A Software Framework for Inverse Modeling and Uncertainty Characterization and Groundwater Assessment Case Study

Carlos A. Osorio-Murillo^{1*}, Matthew Over², Heather Savoy², Daniel P. Ames¹, Yoram Rubin²

¹Brigham Young University, Department of Civil and Environmental Engineering, Provo, Utah

²University of California, Berkeley, Department of Civil and Environmental Engineering, Berkeley, California

*Corresponding author: carosocali@gmail.com

Abstract

Estimation of spatially random fields (SRFs) is required for predicting groundwater flow, subsurface contaminant movement, and other areas of environmental and earth sciences modeling. This paper presents an inverse modeling framework called MAD# for characterizing SRFs, which is an implementation of the Bayesian inverse modeling technique Method of Anchored Distributions (MAD). MAD# allows modelers to “wrap” simulation models using an extensible driver architecture that exposes model parameters to the inversion engine. MAD# is implemented in an open source software package with the goal of lowering the barrier to using inverse modeling in education, research, and resource management. MAD# includes an intentionally simple user interface for simulation configuration, external software integration, spatial domain and model output visualization, and evaluation of model convergence. Two test cases are presented demonstrating the novel functionality of this framework to apply inversion in order to calibrate the model parameters characterizing a groundwater aquifer.

Keywords: Modeling Frameworks, Method of Anchored Distributions, Inverse Modeling, Spatial Random Fields, Model Integration.

26 **Acronyms**

27 FM: Forward Model

28 FMD: Forward Model Driver

29 GIS: Geographic Information System

30 MAD: Method of Anchored Distributions

31 MAD#: Method of Anchored Distributions C# Program

32 pdf: Probability Density Function

33 RFG: Random Field Generator

34 RFGD: Random Field Generator Driver

35 SRF: Spatial Random Field

36

37

38

39

1. Introduction

1.1 Overview

Spatial phenomena variability is typically evaluated through analytical and numerical models that describe the general properties of spatial random fields (SRFs). These models employ parameters and observations to define spatial variability. The characteristics – and hence variability – of an SRF can be discerned by the relationship between model parameters, direct, and indirect information. A number of hydrogeological studies have been conducted using SRF analysis (Delhomme, 1979; Carrera & Neuman, 1986; Dagan, 1987; Bates & Townley, 1988; Bellin & Rubin, 1996; Yeh et al., 2002; Kanso et al., 2003; Gallagher & Doherty, 2007; Farmani et al., 2008). This paper introduces an open source inverse modeling framework, called MAD# (pronounced “mad sharp”), focused on the characterization of SRFs using the Method of Anchored Distributions (MAD), a Bayesian inverse modeling technique (Rubin et al. 2010).

The process of estimating model parameters from the inversion of governing equation(s) and observations is called inverse modeling. For over fifteen years, researchers have advocated for the development of flexible and easy-to-use inverse modeling tools, with the understanding that the shortage of such tools hinders the development of comprehensive and credible uncertainty quantification tools (Poeter & Hill, 1997; Poeter & Hill, 1999; Rubin, 2004; Dagan, 2011). Carrera et al. (2005) identified five features that are needed for broad adoption of inverse modeling tools in hydrogeology: 1) incorporating geological data, 2) improving the flexibility of the code and procedures to handle any and all relevant data types, 3) a complete quantification of uncertainty, 4) reducing the difficulty of code operation, and 5) coupling inverse modeling techniques with a geographic information system (GIS) platform.

A number of existing simulation model software tools include model parameter estimation and uncertainty characterization as embedded functions within the program. For example, WEAP

(Yates et al. 2005) and PMWIN (Chiang & Kinzelbach 2001) both are applications that use forward models (FMs) and model parameter estimation software applications like PEST (Doherty 1994). These and related software tools have aided adoption of uncertainty characterization and inverse modeling to some degree. However, we recognize a need for additional tools that provide a more general set of capabilities and that address the issues raised by Carrera et al. (2005).

MAD has been shown by Rubin et al. (2010), Murakami et al. (2011), and Chen et al. (2012) to be a flexible stochastic inverse modeling technique that addresses the first three challenges posed by Carrera et al. (2005). Specifically, MAD can account for geology (Challenge #1) via the representation of geological features through SRFs modeled using structural parameters; handles multiple relevant data types (Challenge #2) through use of direct measurements and measurements that are indirectly related to the variable modeled; and accommodates uncertainty (Challenge #3) by explicitly incorporating observation uncertainties and quantifying uncertainty of geostatistical structural parameters and a new concept called “anchors”.

1.2 Research Goals

We have endeavored to address Carrera’s Challenges #4 and #5 by implementing and testing MAD in an extensible, user-friendly software framework. Specific goals for the developed framework include:

- 1) It should be capable of generically accommodating FMs that relate target variables with observations.
- 2) It also should be flexible in supporting the use of other user-specified software packages for random field generators (RFGs).
- 3) It should be able to characterize the uncertainty associated with SRFs.
- 4) It should be well documented and transparent with independently verifiable results.

The remainder of this paper presents our approach to meeting the research goals noted above in the form of an open source inverse modeling software framework called MAD#. This new inverse modeling application builds upon a prototype architecture (Osorio et al. 2012), in which MAD was implemented as a HydroDesktop (Ames et al. 2012) plugin using an embedded steady-state head solver written in R statistical software. MAD# is a standalone desktop application and includes an architecture for adding custom random field generator drivers (RFGDs) and forward model drivers (FMDs) for incorporating new models. We present an architectural overview of MAD# and descriptions of drivers currently implemented. We also present a demonstration of MAD# in two synthetic pumping experiments using a MODFLOW (Harbaugh 1996) project created in the PMWIN MODFLOW interface (Chiang & Kinzelbach 2001).

The work presented here is related to an active area of research and development within the broader context of integrated environmental modeling in that our software framework is indeed a method for “integrating” different modeling software packages into a single cohesive environment. This approach is related to the approach supported by as OpenMI (Castronova et al 2013, Knapen et al. 2013). Ridler et al (2014) follow a strikingly similar approach to developing a data assimilation framework using OpenMI and an open data assimilation library, using the C# programming language. Another model integration framework that is rapidly growing in adoption is the Community Surface Dynamics Modeling System (CSDMS) which uses a “wrapper-style” common modeling interface approach which is similar in nature and purpose to the forward model driver approach we present herein (Overeem et al 2013, Castronova and Goodall 2010).

It is worth noting that the OpenMI and CSDMS approaches both presume the existence of software packages that implement particular hydrologic or environmental numerical models. In other words, these models exist as software packages that require specific input and output files

– not simply as conceptual mathematical models. A somewhat different approach has been taken by integrated modeling efforts such as The Object Modeling System (OMS) which is suited to integrating small functions or codes that represent individual physical processes rather than integrating large software packages (David et al 2013).

In comparison to the OpenMI and CSDMS integrated modeling systems, our MAD# approach is relatively simple. Rather than facilitating the transmission of inputs and outputs between various numerical model packages, our approach focuses on a tightly managed system of a single model package connected through a single wrapper (or “driver”) directly to our inversion software. This is much more manageable than the alternative of linking multiple models to multiple models, and helps us avoid the challenges and issues of the so-called “Integronsters” (Voinov and Shugart 2013).

2.0 Methods

2.1 MAD Theoretical Background

Although a complete description of MAD is outside the scope of this paper, a brief introduction to the method is presented here. MAD is a Bayesian inverse modeling technique focused on characterizing SRFs by using Bayes’ theorem and the following concepts intended to address the challenges stated in the previous section:

- Geostatistical models are used to capture large-scale trends and reproduce patterns of spatial variability in terms of SRFs.
- Data classification – MAD classifies data (measurements) in a general format that is not limited (or specific to) any particular discipline or application. MAD categorizes data as:
 - o Type A data, $\mathbf{z}_a = \mathbf{y}(\mathbf{x}_i) + \varepsilon_a$, $i=1, \dots, N$, which could include direct measurements (including measurement error ε) of the target variables (e.g. hydraulic

- conductivity) at location x_i , $i=1,...,N$, or other types of measurements (e.g. transmissivity) at x_i that could be directly related to the target variable at x_i ,
- o Type B data, $\mathbf{z}_b=M(\mathbf{x}_i) + \varepsilon_b$, $j=1,...,M$, which include all measurements (including measurement error ε) that do not conform with Type-A, but are related to the target variable via a forward model, M (e.g. pressure head)
- Localization through anchored distributions (or “anchors”). An anchor is a statistical distribution of a target variable at a given location. Anchors can be employed for multiple target variables and/or locations. Anchors intend to capture local effects in the field of the target variables by conditioning realizations on fields.

MAD defines a target variable as a SRF, which is represented by a vector of geostatistical structural parameters (θ) capturing the global tendency, and anchors (ϑ) for quantifying local variations of the parameter field. MAD relies on the following proportionality (Rubin et al. 2010).

$$p(\theta, \vartheta | z_a, z_b) \propto p(\theta, \vartheta | z_a) p(z_b | \theta, \vartheta, z_a) \quad [1]$$

Where p indicates a probability density function (pdf) and $p(\theta, \vartheta | z_a)$ is the joint prior distribution of the structural parameters and anchors conditional on Type-A data vector $\mathbf{z}_a$, and $p(z_b | \theta, \vartheta, z_a)$ is the likelihood of observing the Type-B data vector $\mathbf{z}_b$ given the structural parameters, anchors and Type-A data. Finally, $p(\theta, \vartheta | z_a, z_b)$ is the joint posterior distribution of the structural parameters and anchors conditional on both Type-A and Type-B data.

2.2 MAD Methodological Approach

MAD is applied in three stages: 1) Strategy, 2) Implementation, and 3) Assessment. These three stages are described in Figure 1 and are discussed in depth in the following three subsections.

Figure 1. Structure of the MAD process.

2.2.1 Strategy

The first stage of applying MAD to a case study is the formulation of a strategy. This strategy entails the following six elements: (1) identifying target variables, (2) selecting appropriate priors for the SRF parameters, (3) identifying types of data available, (4) selecting numerical modeling strategy, (5) selecting locations for anchors, and (6) planning post-calibration model testing. A target variable is typically a heterogeneous variable that needs to be characterized. Measurements of target variables or directly related variables are classified as Type-A data. A SRF model type to describe a target variable's spatial variability in the field needs to be chosen, and this choice can be made from analyzing the measurements and previous literature. The second consideration is choosing appropriate priors for the parameters of the SRF model types chosen for the target variables. After these first two steps, the target variable is set up for MAD.

The third step focuses on identifying the data that could be used for the inversion process. Since the inversion requires indirect data, measurements need to be taken of a variable that the target variable influences via a mathematical model. These measurements are classified as Type-B data. This inversion data type generally describes larger-scale phenomena than the target variable and thus the combination of the two better informs the inversion process.

The fourth step is creating a numerical model. The numerical model needs to take the target variable, or a variable directly related, as an input and produce the inversion data type as an

output. Any relevant environmental influences from the site (e.g. wells or streams) need to be numerically represented. Other considerations that are necessary for building a well-posed model are also required, including dimensionality, boundary conditions, and time dependence. The Type-A and Type-B data need to be collected from within the domain selected and adequately distant from the boundaries to prevent interference.

The fifth step is placing the anchors in locations that are influential in the environmental process being modeled. Optimal selection of anchor locations was discussed by Yang et al. (2012). Each anchor also requires a prior distribution defined by prior knowledge.

Finally, one must choose an evaluation method for assessing the success of the inversion approach and testing calibration. Multiple inversions can be performed and cross-validation can be used to determine which approach is most successful. At this point, the strategy is developed and MAD can be implemented.

2.2.2 Implementation

The four steps of the implementation stage are (1) sampling from prior distributions, (2) creating realizations, (3) executing numerical model simulations, and (4) extracting results.

The first two steps cover the sampling strategy. In the first step, each SRF model parameter and anchor will need to have its prior distribution sampled, creating θ_i and ϑ_i , i in $1, \dots, N$ where N is the number of samples. The number of samples needs to be high enough to cover the parameter space, but choosing the number is not an exact science. Evaluating if there are not enough samples is covered in the assessment stage. In the second step, the samples from the previous step are used to create realizations of the target variable field. Each θ_i will define a SRF model to create realizations conditional on ϑ_i and the Type-A data.

The third step in the implementation stage is to run the simulations. For each realization created, the numerical FM is applied and a simulation is created. In the fourth step, the relevant

Type-B simulated values are extracted and used to calculate likelihoods and posterior distributions. The extracted Type-B values are the z_b vector for which likelihoods are calculated. The method for calculating likelihood is not specific to MAD such that any applicable to the z_b vector and Type-B measurements may be used. The likelihood method is applied to each sample, yielding a likelihood distribution across the sample space. This is multiplied by the prior distributions resulting in the posterior distributions and concluding the implementation stage.

2.2.3 Assessment

The assessment stage of the MAD approach focuses on assessing: (1) convergence and (2) general strategy. Convergence is assessed in two ways: if there were enough realizations and if there were enough samples. Generally, enough realizations are needed to estimate accurate likelihood values and enough samples are needed to resolve accurate posterior distributions. Graphical examples are given in section 3.0 on how to assess convergence.

The second kind of assessment, assessing general strategy, is more open-ended. The success evaluating technique chosen in the strategy stage is applied here. If the chosen success criteria are not met, then the strategy can be modified in several ways, including: increasing the total number of measurements if possible, changing the SRF model type or which parameters are random, or modifying the FM. If the success criteria are indeed met, it is still advisable to compare different parameters of the SRF model (e.g. covariance functions), or FMs to address model uncertainty. With this evaluation and acceptance of success, MAD has been thoroughly applied.

2.3 MAD# Software Framework

This section describes the design and development of an open source software framework that implements the MAD methodological approach described above. The software, MAD#, is

designed as an extensible architecture that uses generic functions for sharing information, executing processes and extracting data of FMs and RFGs, which can be stand-alone software applications, packages of a statistical frameworks, or libraries. To support these diverse applications, MAD# framework uses a driver approach to connect the FMs and RFGs through forward model drivers (FMDs) and random field generator drivers (RFGDs) respectively with the framework. These drivers are libraries that implement a set of interfaces of the *MadInterfaces* library (Figure 2).

Figure 2 General MAD# framework architecture.

This driver-based approach is expected to facilitate adoption and extension of the system by 3rd parties who can create new drivers for supporting new FMs or RFGs. MAD# is programmed using the .NET framework and the open source DotSpatial GIS programming library (www.dotspatial.org) following a similar approach as used in Ames et al. (2012). An open source approach was chosen to support transparency and adaptability of the software (Alexandrov et al. 2011). The DotSpatial library provides the geographic and display functionalities of MAD#, including data management, control, projection, symbology, and

extension management. Statistical libraries including Math.NET (Math.NET, 2014) and ALGLIB (Bochkanov, 2014) are also used in the framework. A customized version of MapWindow 6 (Dunsford & Ames 2011) serves as the final user interface of MAD#

2.3.1 Data structure

The MAD# data structure is based on three entity-relationship models (Figure 3) implemented using open source database SQLite. Three SQLite file databases “XMAD”, “XRESULT” and “XDATA” are created in different stages of the MAD# process.

The XMAD file stores the information generated in pre-processing module. The main tables of the XMAD file structural are explained. The *variable* table contains the list of variables provided by the FMD. These variables are used as key words during whole MAD# process. The *Domain* table stores the geographic information of the FM domain. The *Measure* table manages the Type-A and Type-B data and anchors. The prior information of the structural parameters and anchor are stored in the *PriorData* table. The z_b vector is stored in the *SelectionValues* table. This table is used to link the information generated by the simulation of the FM. In the processing module is generated a XDATA file database for each sample. This structure was selected due to size of the files generated by large simulation projects. These files store the simulated Type-B per realization in the *ResultSelection* table, which is related to the *SelectionValues* table matching each output with the z_b vector. Also, the processing module creates a XRESULT file database that contains a copy of all tables of the XMAD file and the parameter used by the user to execute the simulation in the *ConfigurationResult* table. The XRESULT file is used in the post-processing module to store the likelihood calculation in the *LikelihoodGroupValue* table.

Figure 3 MAD# database structure

2.3.2. Drivers

The FMDs are simple model wrappers that can be developed by 3rd parties to enable use of specific models with MAD#. The FMDs also expose the list of inversion target variables, domain and temporal types. MAD# supports, by default, two forward models: PMWIN – MODFLOW 96 (Chiang & Kinzelbach 2001), and HYDRUS 1D (Šimůnek et al. 1998).

A second type of MAD# driver is intended to support external programs capable of generating random spatial data fields (e.g. using geostatistics). RFGDs are used to define structural parameters (e.g. for a geostatistical model); establish random and deterministic structural parameters; and generate conditional fields. The RFGDs implemented in MAD# are: GSTAT – Based on GSTAT R package and stand-alone (Pebesma & Wesseling 1998), R_Base_Package

278 – Based on Mvtnorm (Genz & Bretz 2009), Msm (Jackson 2011), and Tmvtnorm (Wilhelmm &
279 Manjunath 2013) R packages.

280 Although the likelihood calculation is not a driver, in future releases, the calculation of the
281 likelihood will managed as a driver. This approach will allow to couple different method to
282 calculate the likelihood can be added in the framework. In the current version, the likelihood is
283 calculated using a nonparametric kernel method through the R statistical package *NP* (Hayfield
284 & Racine 2008).

285

286 **2.3.3 Implementation of MAD stages**

287 The strategy stage is addressed in the pre-processing module (Figure 4). This is an input
288 module where the user selects a FMD and RFGD. The selected FMD obtains information about
289 the geographical domain of the FM and the temporal nature of project (steady-state or
290 transient). Geographic domain information is used to reproduce the same domain of the FM. A
291 list of available variables classified in Type-A and Type-B are through FMD. These variables
292 allow users to identify the target variable and type of data available. The target variable is
293 considered a SRF, and defined by SRF parameters through RFGD. The selected SRF
294 parameters are managed as inversion parameters. The module contains tools for introducing
295 the location of observations, and anchors which are also inversion parameters. These
296 parameters should be associated to prior distributions, which can be generated by the MAD#
297 framework or imported by the user.

Figure 4 MAD# pre-processing module flow chart.

The implementation stage is handled by the processing module (Figure 5) which is the core of the MAD# framework. The user defines the number of samples and realizations per samples to be executed. Using each sample, MAD# requests from the RFGD a number of realizations (defined by the user). The RFGD returns the realizations of the sample. MAD# processes the realizations of each sample in the FM using the FMD. The FMD extracts the simulated data at the location of each Type-B data measurement. Finally, the processing module generates a result file, which is a database file with all parameters used in the process, and output files with the information extracted from multiple executions of the FM.

Figure 5 MAD# processing module

The post-processing module comprises the assessment stage (Figure 6). This module begins by assembling the simulated Type-B of the processing module by sample. The MAD# user can define different subsets of the Type-B data vector to evaluate the likelihood. The simulated Type-B and the Type-B subset are compared to calculate the likelihood per sample.

Figure 6 MAD# post-processing flow chart.

The convergence of the likelihood of each sample is evaluated using a graphical tool, which uses the likelihood of a sample with different amount of realizations. The evaluation of convergence of number of samples is done comparing the likelihood of all samples with different number of realizations. When the number of realizations or the number of samples is insufficient, it is necessary to add more realizations or samples. This process is executed again until the likelihood converges adequately. Using the likelihood of all samples is used to generate a posterior of the structural parameters and anchors.

3.0 Example Test Cases

A base synthetic case with lateral confined groundwater flow through a heterogeneous aquifer was used for creating four test cases. These heterogeneous aquifers are of interest because in reality aquifers are heterogeneous and this heterogeneity can affect the travel time of contaminants to as sensitive areas like drinking water wells. A synthetic $\ln(\text{Transmissivity})$ field was created using GSTAT (Pebesma & Wesseling 1998) with isotropic exponential covariance and no trend, which is common on synthetic projects (Li et. al, 2005). The SRF uses the following structural parameters: mean = -2, variance = 0.15, range (length scale) = 28 meters, and nugget = zero meters. The field is 400x400 meters discretized into a 40x40 uniform rectangular grid. From this baseline field, the Type-A data were collected and the anchor values are known. Also, the specific storage was constant of 0.001. The aquifer was evaluated as steady and transient state. Three periods of 3, 10, and 15 days were defined in the transient state. A well was place at center of the domain, pumping only in the second period. The boundary conditions in all test cases were: constant heads 105m and 100m at south and north respectively, and no lateral flow, generating a hydraulic gradient of 1.25%.

3.1 Case Studies: Strategy

The general strategy for this example is as outlined in Section 2.2.1 and starts with choosing the target variables in each test case (Table 1). The structural parameters are not assumed and will

be random variables in the inversion process. The prior distributions of all structural parameter were chosen to be uniform distributions in order to be conservative. The Type-A measurements in all test cases were taken from the three locations that form a triangle placed at the center of the domain (Figure 7). The inversion data type for these test cases is hydraulic head. Head provides ideal Type-B data since head gradients are a function of the hydraulic conductivity field, being this field numerically equals that transmissivity field managed in the test cases, where the thickness is one. The Type-B measurement locations are along the predominant flow path (south to north) for test case I. The test case II, the Type-B measurements are along of the flow direction generated by the pumping. The test cases III and IV use the same Type-B measurement configuration.

Table 1. Description of test cases

Test case	Type	Pumping	Target variable	Structural parameters	Priors bounds	Anchors	Type-B measurements
I	Steady	-	Transmissivity	Mean Partial Sill Range	[-5, -1] [0.1, 0.7] [10, 120]	8	4
II	Transient	5 m ³ /d	Transmissivity	Mean Partial Sill Range	[-5, -1] [0.1, 0.7] [10, 120]	8	12
III	Steady	-	Transmissivity	Mean Partial Sill Range	[-5, -1] [0.1, 0.7] [10, 80]	0	11
IV	Transient	5 m ³ /d	Transmissivity Specific storage	Mean Mean	[-5,-1] [0.0005, 0.002]	0	11

Figure 7 MAD# software showing the study area, Type-B (magenta triangle) and Type-A measurements (blue square), well (green square), and anchors (black star). a) Test case I. Steady state, 3 structural parameters of transmissivity range, partial sill and mean. b) Test case II. Transient scenario with 3 structural parameters of transmissivity range, sill and mean a. c) Test case III. Steady state without anchors and 3 structural parameters of transmissivity range, sill and mean. d) Test case IV. Transient state without anchors, and 2 structural parameters, transmissivity mean and specific storage mean.

361 The FM used was MODFLOW-96 (Chiang & Kinzelbach 2001) since that FM software is well-
362 established in the groundwater community, is easily available which appeals to the MAD#
363 community resource objective, and already has a FMD written for it. The MODFLOW-96 project
364 used for the FM simulations in MAD# is based off of that used for creating the synthetic Type-B
365 data in all test cases. This eliminates the possibility of model uncertainty associated with the
366 FM, a condition which does not reflect reality, but this is an elementary example to show the
367 basic application of MAD#. The one disparate aspect of the FM used in MAD# compared to the
368 baseline FM is that the synthetic transmissivity field is not utilized in the former. The same
369 domain extent and discretization along with boundary and initial conditions were used.

370 The placement of anchors is in a diamond configuration settled between the Type-B data
371 locations. There are eight anchors, and the distances between them and a Type-B data location
372 varies. In the assessment stage, the influence of the distance between the anchors and Type-B
373 data will be analyzed. Strategic placement of anchors is a subject of current research, and is
374 addressed in Yang, Over & Rubin 2012.

375 The goal of this example is to compare the difference between the posterior distributions to the
376 respective prior distributions, for both over the parameter space and in relation to the true
377 values of the synthetic baseline case. Three structural parameters and anchors will be assessed
378 this way. The test case I and II will be compared in order to evaluate the effect of more Type-B
379 measurements and transient state in the inversion. The test cases III and IV are linked with the
380 objective to determine the posterior distribution of the structural parameter mean of
381 transmissivity and specific storage. The maximum likelihood value of the structural parameters
382 range and variance of the test case III were used in the test case IV.

383 All of the information provided in this subsection is entered into the pre-processing module of
384 MAD#, as discussed in Section 2.3.

3.2 Case Studies: Implementation

The implementation of the strategy is carried out in the MAD# processing module; the first step of sampling prior distributions was conducted in each test case using MAD#. To obtain samples for the structural parameters, the prior distributions were independently randomly sampled using Math.NET (Math.NET, 2014) library included in MAD#. The number of unique samples for each structural parameter was 100. To create samples for the anchor location, first a unique sample from the structural parameter set was used to generate conditioned fields of the Type-A data. The conditioned fields at the anchor locations were then used to obtain a vector of values that describe normal distribution. In total, with the 100 unique structural parameter samples each having 10 anchor samples, there were 1000 prior samples.

Each of the 1000 prior samples had 300 realizations generated with the GSTAT as the RFG. Each of these realizations were passed to MODFLOW-96 and the FM was executed a total of 300.000. Since the forward simulations are independent, MAD fits into the ‘embarrassingly parallel’ category of parallel algorithms. The high-throughput computing resource HTCondor was utilized to distribute the forward simulations a computer lab. The total processing time for the simulations is presented in the Table 2. HTCondor controls the number of instances in each simulation process. The computer lab has computers with different CPUs from Intel I7 - 8 cores, 8 GBytes of memory to Intel I5 – 4 GBytes of memory.

Table 2. Description of test cases

Test case	Physical hours	CPU hours	Number of computers
I	3.83	230.1	51
II	6.37	282.9	46
III	1.67	88.4	22
IV	1.40	124.8	30

After the simulations were run and MAD# extracted the simulated Type-B results from the Type-B measurement locations, the likelihood distributions were calculated. The likelihood calculation

method utilized was non-parametric kernel density estimation (Hayfield & Racine 2008). The advantage of the non-parametric method is that there is no assumption on the shape of the distribution, but the disadvantage is that more realizations are needed to resolve the shape compared to parametric methods.

3.3 Case Studies: Assessment

Within the MAD assessment stage, if a sample has consistent likelihood over increasing numbers of realizations, and has the same relative likelihood when compared to other samples, then its likelihood is converged. Figure 8 shows one convergence plot for each of the four test cases. Seven random samples are chosen, and their likelihoods over a range of realizations are calculated. After 260 realizations, all but one sample out of all test cases holds a consistent likelihood value. This suggests that 300 realizations are enough to proceed to calculating posterior distributions.

By multiplying the converged likelihood distributions by the prior distributions, the posterior distributions are calculated. Figures 9-12 compare posterior distributions to their respective prior distributions and the true values. There are two ways to determine if a posterior is an improvement over a prior. Primarily, the posterior should cover a narrower range of values in the parameter space which concentrates the probability density. Second, the posterior can increase the probability of the true value, however this is only applicable in synthetic case studies.

Figure 8. Evaluation likelihood convergence. a) Test case I. b) Test case II. c) Test case III. d) Test case IV.

Comparing the four test cases, some insight can be gained on the influence of Type-B data on posteriors. Test case II – which has more Type-B measurements in both space and time compared to test case I – had posteriors with narrower ranges and higher probabilities for the true values for all but one anchor (Figure 9). The same success can be seen for the partial sill (Figure 10). The mean's posterior is narrower in test case II, but the true value's probability is approximately the same as in the prior, while test case I had a high probability for the true value. For the range structural parameter, the posterior had a narrower range, but the true value does not fall within it. However, in test case III, which has more Type-B variables in space than test case I, the range structural parameter has a posterior with the probability density concentrated closer to the true value. Estimation of the range structural parameter shows a large uncertainty with respect the true value; similar results were found by Firmani et al 2006.

Figure 9. Prior and posterior pdf comparison test cases I and II (red line prior, cyan line posterior test case I, green line posterior test case II, and vertical line true value)

444 The maximum likelihood (ML) value of the structural parameters range and partial sill of the test
 445 case III (Figure 11) were used in the test case IV. The ML value of the range 14.33 m and
 446 partial sill 0.17 were considered constant values in the test case IV. The posterior pdf in test
 447 case IV (Figure 12), where there are transient Type-B measurements but one less than test
 448 case II, the mean has the best posterior out of all the test cases and the storage coefficient's
 449 mean also had a successful posterior.

Figure 10. Prior and posterior pdf comparison of structural parameters between test case I and II. a) Structural parameters test case I. b) Structural parameters test case II.

Figure 11. Prior and posterior pdf of structural parameters of test case III. a) Mean b) Partial Sill c) Range

Figure 12. Prior and posterior pdf of structural parameters of test case IV. a) Structural parameter mean of the transmissivity. b) Structural parameter mean of the primary storage coefficient.

4.0 Discussion and Conclusions

The MAD approach and the MAD# software differ significantly from other common inverse modeling methods and tools. PEST (Doherty 1994), UCODE (Poeter & Hill 1999) and ITOUGH2

(Finsterle & Zhag 2011) have as a primary purpose, deterministic parameter estimation through minimizing linear or nonlinear objective functions and determining a single value per parameter. Conversely, MAD# uses a Bayesian method for transferring information from observations to anchors and structural parameters to obtain posterior distributions for each parameter.

Inverse modeling applications generally require configuring and changing input and output files, executing a forward model, and evaluating results. While methods such as Joint Universal Parameter Identification and Evaluation of Reliability (JUPITER) (Banta et al. 2008) have been developed to reduce the complexity of parameter estimation, users are required to alter input files of the FM to fit within the specific parameter estimation framework. MAD# implements a new approach, in which users are not required to create template files. Using a plug and play approach, the user just specifies the FMD for executing the inverse process. The expectation is that this approach will increase the ease of using new FMs and simplify the application of inverse modeling techniques.

The generic configuration of the MAD# framework allows accommodation of different FMs via drivers. This paper demonstrated how a FM can be linked with the MAD# framework. The generation of realizations in MAD# is also managed via a driver. The MAD# user can select the appropriate generator for each project. The generated information in the inversion process is stored in simple SQLite databases, which can be accessed by generic SQLite manager applications and integrated with the GIS environment. The inversion process required a long processing time, which suggests that MAD# team should work in mechanism to execute in High Performance Computing frameworks.

In the case studies presented here, the PMWIN-MODFLOW driver manages the MODFLOW project generated by PMWIN without changing information in the original model control files. The SRF Ln(Transmissivity) is characterized using the indirect measurements of pressure head

via solving head equation implemented in MODFLOW. The anchors defined in the SRF domain show the posterior distribution of Ln(Transmissivity) and the uncertainty at the anchor locations. The posterior pdfs of structural parameters and anchors characterized the global and Ln(Transmissivity) field, respectively. The posterior pdfs obtained with more Type-B measurements produced values closer to true values. MAD# provides a user interface that allows comparison of multiple scenarios. The range structural parameter

In summary, the MAD# software framework was designed, developed, and described here to aid scientists, modelers, and students in the application of inverse modeling and SRF characterization. MAD# specifically addresses the five criteria proposed by (Carrera et al. 2005) and represents a potentially valuable step forward for inverse modeling in general and the MAD method specifically.

Software Availability

MAD# is made available through collaboration with the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) Hydrologic Data Center. MAD# source code and documentation can be accessed at the MAD code repository website <http://mad.codeplex.com>. MAD# and its source code are released under the New Berkeley Software Distribution (BSD) License which allows for liberal reuse of the software and code.

Acknowledgements

This work was supported by the National Science Foundation under grant EAR-1011336, "The Method of Anchored Distributions (MAD): Principles and Implementation as a Community Resource." Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This research used resources of the National Energy Research Scientific

Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We also acknowledge the extensive programming, testing, and documentation work of Daniel Gunnel, and others.

References

- Alexandrov, G, Ames, D, Bellocchi, G, Bruen, M, Crout, N, Erechtkhoukova, M, Hildebrandt, A, Hoffman, F, Jackisch, C & Khaite, P 2011, 'Technical assessment and evaluation of environmental models and software: Letter to the Editor', *Environmental Modelling & Software*, vol 26, pp. 328-336.
- Bochkanov, S 2014, 'ALGLIB' [Computer program], Available at www.alglib.net (Accessed 03 September 2014)
- Ames, DP, Horsburgh, JS, Cao, Y, Kadlec, J, Whiteaker, T & Valentine, D 2012, 'HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis', *Environmental Modelling & Software*, vol 37, pp. 146-156.
- Banta, ER, Hill, MC, Poeter, E, Doherty, JE & Banbendreier, J 2008, 'Building model analysis applications with the Joint Universal Parameter Identification and Evaluation of Reliability (JUPITER) API', *Computers & Geosciences*, vol 34, pp. 310-319.
- Bates, BC & Townley, LR 1988, 'Nonlinear, discrete flood event models, 1. Bayesian estimation of parameters', *J. Hydrol.* 99, pp. 61-76, doi: 10.1016/0022-1694(88)90078-9.
- Bellin, A & Yoram R 1996. 'HYDRO_GEN: A spatially distributed random field generator for correlated properties.' *Stochastic Hydrology and Hydraulics*, vol 10. no. 4, pp 253-278.
- Carrera, J, Alcolea, A, Medina, A, Hidalgo, J & Slooten, LJ 2005, 'Inverse Problem in hydrogeology', *Springer-Verlag*, vol 13, pp. 206-222.
- Carrera, J & Neuman, SP 1986, 'Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information', *Water Resour. Res.*, vol 22, no. 2, pp. 199-210, doi: 10.1029/WR022i002p00199.
- Castronova, A. M., Goodall, J. L., & Ercan, M. B 2013, 'Integrated modeling within a hydrologic information system: an OpenMI based approach', *Environmental Modelling & Software*, 39, 263-273.
- Castronova, A. M., & Goodall, J. L 2010, 'A generic approach for developing process-level hydrologic modeling components', *Environmental Modelling & Software*, 25(7), 819-825.
- Chen, XM, Hahn, MS, Hammond, GE, Rockhold, ML, Zachara, JM & Rubin, Y 2012, 'Three-dimensional Bayesian geostatistical aquifer characterization at the Hanford 300 Area using tracer test data', *Water Resour. Res.*, vol 48, no. W06501, doi: 10.1029/2011WR010675.
- Chiang, W-H & Kinzelbach, W 2001, *3D-Groundwater Modeling With Pmwin: A Simulation System for Modeling Groundwater Flow and Pollution*, Springer Verlag.

544 Dagan, G 1987, 'Theory of solute transport by groundwater', *Ann. Rev. Fluid Mech.*, vol 19, pp.
545 183-215.

546 Dagan, G 2011, 'An overview of stochastic modeling of groundwater flow and transport: From
547 theory to applications', *Eos, Transactions American Geophysical Union*, vol 83, no. 53,
548 pp. 621-625.

549 David, O., Ascough II, J. C., Lloyd, W., Green, T. R., Rojas, K. W., Leavesley, G. H., & Ahuja, L.
550 R. (2013). A software engineering perspective on environmental modeling framework
551 design: The Object Modeling System. *Environmental Modelling & Software*, 39, 201-213.

552 Delhomme, JP 1979. "Spatial variability and uncertainty in groundwater flow parameters: A
553 geostatistical approach." *Water Resour. Res.*, vol 15, pp. 269-280.

554 Doherty, J 1994, 'PEST: a unique computer program for model-independent parameter
555 optimisation.', *Water Down Under 94: Groundwater/Surface Hydrology Common Interest
556 Papers*, p. 551.

557 Dunsford, H & Ames, DP 2011, 'MapWindow 6.0: an extensible architecture for cartographic
558 symbology', *OSGeo Journal*, vol 8, pp. 31-36.

559 Farmani, MB, Kitterød, N-O & Keers, H 2008, 'Inverse modeling of unsaturated flow parameters
560 using dynamic geological structure conditioned by GPR tomography', *Water Resour.
561 Res.*, vol 44, no. W08401.

562 Finsterle, S & Zhag, Y 2011, 'Solving iTOUGH2 simulation and optimization problems using the
563 PEST protocol', *Environmental Modelling & Software*, vol 26, pp. 959-968.

564 Firmani, G, Fiori, A & Bellin A 2006, 'Three-dimensional numerical analysis of steady state
565 pumping test in heterogeneous confined aquifers', *Water Resour. Res.*, vol 42, W03422,
566 doi:10.1029/2005WR004382.

567 Gallagher, M & Doherty, J 2007, 'Parameter estimation and uncertainty analysis for a watershed
568 model', *Environmental Modelling & Software*, vol 22, no. 7, pp. 1000-1020.

569 Genz, A & Bretz, F 2009, *Computation of Multivariate Normal and t Probabilities*, Springer-
570 Verlag, Heidelberg.

571 Harbaugh, AW 1996, *Programmer's documentation for MODFLOW 96, an update to the US
572 Geological Survey Modular Finite-Difference Ground Water Flow Model+ MODFLOW
573 2000.*, USGS, Washington D.C.

574 Hayfield, T & Racine, JS 2008, 'Nonparametric Econometrics: The np Package', *Journal of
575 Statistical Software*, vol 27, no. 5.

576 Jackson, C 2011, 'Multi-State Models for Panel Data: The msm Package for R', *Journal of
577 Statistical Software*, vol 38, no. 8, pp. 1-29.

578 Kanso, A, Gromaire, MC, Gaume, E, Tassin, B & G., C 2003, 'Bayesian approach for the
579 calibration of models: application to an urban stormwater pollution model', *Water
580 Science and Technology*, vol 47, no. 4, pp. 77-84.

581 Knapen, R., Janssen, S., Roosenschoon, O., Verweij, P., De Winter, W., Uiterwijk, M., & Wien,
582 J. E 2013, 'Evaluating OpenMI as a model integration platform across disciplines',
583 *Environmental Modelling & Software*, 39, 274-282.

584 Li, W, Nowak, W & Cirpka, O, A 2005, 'Geostatistical inverse modeling of transient pumping test
585 using temporal moments of drawdown', *Water Resour. Res.*, vol 41,
586 W08403, doi:10.1029/2004WR003874

Math.NET team (2014), '*Math.NET Numerics (version 2.2.0)*' [Computer program], Available at <https://mathnetnumerics.codeplex.com/releases/view/126119> (Accessed 03 September 2014)

Murakami, H, Chen, X, Hahn, MS, Liu, Y, Rockhold, ML, Vermeul, VR, Zachara, JM & Rubin, Y 2011, 'Bayesian approach for three-dimensional aquifer characterization at the Hanford 300 Area', *Hydrol. Earth Syst. Sci.*, vol 14, pp. 1989-2001, doi: 10.5194/hess-14-1989-2010.

Osorio, C, Over, M, Ames, DP & Rubin, Y 2012, An Extensible Inverse Modeling Software Architecture for Parameter Distribution Estimation, 2012 *International Congress on Environmental Modelling and Software*, <http://www.iemss.org/sites/iemss2012/proceedings/D6_1005_Osorio_et_al.pdf>.

Overeem, I., Berlin, M. M., & Syvitski, J. P 2013, 'Strategies for integrated modeling: The community surface dynamics modeling system example', *Environmental Modelling & Software*, 39, 314-321.

Pebesma, EJ & Wesseling, CG 1998, 'Gstat: a program for geostatistical modeling, prediction and simulation', *Computer & Geosciences*, vol 24, no. 1, pp. 17-31.

Poeter, EP & Hill, MC 1997, 'Inverse Methods: A necessary next step in groundwater modeling', *Ground Water*, vol 35, pp. 250-260.

Poeter, EP & Hill, MC 1999, 'UCODE, a computer code for universal inverse modeling', *Computers & Geosciences*, vol 25, pp. 457-462.

Ridler, M. E., van Velzen, N., Hummel, S., Sandholt, I., Falk, A. K., Heemink, A., & Madsen, H 2014, 'Data assimilation framework: Linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenMI)'. *Environmental Modelling & Software*, 57, 76-89.

Rubin, Y 2004, 'Stochastic hydrogeology – challenges and misconceptions', *Stochastic Environmental Research and Risk Assessment*, vol 18, pp. 280-281.

Rubin, Y, Chen, X, Murakami, H & Hahn, M 2010, 'A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields', *Water resources research*, vol 46, no. W10523.

Šimůnek, J, K., H, Šejna, M & van Genuchten, MT 1998, 'The Hydrus-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media Version 1.0', *International Ground Water Modeling Center*, p. 186.

Voinov, A., & Shugart, H. H 2013, 'Integronsters, integral and integrated modeling', *Environmental Modelling & Software*, 39, 149-158.

Wilhelmm, S & Manjunath, BG 2013, '*tmvtnorm*' [computer program]. Available at <http://cran.r-project.org/web/packages/tmvtnorm/index.html> (Accessed 03 September 2014).

Yang, Y, Over, M & Rubin, Y 2012, 'Strategic placement of localization devices (such as pilot points and anchors) in inverse modeling schemes', *Water Resources Research*, p. 48.

Yates, D, Sieber, J, Purkey, D & Huber-Lee, A 2005, 'WEAP21—A Demand-, Priority-, and Preference-Driven Water Planning Model', *Water International*, vol 30, no. 4, pp. 487-500.

629 Yeh, T-CJ, S., L, J., GR, Baker, K, Brainard, JR, D., A & D., L 2002, 'A geostatistically based
630 inverse model for electrical resistivity surveys and its applications to vadose zone
631 hydrology', *Water Resour. Res.*, vol 38, no. 1278.

632

633 **Appendix**

634

635 The spatial random function used in the example test case was the exponential covariance
636 function, given below:

$$C(h) = \sigma^2 \exp\left(-\frac{h}{a}\right)$$

637 also written as a variogram:

$$\gamma(h) = \sigma^2 \left(1 - \exp\left(-\frac{h}{a}\right)\right)$$

638 with σ^2 being the variance (also known as the partial sill) and a being the length scale (also
639 called range in the gstat documentation, but not to be confused with the term 'effective range'
640 which is used to refer to the distance at which the variogram approaches its maximum or 95% of
641 its asymptote). The other structural parameter mentioned in the example test case was the
642 mean, μ . This represents the field mean and is not seen in the covariance function because the
643 function represents how the values in the field fluctuate about the mean over different spatial
644 distances, h .

645