

TITLE: Effective and efficient methods to calibrate the Gridded Surface Subsurface Hydrologic Analysis hydrology model

AUTHORS: Brian E. Skahill and Charles W. Downer

DATE: August 19, 2013

PURPOSE: The purpose of this document is to describe how to use two similar functionalities that were recently developed to support the effective and efficient computer-based calibration of the Gridded Surface Subsurface Hydrologic Analysis (GSSHA) model.

INTRODUCTION: Spatially explicit physics-based models such as GSSHA (Downer and Ogden, 2003a, b) support a more realistic characterization of the physical aspects of the watershed system and a more transparent simulation and evaluation of project alternatives than is possible with traditional hydrologic simulation models (viz., lumped and semi-distributed model structures). And they have the potential to predict with greater reliability than lumped hydrologic model structures (Moore and Doherty, 2005). But, they also have the potential to easily become highly parameterized, particularly when they are deployed to simulate on a continuous basis heterogeneous watershed systems. Moreover, their model run times are often far greater than lumped and semi-distributed hydrologic models. It is this combination of computationally intensive forward model run times and the potential for a highly dimensional specified adjustable model parameter space which present a unique challenge for the computer-based calibration of spatially explicit physics-based hydrologic models. In particular, their combination necessitates the use of a calibration method that is as efficient as possible. Moreover, highly parameterized model deployments can make calibration problematical in that the information content encapsulated in the available observation dataset may not support the unique estimation for each of the specified adjustable model parameters, resulting in poor fits between the observations and their model simulated counterparts and/or non-physical models (i.e., estimated parameter sets). This article describes how to use two separate but closely related methods of computer-based parameter estimation either of which can serve as an effective and efficient means to support the practical calibration of a GSSHA hydrologic model. The two methods are adaptations to the “efficient local search” calibration methodology whose use as an alternate GSSHA run mode was originally described by Skahill and Downer (2012) and also more recently via use of the Watershed Modeling System (http://www.xmswiki.com/xms/WMS:GSSHA_Calibration and http://www.xmswiki.com/xms/WMS:WMS_Tutorials#Distributed_Hydrologic_Modeling_using_GSSHA). The Watershed Modeling System (WMS) is a graphically-based spatial data processing environment which supports the complete GSSHA model deployment process. The two methods whose use will be described herein are both highly efficient much in the same way as the “efficient local search” method is as described in Skahill and Downer (2012), Skahill et al. (2012), and Skahill et al. (2009). And they both make (highly parameterized) calibration contexts which may otherwise be problematical by virtue of parameter insensitivity effective by incorporating additional data into the estimation process which pertains directly to the specified adjustable model parameters. The extra “observations” imparted to effectively stabilize the model calibration process which pertain directly to the parameters requiring estimation are often referred to as prior information or regularization constraints.

DESCRIPTION OF USE: This section will succinctly describe how to use the two previously mentioned and closely related efficient GSSHA model calibration functionalities which employ prior information or regularization relationships. We will assume that one has already prepared the files necessary to calibrate a GSSHA model with the “efficient local search” method using the previously mentioned available guidance (Skahill and Downer, 2012; http://www.xmswiki.com/xms/WMS:GSSHA_Calibration; http://www.xmswiki.com/xms/WMS:WMS_Tutorials#Distributed_Hydrologic_Modeling_using_GSSHA). For each of the two methods, we will consider the same two example problems and describe in a step-by-step manner what modifications are required to a problem that has already been set up to use the “efficient local search” GSSHA alternate run mode calibration functionality. The control file named “cal_003.pst”, which is part of the [example](#) provided with the “efficient local search” functionality that is described on the GSSHA wiki (<http://gsshawiki.com>) will be the basis for the beginning for the two noted example problems. For each of the two methods, the two examples will present two different approaches to consider for specifying prior information or regularization relationships.

Method 1: Secant Levenberg Marquardt method with prior information

Example 1: This example employs the “efficient local search” method adapted to include prior information. The prior information is specified to effect a preferred homogeneity condition; viz., that all adjustable model parameters of the same type are specified to be the same value.

1. Copy the contents of the [example](#) problem to a new and empty directory on your local hard drive.
2. Make a copy of the control file named “cal_003.pst” and name the copy “cal_003_pi.pst”.
3. Add the following to the end of the control file named “cal_003_pi.pst”. In so doing, eight individual pieces of prior information for the eight specified adjustable model parameters now enforce the previously mentioned preferred homogeneity condition. The log of each adjustable model parameter is used in each prior information equation since each parameter is log transformed as indicated in the second entry on each of the eight rows which constitute the “parameter data” section of the control file. The eight pieces of prior information are uniformly weighted with a value of 1, and each of their individual contributions to now be added to the global quantitative measure of model to measurement misfit are aggregated into a single new observation group named “pinfo”.

* prior information

```

pi1 1.0 * log(rough1) - 1.0 * log(rough2) = 0.0 1 pinfo
pi2 1.0 * log(rough2) - 1.0 * log(rough3) = 0.0 1 pinfo
pi3 1.0 * log(rough3) - 1.0 * log(rough1) = 0.0 1 pinfo
pi4 1.0 * log(ksat1) - 1.0 * log(ksat2) = 0.0 1 pinfo
pi5 1.0 * log(ksat2) - 1.0 * log(ksat3) = 0.0 1 pinfo
pi6 1.0 * log(ksat3) - 1.0 * log(ksat4) = 0.0 1 pinfo
pi7 1.0 * log(ksat4) - 1.0 * log(ksat5) = 0.0 1 pinfo
pi8 1.0 * log(ksat5) - 1.0 * log(ksat1) = 0.0 1 pinfo

```

4. Modify the “observation groups” section of the control file to now also include the single prior information observation group named “pinfo” on a new line after the observation group named “tmf1”.
5. Modify the second line of the control data section of the control file to indicate that there are now **8** pieces of prior information and **2** observation groups as indicated directly below

```
* control data
restart estimation
  8  97  8  8  2
```

6. Save the control file
7. Run the GSSHA model by opening up a command window, typing the following at the prompt, and pressing enter:

```
gssha -slm cal_003_pi.pst
```

8. The computer-based GSSHA model calibration run for this example problem reduced the global quantitative measure of model to measurement misfit from an original value of approximately 34.84 to a final value of 3.347E-002 upon the completion of five optimization iterations (the maximum number of optimization iterations to be performed as specified in the control data section of the control file), which equated to twenty-eight forward GSSHA model calls. The final estimated values for the eight specified adjustable model parameters ((rough1, rough2, rough3, ksat1, ksat2, ksat3, ksat4, ksat5) = (0.100364, 0.095635, 0.096860, 0.105406, 0.132966, 0.144165, 0.150000, 0.113480)) and their computed confidence limits are listed in the record file named “cal_003_pi.rec”, and they together clearly underscore enforcement of the prior information. The adjustable model parameter named “ksat4” hit its lower bound as the method attempted to simultaneously enforce fit to the observation data and the preferred homogeneity condition encapsulated in the specified prior information. The same model was also calibrated using the original control file named “cal_003.pst” by typing “gssha -slm cal_003.pst” at the command prompt and pressing enter. The final estimated model for this GSSHA model calibration run is (rough1, rough2, rough3, ksat1, ksat2, ksat3, ksat4, ksat5) = (0.091778, 0.096802, 0.491869, 0.084113, 0.287828, 1.211571, 0.336561, 2.758958). And during execution of the model calibration run with the original control file named “cal_003.pst”, four of the eight specified adjustable model parameters were indicated to have no effect on observations (viz., “rough3”, “ksat3”, “ksat4”, and “ksat5”). Clearly, the GSSHA model calibration problem encapsulated in the original control file named “cal_003.pst” was in fact a good candidate for stabilization via the specification of prior information or regularization relationships.

Example 2: This example also employs the “efficient local search” method adapted to include prior information. The prior information equations in this example define preferred parameter values; viz., (rough1, rough2, rough3, ksat1, ksat2, ksat3, ksat4, ksat5) = (0.2, 0.3, 0.4, 0.1, 0.5, 1.5, 0.3, 0.2).

1. Copy the contents of the [example](#) problem to a new and empty directory on your local hard drive.

2. Make a copy of the control file named "cal_003.pst" and name the copy "cal_003_pi2.pst".
3. Add the following to the end of the control file named "cal_003_pi2.pst". In so doing, eight individual pieces of prior information for the eight specified adjustable model parameters now enforce the previously mentioned preferred parameter values. As with example 1, the log of each adjustable model parameter is used in each prior information equation since each parameter is log transformed as indicated in the second entry on each of the eight rows which constitute the "parameter data" section of the control file. The eight pieces of prior information are uniformly weighted with a value of 1, and each of their individual contributions to now be added to the global quantitative measure of model to measurement misfit are aggregated into a single new observation group named "pinfo".

```
* prior information
pi1 1.0 * log(rough1) = -0.69897000 1 pinfo
pi2 1.0 * log(rough2) = -0.52287875 1 pinfo
pi3 1.0 * log(rough3) = -0.39794001 1 pinfo
pi4 1.0 * log(ksat1) = -1.00000000 1 pinfo
pi5 1.0 * log(ksat2) = -0.30103000 1 pinfo
pi6 1.0 * log(ksat3) = 0.176091260 1 pinfo
pi7 1.0 * log(ksat4) = -0.52287875 1 pinfo
pi8 1.0 * log(ksat5) = -0.69897000 1 pinfo
```

4. Modify the "observation groups" section of the control file to now also include the single prior information observation group named "pinfo" on a new line after the observation group named "tmf1".
5. Modify the second line of the control data section of the control file to indicate that there are now **8** pieces of prior information and **2** observation groups as indicated directly below

```
* control data
restart estimation
  8  97  8  8  2
```

6. Save the control file
7. Run the GSSHA model by opening up a command window, typing the following at the prompt, and pressing enter:

```
gssha -slm cal_003_pi2.pst
```

8. The computer-based GSSHA model calibration run for this example problem reduced the global quantitative measure of model to measurement misfit from an original value of approximately 32.86 to a final value of 4.123E-001 upon the completion of five optimization iterations (the maximum number of optimization iterations to be performed as specified in the control data section of the control file), which equated to twenty-four forward GSSHA model calls. The final estimated values for the eight specified adjustable model parameters ((rough1, rough2, rough3, ksat1, ksat2, ksat3, ksat4, ksat5) = (0.092067, 0.287423, 0.418490, 0.071026, 0.505346, 1.431598, 0.307636, 0.354920)) and their computed confidence limits are listed in the record

file named "cal_003_pi2.rec". The weight assigned to each piece of prior information in this example was one, just as with the first example. Clearly, the same weighting strategy more easily enforced the prior information that was specified with the first example. Of course, stronger enforcement of the prior information via a different weights assignment strategy would yield a final estimated model most likely in closer agreement with the preferred model specified in the prior information, but possibly at a cost with respect to fit to the observed data.

Method 2: Secant Levenberg Marquardt version of the Tikhonov solution

This method differs from the first in that it has a regularization weight multiplier, β^2 , which permits the modeler to explicitly explore the tradeoff between fit to the observed data and fit to the regularization relationships. Just as with the first method with prior information, the regularization relationships are weighted; however, with this approach they are subsequently multiplied by a regularization weight factor, β^2 . Selection of an appropriate value for β^2 is critical. If its value is too high the parameter estimation process will ignore the measurement dataset in favor of fitting the regularization observations. If it is too small, the regularization observations will not endow the parameter estimation process with the numerical stability which it needs in order to obtain estimates for the parameters. Alternatively, the assignment of a value to β^2 can be viewed as a mechanism for trading parameter reasonableness against goodness of fit. There are several means by which to estimate an "optimal" fixed value for β^2 , but in the two examples below we will demonstrate one approach (Hansen, 2000).

Example 1: This is the same example 1 calibration problem previously considered with the first method.

1. Copy the contents of the [example](#) problem to a new and empty directory on your local hard drive.
2. Make a copy of the control file named "cal_003.pst" and name the copy "cal_003_tik_001.pst".
3. Add the prior information to the control file in exactly the same way as it was added above in step 3 of Example 1 with the first method.
4. Modify the observation groups section of the control file in exactly the same way as it was in step 4 of Example 1 with the first method.
5. Modify the second line of the control data section of the control file in exactly the same way as it was in step 5 of Example 1 with the first method.
6. On the first line of the control data section change "estimation" to "regularization" to activate use of the second method.
7. Add the following to the end of the control file, the key input being specification of the regularization weight factor which in this case is specified to be **1**

```
* regularisation
10.0 11.0 0.0
1.0e0 1.0e0 1.0000000001e0
1.3 1.0e-2 0
```

8. While it was not necessary, in this case for this method, in attempts to focus on the influence of the regularization relationships, during testing we set the de-facto stabilization device that is a

part of both methods effectively to zero (**1E-14**), and also modified the input control setting such that only **1** parameter upgrade was to be tested at each optimization iteration. These two changes are made on the fourth line of the control data section of the control file as indicated in bold below

0.00000000000001 2.0 0.3 0.03 **1**

9. And while it was also not necessary, the parameter data section of the control file was also modified such that (a) the initial parameter set was 0.1, and (b) the lower and upper bounds for each of the two parameter types were uniformly the same, as indicated below in the fourth and also the fifth and sixth columns, respectively

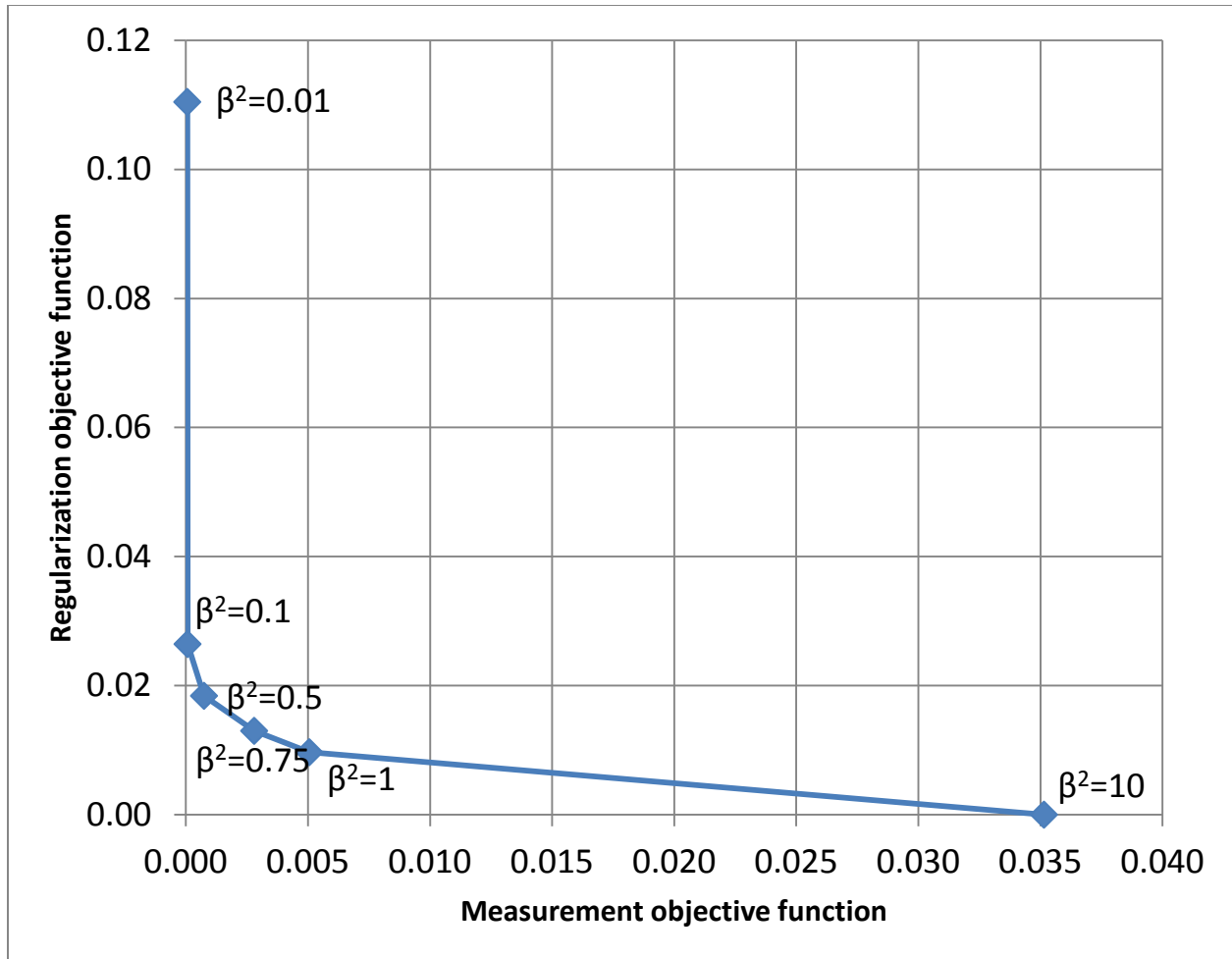
* parameter data

rough1	log	factor	0.1	1.00000E-02	5.00000E-01	rough1	1.000	0.000	1
rough2	log	factor	0.1	1.00000E-02	5.00000E-01	rough2	1.000	0.000	1
rough3	log	factor	0.1	1.00000E-02	5.00000E-01	rough3	1.000	0.000	1
ksat1	log	factor	0.1	6.60000E-03	2.99	ksat1	1.000	0.000	1
ksat2	log	factor	0.1	6.60000E-03	2.99	ksat2	1.000	0.000	1
ksat3	log	factor	0.1	6.60000E-03	2.99	ksat3	1.000	0.000	1
ksat4	log	factor	0.1	6.60000E-03	2.99	ksat4	1.000	0.000	1
ksat5	log	factor	0.1	6.60000E-03	2.99	ksat5	1.000	0.000	1

10. Save the control file
11. Run (Calibrate) the GSSHA model by opening up a command window, typing the following at the prompt, and pressing enter:

```
gssha -slm cal_003_tik_001.pst
```

12. The GSSHA model was also calibrated with this second method five additional times with five separate and unique values for the regularization weight factor; viz., 0.01, 0.1, 0.5, 0.75, and 10. The plot directly below presents the final estimated values which quantified model to measurement misfit with respect to the observations and also the regularization relationships for each of the six GSSHA model calibration runs. Clearly, the plot indicates the previously mentioned explicit tradeoff associated with varying the value for the regularization weight factor. The L-curve criterion would suggest that an optimal value to select for β^2 to calibrate this problem as otherwise designed is approximately 0.75 (Hansen, 2000).



Example 2: This is the same example 2 calibration problem previously considered with the first method.

1. Copy the contents of the [example](#) problem to a new and empty directory on your local hard drive.
2. Make a copy of the previously prepared control file named "cal_003_tik_001.pst", name the copy "cal_003_tik_101.pst", and place this file in the new directory created in the previous step.
3. Change the seventh line of the control data section of the control file to be the following, which modifies the termination criteria for the method

```
30 .001 4 4 .001 4
```

4. Change the parameter data section of the control file to the following

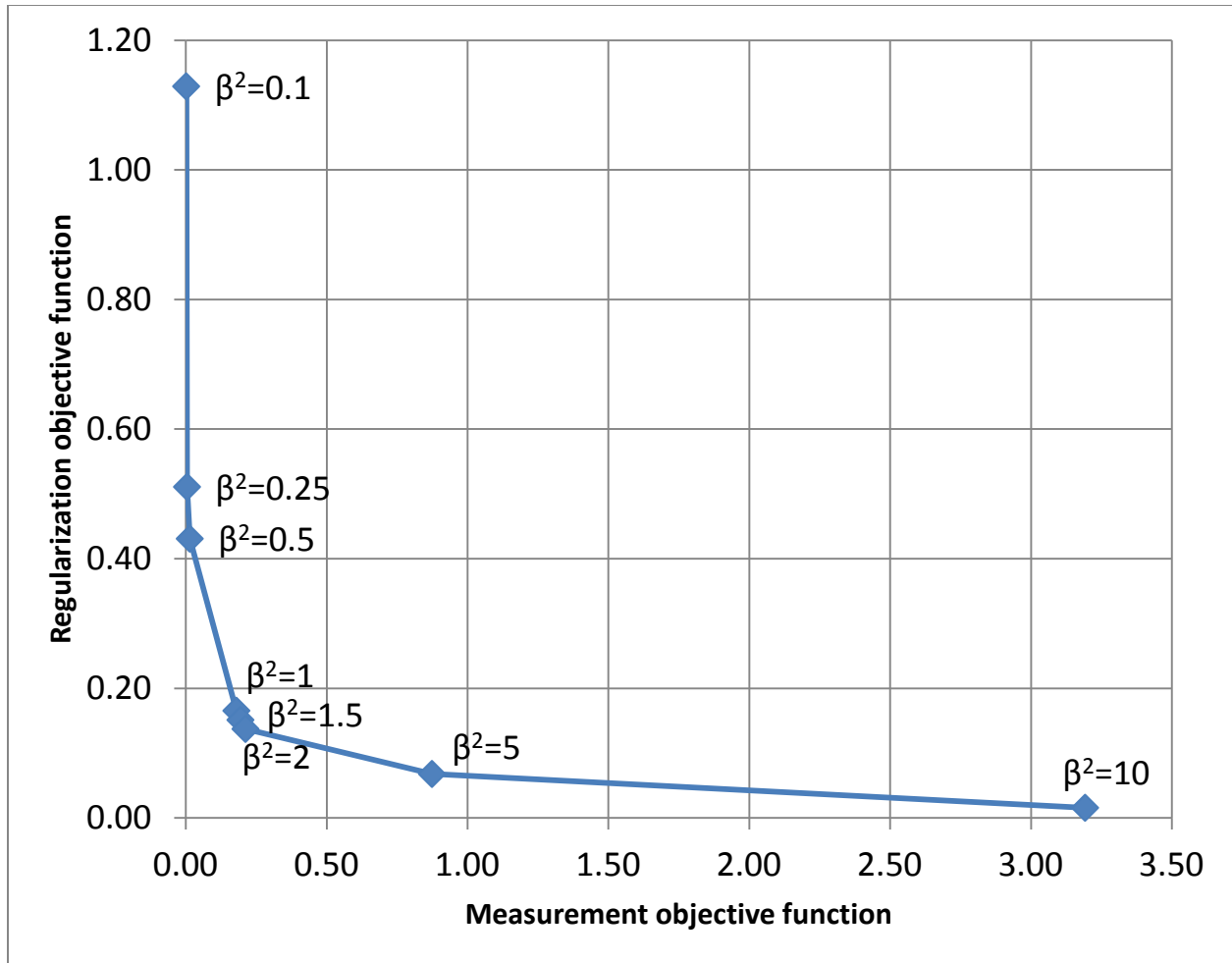
```
* parameter data
rough1 log factor 0.4 1.00000E-02 5.00000E-01 rough1 1.000 0.000 1
rough2 log factor 0.2 1.00000E-02 5.00000E-01 rough2 1.000 0.000 1
rough3 log factor 0.3 1.00000E-02 5.00000E-01 rough3 1.000 0.000 1
ksat1 log factor 0.5 1.50000E-02 0.61 ksat1 1.000 0.000 1
ksat2 log factor 0.1 1.50000E-02 0.61 ksat2 1.000 0.000 1
```

ksat3	log	factor	0.1	6.60000E-02	2.99	ksat3	1.000	0.000	1
ksat4	log	factor	0.4	1.50000E-01	0.61	ksat4	1.000	0.000	1
ksat5	log	factor	0.1	6.60000E-03	2.99	ksat5	1.000	0.000	1

5. Change the prior information section to be the exactly the same as that specified in item 3 for the second example problem considered with the first method
6. Save the control file
7. Run (Calibrate) the GSSHA model by opening up a command window, typing the following at the prompt, and pressing enter:

```
gssha -slm cal_003_tik_101.pst
```

8. The GSSHA model was also calibrated with this second method seven additional times with seven separate and unique values for the regularization weight factor; viz., 0.1, 0.25, 0.5, 1.5, 2, 5, and 10. The plot directly below presents the final estimated values which quantified model to measurement misfit with respect to the observations and also the regularization relationships for each of the eight GSSHA model calibration runs. Clearly, the plot below indicates the previously mentioned explicit tradeoff associated with varying the value for the regularization weight factor. The L-curve criterion would suggest that an optimal value to select for β^2 to calibrate this problem as otherwise designed is approximately 2 (Hansen, 2000).



DISCUSSION AND SUMMARY: This article described by way of example how to use two new closely related methods that can be employed to effectively and efficiently calibrate a GSSHA model. They are both adaptations to the previously reported upon “efficient local search” GSSHA model calibration methodology (Skahill and Downer, 2012, and references cited therein; http://www.xmswiki.com/xms/WMS:GSSHA_Calibration; http://www.xmswiki.com/xms/WMS:WMS_Tutorials#Distributed_Hydrologic_Modeling_using_GSSHA). The two new methods both achieve calibration run efficiency in the same way as the “efficient local search” method. And they both make what may otherwise be problematical calibration settings effective through the introduction of extra observations which pertain directly to the estimable parameter set. Either approach is a local search biased toward a specified preferred system state. And the degree to which the model is biased is influenced by the prior information weights assignment strategy, and in the case of the second method, also the value for the regularization weight multiplier which allows for the explicit evaluation of the tradeoff between fit to the observed data and fit to the regularization relationships. We are of the mindset that efficient methods which permit the specification of regularization constraints hold strong promise as practical means for the effective and efficient calibration of potentially computationally expensive and highly parameterized modeling contexts which can be associated with physics-based spatially explicit surface hydrology models such as GSSHA. And it is

for this reason that we recommend use of either of these two methods for the calibration of most practical applications of the GSSHA model.

Two separate preferred models were presented for each of the two methods in the example problems. We encourage the active reader to explore further by experimenting with the example problem datasets. Planned future work will focus on updating the Watershed Modeling System interface to support these two new recommended GSSHA model calibration functionalities.

ADDITIONAL INFORMATION: For additional information, contact Dr. Brian E. Skahill or Dr. Charles W. Downer, at Brian.E.Skahill@usace.army.mil and Charles.W.Downer@usace.army.mil, respectively. This CHETN should be cited as follows:

Skahill, B.E., and Downer, C.W. (2013). *Effective and efficient methods to calibrate the Gridded Surface Subsurface Hydrologic Analysis hydrology model. Coastal and Hydraulics Engineering Technical Note*, ERDC/CHL CHETN-AA-BB. Vicksburg, MS: U.S. Army Engineer Research and Development Center. An electronic copy of this CHETN is available from <http://chl.erdc.usace.army.mil/chetn>.

REFERENCES:

Downer, C.W., and F.L. Ogden. 2003a. *GSSHA user's manual: Gridded surface subsurface hydrologic analysis, version 1.43 for WMS 6.1*. Technical Report. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Downer, C. W., and F. L. Ogden. 2003b. Prediction of runoff and soil moistures at the watershed scale: Effects of model complexity and parameter assignment. *Water Resources Research*, 39(3): 1045.

Hansen, P.C. 2001. The L-Curve and its Use in the Numerical Treatment of Inverse Problems. in *Computational Inverse Problems in Electrocardiology*, ed. P. Johnston, Advances in Computational Bioengineering, WIT Press, ISBN-10: 1853126144.

Moore, C., and J. Doherty. 2005. Role of the calibration process in reducing model predictive error. *Water Resources Research*, 41, W05020, doi:10.1029/2004WR003501.

Skahill, B., Baggett, J., Frankenstein, S., and Downer, C.W., 2009. "More efficient PEST compatible model independent model calibration." *Environmental Modelling & Software* (24), 517–529.

Skahill, B.E., Downer, C.W., and Baggett, J.S., 2012. "A Practical Guide to Calibration of a GSSHA Hydrologic Model using ERDC Automated Model Calibration Software – Efficient Local Search." ERDC/CHL TR-12-3 Technical Report, U.S. Army Engineer Research and Development Center, Vicksburg, MS.

Skahill, B.E., and Downer, C.W. 2012. *New Gridded Surface Subsurface Hydrologic Analysis Computer-based Calibration Capabilities. Coastal and Hydraulics Engineering Technical Note*, ERDC/CHL CHETN-XI-16. Vicksburg, MS: U.S. Army Engineer Research and Development Center. An electronic copy of this CHETN is available from <http://chl.erdc.usace.army.mil/chetn>.