance among possible combinations of sets of a fixed number of measurements. For example, we show the ranked absolute prediction error for each of the 70 unique measurement sets that can be formed from four measurements according to the absolute error of $t_{wf}$ (Figure 3). For convenience of plotting, we normalize the errors from zero to one. That is, a rank of 0 represents the best-ranked monitoring network and 1 represents the worst-ranked monitoring network; the ranking interval is $1/(n-1)$ where $n$ is the number of measurement sets with a fixed number of measurements. For texture A, the coarsest sediment considered, the variation in performance among measurement sets is small; that is, many of the best performing measurement sets have the same expected performance. For the finer soil textures (B and C), there is significant variation among measurement sets; as a result, a subset of the possible measurement sets offers improved performance. The best performing measurement set for soil texture A is g3-g4-g5-g6, which includes two measurements collected during infiltration and two collected drainage. As might be expected, this measurement set contains the three measurements (g4, g5, g6) with the largest signal to noise ratios. Again, this demonstrates the ability of our approach to recover design elements that would otherwise rely on the intuition of an expert in both hydrologic and geophysical responses. As the soil texture becomes finer, the best measurement set includes more late-time drainage measurements: g4-g5-g7-g8 for texture B; and g3-g5-
g7-g8 for texture C. This result may also be expected because the late time measurement for finer soils still has a relatively large signal to noise ratio, but it maximizes the change from the condition at the end of drainage. In this case, even an expert would not be able to determine the optimal balance of measurements without conducting an analysis similar to that shown here. The late-time drainage measurements are important for constraining estimates of $\varepsilon$, which has a larger percent bias than $K_s$ (not shown). Although the three optimal measurement sets appear similar, their performance for the other measurement sets is far from optimal (Figure 3). For example, choosing the optimal set for texture A results in an order of magnitude increase in the expected absolute error for textures B and C. Conversely, because texture A is relatively insensitive to the measurement set selected, it performs well using the optimal sets selected for B or C. The measurement set that has the minimum error across all three soil types is the same set selected for texture C (g3-g5-g7-g8); this is not true for all measurement set sizes. Such an analysis could be used to select some measurements as common to many or most optimal measurement sets. This pre-selection of some measurements may be sufficient for some monitoring applications; for example, if the choice is made a priori to use a fixed number of the candidate measurements. Or, this screening approach could be used to greatly reduce the number of possible combinations, and therefore, the computation effort for problems considering large numbers of candidate measurements and/or large measurement sets.

**Expected Predictions and Uncertainty of Specific Measurement Sets**

The previous analyses have only considered a single prediction based on each
measurement set, determined using the Nelder-Mead simplex algorithm. Once we narrowed the number of measurement sets using this screening approach, we used SCEM to estimate the distributions of $K_s$ and $\varepsilon$ values. Specifically, we considered the three measurement sets chosen as optimal for each soil for each number of measurements in the set. We also considered the optimal set selected for all three soils combined for each number of measurements in the set. We examined the ability of each set to recover the value of $t_{wf}$ for each of the three “truth” soils. For each realization of measurement error, the 90% confidence intervals on the estimated parameter sets were determined and these values were used to examine the variability in the prediction of interest (Figure 4). In general, as soil textures become finer (from A to C) the bias of the prediction (difference between the correct value and the mode of the prediction ensemble) increases. In addition, the uncertainty of the prediction increases, as shown by the increased range of prediction distribution. For the coarsest soil, texture A, the most likely values of the prediction of interest were close to the true value for measurement sets with more than 2 measurements and the variation in prediction uncertainty among measurement sets of size 4, 6, and 8 is small. For coarser alluvium B, the distribution of $t_{wf}$ is not symmetric: the result is skewed toward underprediction. In this case, the estimates of log $K_s$ are approximately normally distributed, but the predicted $t_{wf}$ decreases more rapidly with decreasing $K_s$. The uncertainty in the prediction of $t_{wf}$ increases continuously with a reduction in the number of measurements considered (Figure 4). For the finest alluvium (texture C), even with 8 measurements the mostly likely parameter estimate produces a prediction more than 45 days smaller than the true value of 93 days. This large error
occurs even though the value of $K_s$ is estimated within 0.2 % of its correct true value; the error is due to a relative error of $\varepsilon$ of 8 percent. These results illustrate one of the most challenging aspects of monitoring network design: the optimal monitoring network, and the expected performance of that network, depend critically on the unknown hydrologic conditions. The previous analysis showed that it can be difficult to design a monitoring network given the uncertainty of the hydrologic conditions. But, in fact, hydrologists must do this all the time. The next step of our approach considers this uncertainty explicitly.

**Expected Performance Cost of Specific Measurement Sets**

We design the monitoring network that results in the minimum performance cost across the plausible hydrologic conditions. For comparison, we consider the performance cost in two ways. First, we assume that the most likely prediction made for each error realization is equally likely; we then calculate the performance cost for each error realization and average them to give the expected performance cost. Referring to Figure 4, this first approach is equivalent to choosing a monitoring network based on the performance cost associated with the likely difference between the most likely prediction and the correct value, without consideration of the uncertainty (range) of the predictions. We use SCEM to identify the most likely prediction for this approach, although a simpler inverse method (e.g., the Nelder-Mead simplex) could be used. In the second approach, we calculate the performance cost for each parameter set within the 90% confidence interval for each error realization. We then multiply each of these by the likelihood of the
parameter set, based on the goodness of fit to the measurement set, which is provided by SCEM. Then, we sum these probability-weighted performance costs to find the performance cost for that error realization. Finally, we assume that each measurement error realization is equally likely, and we average the performance cost across all error realizations to define the expected, probability weighted performance cost for each measurement set. Referring to Figure 4, this second approach attempts to account for both the bias and the uncertainty shown in the prediction ensemble. In both cases, we consider each of the three truth soils to be equally likely. So, we average the performance costs calculated for the three truth soils.

Our approach is based on the assumption that, in practice, hydrologic predictions of interest must be translated to a measure of the acceptability of the risk of failure for a hydrologic assessment. To consider the trade-off between the quality of hydrologic predictions and measurement costs, we apply a performance cost function that translates a prediction error to a performance cost, \( p_c \). We use a simple cost function, with zero cost associated with a perfect prediction, and a symmetric, linear performance cost calculated as a function of the absolute bias of the prediction of \( t_{wfg} \) as

\[
\begin{align}
p_c &= \left( \frac{p_c^m}{\Delta t} \right) \left| t_{wfg} - \hat{t}_{wfg} \right| \quad \text{if} \quad \left| t_{wfg} - \hat{t}_{wfg} \right| < \Delta t ; \\
p_c &= p_c^m \quad \text{if} \quad \left| t_{wfg} - \hat{t}_{wfg} \right| \geq \Delta t ;
\end{align}
\]
where \( \hat{t}_{zwf} \) is the prediction estimate and \( t_{zwf} \) is the error-free prediction. In many cases, there is a threshold error, \( \Delta t \), beyond which a hydrologic assessment has no value. In this case, we set this threshold error at 100 days; predictions that have an absolute error greater than this amount have a fixed maximum cost, \( p_c^m \).

Combining the Measurement and Performance Costs to Define the Total Cost

Following either of the two approaches described above, we can define an expected performance cost for each measurement set. For a real-world problem, the measurement cost could be defined exactly. These costs could include installation, operation, and analysis costs. But to make our synthetic problem more general, we consider a range of measurement costs. Specifically, instead of using an application-specific absolute measurement cost, we consider the measurement cost, \( c \), to be a fraction of the a priori performance cost. This ensures that the maximum cost of collecting all eight measurements will be less than the a priori performance cost. As stated above, if this condition were not met, only a measurement set with a smaller maximum number of measurements should be considered. The a priori performance cost used to define the measurement cost for a specific measurement set, \( m_c \), is the average of the initial costs (zero measurements) for all three truth soils, \( \bar{p}_c^a \):

\[
m_c = \frac{\bar{p}_c^a cn}{n_{\text{max}}}
\]  

(6)
where $n$ is the number of measurements in the measurement set and $n_{\text{max}}$ is the maximum measurement set size (8 in our analysis). We consider the impact of the per-measurement cost ranging from a total cost of 0% to 100% of the a priori performance cost for all eight measurements.

The measurement cost is a real cost, whereas the performance cost is a risk-based cost. Therefore, these costs may not be considered to be equivalent in all cases. For simplicity, we assume that the cost function is constructed such that these costs are equivalent. This allows us add the performance and measurement costs to define the total cost. Once the costs are combined, the multi-objective design problem (measurement cost versus the performance cost) reduces to a single objective problem of minimizing the total cost. As an example, the performance, measurement, and total costs are shown for the initial conditions or zero measurements (as described in the Initial Uncertainty section) and for measurement sets comprised of 2 through 8 measurements with a cost fraction of $c = 0.40$ for texture A using the most-likely prediction to calculate the performance cost (Figure 5A). The initial performance cost was determined using the Nelder-Mead simplex method of calculation, which always returns a lower cost estimate than the probability-weighted approach. The results indicate that adding measurements reduces the total cost from the initial conditions because our initial prediction uncertainty is so high. The optimal monitoring network for these conditions based on the probability-weighted approach (Figure 5B, open triangles) has three measurements (g3-g4-g5). Note that, for this soil, there is very little error in the most likely prediction, even for measurement sets including only two measurements. This can be seen in very similar
values of the measurement costs (Figure 5A) and the total cost using the most-likely approach (Figure 5B, filled triangles). Therefore, using the performance cost based on the most likely response suggests that even fewer measurements should be collected (two measurements: g4 and g5).

As expected, the most-likely approach to assessing the performance cost resulted in the same or fewer measurements than the risk-based assessment for a given measurement cost. This is because the most-likely approach does not consider the possible costs of the more-costly, but less likely, outlying predictions. By considering a range of predictions, over-confidence in a poor prediction is avoided in the cost analysis. This is particularly important when predictions have significant errors. Only when the true prediction is known, as in this synthetic case study, is it possible to compare optimal networks based on the most likely and probability-weighted approaches. In general, using the risk-based approach will minimize the impact of poor predictions on network design. We believe that it is more appropriate to use the risk-based assessment approach, because this approach captures the reduction in prediction uncertainty due to adding measurements, which is often as important and the impacts of reduced prediction error. However, we present the most-likely approach because it is often used, albeit implicitly, in defining the performance of a hydrologic assessment.

To examine the impact of the measurement cost, we examined how the optimal number of measurements changes with the measurement cost fraction \( c \) for texture A using the prediction based performance cost (Figure 6). The performance cost does not depend on the measurement cost, so the total cost for any number of measurements
increases linearly with an increase in the per measurement cost. The results show that, as expected, when the measurement cost is low ($c < 0.35$) all eight measurements should be collected. As the cost of the measurements increases ($0.35 < c < 0.85$), three measurements should be collected. If the $c$ value exceeds 0.85, no measurements should be collected. Again, this demonstrates that our design approach can capture general design elements that could be determined using expert input (e.g., reduce the number of measurements as measurement cost increases), but our approach adds quantitative information regarding the threshold measurement costs and the associated numbers of measurements to collect. In fact, if $c$ exceeds 0.9, the a priori cost for soil texture A is less than the measurement cost of 7 or 8 measurements. The recommendation that no measurements should be collected does not indicate that the prediction will not be improved by adding measurements; rather it shows that the improvement will not be worth the cost of collecting the data. This conclusion is very difficult to draw without applying a quantitative network design analysis such as ours.

Choosing a Monitoring Network

The preceding analysis was presented for texture A. However, when selecting a monitoring network for future measurements, the specific subsurface conditions are unknown. Here we show how the design results vary among soil types and for the two different approaches to defining the performance cost. For simplicity of presentation, we show only the number of measurements selected as a function of the per measurement cost (Figure 7). These numbers of measurements are identified using the approach.
leading to the results shown in Figure 6. As a consequence, measurement sets with the same number of measurements may be composed of different observations for different textures. As expected, the optimal number of measurements suggested increased with decreasing measurement cost; this result holds for the three truth-textures considered.

In most cases, the finer soils, which had less certain and more biased predictions, support selection of fewer measurements. At first, this may seem counterintuitive. But, this demonstrates that the design approach is balancing the likelihood of improving the prediction against the cost of that improvement. That is, for the finer soils, the predictions are so poor that even with more measurements little improvement occurred in the prediction of $t_{w}f$.

Finally, we calculated the average performance cost across all three truth soils using the measurement sets that had the minimum error across all three truth soils and identified the number of measurements that led to the minimum total cost (Figure 7). In this case we assumed that each of the three truth soils was equally likely; but if prior information supported an estimate of the likelihood of each truth soil, these likelihoods could be included in the cost averaging. We expected that this approach would give greater weight to the most costly plausible conditions (in our case, texture C). In general, this expectation was confirmed; the number of measurements selected for all soils was equal to or less than the number selected for texture C. However, it is important to remember that the specific observations selected may be different for texture C or for all soils.
Conclusions

Our proposed network design approach assesses the value of hydrologic measurements in the context of the model(s) used for interpretation and the specific question(s) posed. To evaluate the performance of our static network design approach, we used a synthetic case study of monitoring artificial recharge with time-lapse gravity change observations. In this study we employed an ensemble approach that, while more computationally demanding, focused on the consideration of the effect of measurement errors on the design process. The goal of this approach was to develop a method to make informed choices among the many available models and measurements by comparing the cost of the measurements with the benefit of improved hydrologic assessment. Specifically, the goal was to identify which measurement sets are most likely to provide an optimal trade-off between the reduction of the costs associated with the error and uncertainty of a hydrologic prediction of interest and the cost of collecting additional measurements.

The method captures expected design elements that could be predicted by someone with expertise in hydrologic and geophysical responses. For example, our approach showed that measurements with high signal to noise should be included in the measurement set. In addition, there is a diminishing return in prediction improvement as more measurements are added to the monitoring network. Furthermore, the optimal number of measurements increased with decreasing measurement cost. Finally, as expected, designing a network based on the most-likely prediction leads to fewer measurements than designing the network to consider prediction uncertainty. Our
analysis also confirmed some general expectations regarding the design of monitoring networks. For example, the best measurement set is a function of the particular prediction of interest and of the actual subsurface conditions. However, our approach provides a method to consider these influences quantitatively and to use that analysis to guide network design.

A more subtle aspect of our design approach is that it captures the trade-off between prediction uncertainty and the cost of measurements. By comparing measurement sets for three different soil textures, or truth models, we saw that the size and composition of the best measurement set also depended on relationship between the prediction of interest and the hydrologic conditions. For example, finer soils, which had less certain and more biased predictions, support selection of fewer measurements because predictions are so poor that more measurements result in little improvement of the prediction of interest. This is an unexpected result from our approach to monitoring network design that would not be evident with other approaches that do not allow for the possibility that additional measurements will not be worthwhile.

The network design approach that we describe can be applied to any hydrologic problem, regardless of its complexity. The general approach presented herein could be modified easily to incorporate additional complexities such as model structure errors, multiple measurement methods, multiple measurement locations, competing hydrologic process models, and more complete measurement costs (e.g., instrument installation, operation, and analysis). However, as the network design becomes more complex, it is important to consider whether this exhaustive analysis presented here is warranted or if it
could be replaced by a faster, but still robust technique. For example, as part of this analysis, we showed several intermediate results that could have been used for approximate network designs. In the end, it will be important to match the complexity and available budget of a project with an appropriate design approach. Our future efforts will be aimed at developing such a set of approaches with varying complexity and computational cost.

Acknowledgements

The authors would like to thank Kevin Mickus for graciously supplying the FORTRAN code for the solution of the gravity field with right rectangular prisms. Additionally, the authors that Andrew Hinnell for providing his modification to the Jury and Horton rectangular drainage model. We gratefully acknowledge support from the University of Arizona, Technology and Research Initiative Fund, Water Sustainability Competitive Grants Program.

References


Scintrex CG-5 brochure, Micro-g LaCoste, Lafayette, Colorado, 2 p.


### Tables

<table>
<thead>
<tr>
<th>Property or parameter, notation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied flux at ground surface (m/s) during infiltration</td>
<td>(^a) (1.0 \times 10^{-4}), (^b) (7.92 \times 10^{-5}), (^c) (5.83 \times 10^{-5})</td>
</tr>
<tr>
<td>Applied flux at ground surface (m/s) during drainage</td>
<td>0</td>
</tr>
<tr>
<td>Brooks-Corey exponent, (\varepsilon)</td>
<td>(3.25^1, 3.75^2, 4.25^3)</td>
</tr>
<tr>
<td>Gravity observation location (m) from edge of detention basin</td>
<td>2</td>
</tr>
<tr>
<td>Initial uniform water content with depth (cm(^3)/cm(^3))</td>
<td>0.05</td>
</tr>
<tr>
<td>Measurement times (days elapsed since start of infiltration)</td>
<td>1.25, 2.5, 3.75, 5, 8.25, 11.5, 14.75, 18</td>
</tr>
<tr>
<td>Residual water content (cm(^3)/cm(^3)), (\theta_r)</td>
<td>0.05</td>
</tr>
<tr>
<td>Saturated hydraulic conductivity (m/s), (K_s)</td>
<td>(^a) (1.0 \times 10^{-4}), (^b) (7.92 \times 10^{-5}), (^c) (5.83 \times 10^{-5})</td>
</tr>
<tr>
<td>Saturated water content (cm(^3)/cm(^3)), (\theta_s)</td>
<td>0.40</td>
</tr>
<tr>
<td>Time of drainage (days elapsed since start of infiltration), (t_d)</td>
<td>5 to 18</td>
</tr>
<tr>
<td>Time of infiltration (days elapsed since start of infiltration)</td>
<td>0 to 5</td>
</tr>
</tbody>
</table>

\(^a\) Alluvium texture A.

\(^b\) Alluvium texture B.

\(^c\) Alluvium texture C.

Table 1. Parameter values and parameter ranges used for synthetic gravity measurements.
Figure 1. Times series of hydrologic and gravity response to infiltration and drainage. A. Cumulative infiltration. B. Depth to the wetting front (m). C. Gravity change with error bars representing a spread of two standard deviations of measurement error for $K_s = 1.0 \times 10^{-4}$ m/s and $\varepsilon = 3.25$ (solid line), $K_s = 7.92 \times 10^{-5}$ m/s and $\varepsilon = 3.75$ (dashed line), and $K_s = 5.83 \times 10^{-5}$ m/s $\varepsilon = 4.25$ (dotted line). The vertical solid line represents $t_d$, the time of drainage.
Figure 2. A. Relative error of the depth to the wetting front at the end of the monitoring period. B. Time elapsed (days) for the wetting front to reach 200 m as a function of the number of measurements in the measurement set.

Figure 3. Absolute error of the time for the wetting front to reach 200 m for textures A, B, and C for 4 measurements. The symbols identify the best measurement set for each soil texture. The best measurement set for texture C also has the minimum error across all three soil textures.
Figure 4. Predictions of the time for the wetting front to reach a depth of 200 m below land surface for alluvium textures A, B, and C with overall best measurement sets: g3-g8, g3-g5-g7-g8, g2-g4-g5-g6-g7-g8, g1-g2-g3-g4-g5-g6-g7-g8.
Figure 5. Cost of the prediction of the time for the wetting front to reach a depth of 200 m below land surface for alluvium texture A as a function of the number of measurements used in parameter estimation for a linear cost function for the best measurement sets across all soil textures. For 8 measurements the measurement cost is 40% of the a priori performance cost. A. Performance cost, measurement cost, and total cost (height of the stacked bars). B. Total cost determined using the most-likely (filled triangles) and probability-weighted (open triangles) approaches.

Figure 6. Total cost for linear cost function cf1 for texture A as a function of the total measurement cost for 8 measurements for prediction based on SCEM with uncertainty.
Figure 7. The optimal number of measurements for the lowest total cost as a function of the total measurement cost for 8 measurements. The overall best measurement set is the same for each texture.