Quantifying and Communicating Uncertainty in Products of the USGS National Water Census

Project Report: Assistance Award Number G14AP00115

Kenneth W. Potter and Zachary Schuster
University of Wisconsin

January 14, 2015

Introduction

The United States Geological Survey (USGS) Water Census Project is an ambitious program that will provide unprecedented information on the availability, movement, and use of water in the United States. This information will make it possible for water managers and users at local, state, regional, and national scales to significantly improve the sustainability of the nation's water resources. To make best use of this information, it is critical that associated uncertainties be quantified and communicated in ways that serve the needs of the users as well as the USGS. This report provides advice on characterizing, quantifying and communicating the uncertainty in Water Census information and in products based on Water Census information.

Terminology

The Water Census will yield quantitative and qualitative information as well as tools that can be used to produce additional information and/or support water management. This paper focuses on products that are in the form of quantitative information, which includes measurements and estimates, the latter of which may be based on modeling. The Water Census will both use and produce such quantitative information. For example, quantitative information on irrigation water use could be based on estimates of the irrigated land area in various crops, measurements of diversions, or a combination of both. We will use the term "data" to refer to such quantitative information as well as to quantitative products.

Uncertainty in Water Census Data

Virtually all data are uncertain, in that they can deviate from the true values. The development, release, and use of Water Census data should be informed by the associated degree of uncertainty. USGS analysts are using a wide range of uncertainty measures to select methods for developing each of the proposed products and will likely use some of these methods to decide whether a product is suitable for release.

Water Census users will use uncertainty information in a variety of ways. Ideally, the decision to use Water Census data should depend on the level of uncertainty. For example, a state regulatory agency would likely consider the uncertainty in USGS estimates of "natural" streamflows before choosing to use them to regulate instream flows. A Water Census data user might also incorporate uncertainty information in a formal environmental or economic analysis.

Many Water Census data users, however, will ignore the uncertainties. This could have unfortunate consequences. For example, state regulations based on highly uncertain estimates of streamflows could be successfully challenged in court. Hence it is critical that the USGS quantify and publish uncertainty information for all of its water census products. It would also be prudent for the USGS to track major uses of its products and even provide assistance on the use of uncertainty information in the case of heavily-used products.

We most commonly use statistical methods to characterize uncertainties in data. These characterizations range from very simple to complex. The general rule is to use the simplest approach that meets the needs of the data user.

Characterization of Uncertainty- Basic Error Structures

Single Data Values

First consider a single data value, such as the consumptive use at a powerplant during a given year. Errors in such data are typically modeled as additive or multiplicative. For additive random error the simplest model is

Data value = true value + bias + random error with zero mean

A bias can be positive or negative. Whenever possible, bias should be eliminated by some kind of calibration process. (Occasionally, known biases are left uncorrected. For example, raingage data are biased downward; however they are not commonly corrected, even though methods to correct bias are available.) The variability of the random error is characterized by its standard deviation. If the probability distribution of the error is known and there is no bias, a 95% confidence interval can be calculated. The normal distribution is most commonly assumed for computing confidence intervals. The bounds of confidence intervals are random variables. For example, on average, ninety-five percent of 95% confidence intervals contain the true value. Note, however, that many users will be unfamiliar with standard deviations and confidence intervals. Hence it would be useful for the USGS to provide information on their meaning and uses.

For multiplicative error the simplest models are

Data value = (true value x random error with a mean of 1) + bias

Data value = true value x (random error with mean greater than zero)

For either model there can be either a positive or negative bias. In the first multiplicative model the bias is constant. In the second model, the bias is proportional to the true value. As with additive error, bias should be eliminated by calibration. Multiplicative error is commonly assumed to have a lognormal distribution. The variability of multiplicative error is fully characterized by its standard deviation (or coefficient of variation). Note that the two multiplicative models are equivalent if there is no bias. Also, if there is no bias and the errors are lognormally distributed, the error structure of natural logarithm of the data is the same as the unbiased additive case with normal error. As in the case of additive error, confidence intervals can be constructed if there is no bias.

In the case of instrument measurement errors, bias is sometimes treated as a random variable (Joint Committee for Guides in Metrology, 2008). For example, a particular current meter may have an unknown bias. Hence any given meter might consistently underestimate or overestimate flow. In the case of a randomly chosen current meter the variance of biases over all meters could be added to the measurement variability. However, periodic calibration of a meter would reduce the possibility of bias. A similar approach could be applied to the bias in rain gage catch. In the water resource arena, biases are generally controlled by calibration or ignored. However, it would be prudent for the USGS to consider bias in all of its Water Census products and correct when possible.

Data Vectors or Matrices

For most applications users will be interested in a vector of data (e.g. daily precipitation at a specific location), or a matrix of data (e.g. daily precipitation at geographically gridded locations over the same time period). For many applications, it would be appropriate to assume simple additive or multiplicative errors for each data value in the vector or matrix, where the errors are uncorrelated.

For some applications errors will be correlated. For example, errors in a vector of annual water budget estimates of changes in estimated groundwater storage could be negatively correlated, as overestimation one year could contribute to underestimation in the succeeding month. There could also be correlation between vectors in a matrix of data. An example is a matrix of monthly water storages in nearby lakes, where the monthly storage at each lake is based on a single water level. Wind setup is likely to introduce errors in the water level data used to compute storages. These errors could be correlated across lakes, depending on the location of the measurements. It is relatively easy to accommodate correlation in simple additive or multiplicative error models.

Other Commonly-Used Measures of Error

In addition to bias and standard deviation, other measures are sometimes used to characterize errors. For example, mean squared error (MSE) and mean absolute error (MAE) are commonly used to quantify measurement or model errors when a set of true values is available. For example, pit gages are assumed to accurately measure rainfall, and hence are commonly used to characterize errors in operational rain gages. The mean absolute error of measured or modeled values is the arithmetic average of the absolute value of the differences between the modeled or measured values and the true values. It depends on both the bias and the standard deviation of the errors. The mean square error of measured or modeled values is the arithmetic average of the squared differences between the modeled or measured values and the true values. Note that the mean squared error is equal to the sum of the square of the bias and the variance of the errors. Variations of the mean squared error are the root mean squared error (RMSE) and the relative root mean squared error (RRMSE). The RMSE is simply the square root of the mean squared error. The RRMSE is the square root of the arithmetic average of the squared errors divided by the true value.

The accuracy of predictions is often quantified by the Pearson correlation coefficient (R^2), which equals the fraction of variance in the true values that is accounted for by the predictions. The Spearman correlation coefficient is Pearson correlation coefficient applied to the ranks of the data.

Another measure that is commonly used to estimate the accuracy of hydrologic models is the Nash-Sutcliffe Efficiency (NSE), as defined below

$$NSE = 1 - \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \mu_O)^2}$$

where S_i and O_i are the model and observed values, respectively.

The NSE is essentially 1 minus the MSE standardized by the estimated variance of the errors.

MSE, RMSE, RRMSE, MAR, R^2, and NSE are all useful for characterizing errors in predictions when true values are known. However, bias and standard deviation are the most fundamental measures and are most useful in evaluating how errors "propagate" in calculations based on data.

Complex Error Structure

A simple error structure is clearly inadequate for some Water Census data. An example is a vector or matrix of daily streamflows obtained from a deterministic rainfall-runoff model. It is well known that rainfall-runoff models are vulnerable to "structural" errors that are not well described statistically (Doherty and Welter, 2010). For example, deterministic rainfall-runoff models commonly use simple methods to model baseflow, leading to systematic time-varying errors. Although simple measures are commonly used for data with complex errors, they do not characterize the errors well.

Estimation of Uncertainty of Water Census Data - Water Budget Components

A major goal of Water Census is to provide estimates of the major hydrologic water budget components for all HUC-12 watersheds in the U.S. Currently the Water Census provides gridded estimates of precipitation and evapotranspiration (ET), and in the future will provide daily streamflows at the outlet of all U.S. HUC-12 watersheds.

Evapotranspiration

The Water Census estimates monthly ET on a 1-km grid using the Simplified Surface Energy Budget Model (SSEBM)(Senay, et al., 2013), which makes use of remotely-sensed thermal data and model-assimilated weather fields. Currently the Water Census webpage does not provide estimates of the uncertainty of the ET estimates. However, Senay et al. (2013) used eddy covariance flux tower data from 45 stations across the contiguous United States to estimate errors in monthly ET for seven ecosystem types for the period 2000-2011. The variance explained (R^2) by the model ranged from 0.43 for cropland/natural vegetation mosaic to 0.9 for the urban landscape, and averaged 0.64 over all land use/land covers. The comparisons with flux tower measurements also indicate that the SSEBM estimates were biased high for forest, shrubland, and urban lands, and low for cropland and grassland. However, it is important to note that flux tower measurements are indirect, and their accuracy is not well understood.

Of the water budget components, ET is probably the most uncertain. Large uncertainties in the ET data, if not clearly acknowledged, could undermine efforts to resolve water conflicts relating to irrigation. (See the Appendix for further discussion.) Furthermore, the uncertainty of ET estimates has rarely been quantified. The USGS should use eddy flux tower data to develop general estimates of the error bias and standard deviation of Water Census estimates of ET at both monthly and annual time scales. In addition, the USGS should use lysimeter data to investigate uncertainties in eddy flux tower measurements.

Precipitation

The Water Census provides access to 1-km gridded daily precipitation produced and archived by the Oak Ridge National Distributed Active Archive Center (ORNL DACC). These data were generated using the DAYMET model (Thornton *et al*., 1997), which is based on the spatial convolution of a truncated Gaussian weighting filter applied to National Weather Service rain gage data for the period 1980 to the present.

It does not appear that there is any published information on the accuracy of DAYMET precipitation. Because DAYMET estimates are based on rain gage data, they will be biased low. For summer precipitation, the bias in rain gage data is about 5%, while for winter precipitation the bias range from 3% to 28% (Legates and DeLiberty (1993). The DAYMET data will be subject to the same biases. The

USGS should consider applying bias corrections to the DAYMET data. Otherwise, it should advise users of the bias.

In additional to bias in the DAYMET data, there are random errors, the variance of which depends on the spatial density of the raingage data and the spatial variability of the rainfall. At individual grid points, the errors could be very large, particularly for convective events (Goodrich, *et al*, 1995). Users should be cautioned to be wary about the accuracy of DAYMET rainfall amounts at individual grid points, particularly for convective rainfall events.

NEXRAD radar data is commonly used conjunctively with rain gage data to improve the accuracy of representations of the space-time distribution of storm rainfall (e.g., Krejewski et al., 2010). The resulting estimates of the spatially distributed rainfall could be used to evaluate the accuracy of the DAYMET precipitation data (Villarini et al., 2014). As more radar data is collected, it could be used to improve the accuracy of the DAYMET data. The USGS should consider partnering with the National Weather Service to produce a large storm catalog based the conjunctive use of rain gage and NEXRAD data.

Water Budget for a Gaged Watershed

In stream-gaged watersheds for which there are no significant consumptive uses and no significant groundwater flows in or out of the watershed (such as a groundwater underflow or ET directly from groundwater), the annual water budget is given by

$$S(t+1) - S(t) = \Delta S(t) = P(t) - ET(t) - Q(t)$$

where: S(t) and S(t+1) are the watershed storages at the beginning and end of year t

ΔS(t) is the change in watershed storage in year t

P(t) is the total watershed precipitation during year t

ET(t) is the total watershed evapotranspiration during year t

Q(t) is the streamflow out of the watershed

For any gaged watershed in the nation, USGS Water Census data can be used to estimate annual watershed precipitation and evapotranspiration for the period 2000-2011 (the current duration of the ET data). Hence it is possible to estimate the annual change in watershed storage for the same period.

It is also possible to *directly* estimate the annual change in watershed storage over the same period. The main components are changes in the water content of groundwater, lakes, ponds, wetlands, soil moisture, and snowpack. The water content of large surface storages is commonly monitored. For those that are not, estimates of annual changes in their water storage could be estimated from remotely sensed data and supplemented by hydrologic modeling, as could annual changes in soil moisture and snowpack. Estimates of the annual water budget for gaged watersheds throughout the United States would provide useful insight into the accuracy of Water Census data on water budget terms.

The Appendix provides a simple example of a water budget using Water Census and stream gage data to estimate the annual change in groundwater storage contributing to baseflow. For this example, the Water Census ET data appears to significantly underestimate actual ET. The USGS should conduct annual water budgets for gaged watersheds throughout the U.S. to obtain insight regarding the accuracy of water budget data produced by the Water Census data.

Characterization of Errors in Water Census Data with Complex Error Structure- Historical Daily Streamflows

USGS researchers have recently completed a study of alternative methods for estimating historical daily streamflows at ungaged locations in the southeast United States (Farmer *et al*., 2014). The study team used a wide range of metrics to evaluate the performance of the methods, applied to the estimated daily streamflows, no-failure storage-yield curves, and streamflow statistics (including fundamental daily statistics and flow-duration curves). Seven metrics were used, including the Nash-Sutcliffe efficiency applied to untransformed and log-tranformed predictions, root-means squared error, and the Pearson correlation coefficient.

The approach used by Farmer *et al.* (2014) wisely acknowledges the fact that Water Census estimated daily flows will be used to generate a variety of products (e.g. flow-duration curves and storage-yield curves) that will be used in a variety of ways. In most cases the only practical way to estimate the uncertainties in such products will be through regional studies that use methods similar to those in Farmer *et al.* (2014). The difference will be that the purpose of such studies would be to evaluate uncertainties, rather than to select estimation methods. Also, such studies may need to consider other products, such as the seven-day, 10-year low flow ($Q_{7,10}$) and Indicators of Hydrologic Alteration (Richter et al., 1996).

Farmer *et al*. (2014) demonstrated that there was no one modeling method that produced the best results for all metrics and products. This raises the question of whether the USGS should provide estimates of certain high-demand products, such as $Q_{7,10}$, or flow-duration curves, rather than letting the user extract these products from estimated daily flows.

Communication of Uncertainty

The USGS Water Census will have many customers, and its products will be used in a wide variety of predicable and unpredictable ways. It is essential that users of Water Census data be provided with uncertainty information that enables them to understand the limitations of the data. In particular, the USGS should specifically mention inappropriate uses of Water Census data. Sophisticated users should be provided with information that enables them to estimate how water census data uncertainty propagates through analyses conducted with the data. We also recommend that the USGS to develop the capacity to track the use of Water Census data so that it can detect inappropriate uses of the data as well as improve products and perhaps develop new ones.

For all Water Census data with simple error structures, the USGS should provide information on biases as well as confidence limits and standard deviations (in real-space or log-space, as appropriate). This information should be accompanied by a carefully written narrative on uncertainty. The USGS has always done an exceptional job of communicating uncertainties associated with all its paper and web-based products. Given the likely diversity of Water Census users and uses, the communication challenge will be much greater.

Water Census products with complex error structures are more challenging. The estimated historical daily streamflows are the main such products. It is not yet clear how these streamflows will be estimated, or exactly how they will be used. We recommend that the USGS develop these products regionally, using methods similar to those used by Farmer *et al.* (2014). The use of such methods will enable the estimation of the uncertainties associated with the most likely products that will be based on the estimated streamflows. The USGS should also track the use of this product to help discourage misuse.

Recommendations

Throughout this report we have made recommendations regarding uncertainty and Water Census products. These are summarized below.

Data with Simple Error Structures

- Use bias, standard deviation and confidence intervals as the primary measures of uncertainty; associate descriptive terms based on confidence limits (e.g., Excellent, Very Good, Fair and Poor)
- Use as much data as possible to "refine" Water Census data and estimate uncertainties
    - Conduct annual water budgets for gaged watersheds throughout the U.S. to obtain insight regarding the accuracy of water budget data produced by the Water Census data.
    - Consider de-biasing DAYMET precipitation data using the de-biasing methods in the literature.
    - Consider using NEXRAD data conjunctively with DAYMET data to refine the DAYMET data for major storms.

- Use flux tower data to develop general estimates of the error bias and standard deviation of Water Census estimates of ET at both monthly and annual time scales.
- Use lysimeter data to investigate biases in flux tower measurements

Data with Complex Error Structures

- Use regional studies such as Farmer *et al*. (2014) to develop and evaluate the uncertainty in Water Census data using metrics that reflect their most likely uses.
- Consider providing estimates of certain high-demand products based on daily streamflows, such as $Q_{7,10}$, or flow-duration curves, rather than letting the user extract these products from estimated daily flows.

Communication

- Develop and implement a consistent strategy for communicating the uncertainty associated with each Water Census product.
- Track the use of Water Census products and consider providing assistance (e.g. guidance documents) on the use of uncertainty information in the case of heavily used products.
- Use the tracked information on the use of Water Census data to improve the uncertainty information.
- Specify inappropriate uses of Water Census Data

Appendix- Example Water Budget for a Gaged Watershed

As discussed in the body of this report, the USGS Water Census data can be used to estimate annual precipitation and evapotranspiration for the period 2000-2011, enabling calculation of the annual change in in water storage for the same period. It is also possible to directly estimate the change in storage annually and over the 12-year period. Here we conduct a water budget for a watershed in which the annual change in water storage is dominated by the annual change in groundwater storage contributing to baseflow. The results provide useful insights regarding the accuracy of the water budget terms.

We estimate the change in groundwater storage by assuming that the groundwater system acts like a linear reservoir, such that groundwater flow contributing to streamflow at time t is proportional to the groundwater storage. That is,

$Q_g(t) = wS_g(t)$

where: $Q_{gw}(t)$ is the groundwater flow contributing to sreamflow (cfs)

S_g (t) is the groundwater storage contributing to streamflow (cfs-days)

w is a constant (days.)

During periods when there is no recharge and streamflow is entire composed of baseflow, it is easily shown that

$Q(t) = k^t Q(0)$

where $k = \exp(-t/w)$.

Based on this result one can estimate the annual change in groundwater storage contributing to baseflow as

$S_g(t+1) - S_g(t) = w[Q(t+1) - Q(t)]$

We have applied this approach to the Eau Pleine River in Wisconsin. The gaged portion of watershed has an area 220 square miles. The land use is dominantly agricultural, and there are no large surface water storages in the watershed.

Based on the streamflow on October 1 of each year, we estimated the annual water-year change in groundwater storage assuming values of k equal to 0.95, 0.99, and 0.999, a range of k that covers most all streams in Wisconsin. The data and results are given in Table A-1.

Table A-1

DATA						
Water Year	Precip. (in)	ET (in)	Q (in)	ΔS (in)	Q (Oct 1) (cfs)	Q (Sept 30) (cfs)
2001	33.4	20.7	12.4	0.4	9.4	15.0
2002	46.0	20.9	18.7	6.3	15.0	90.9
2003	25.0	19.4	10.6	-5.0	92.1	5.2
2004	29.9	19.9	9.3	0.7	5.3	6.6
2005	28.7	20.3	6.9	1.5	6.7	11.7
2006	28.4	19.8	4.5	4.1	12.5	8.7
2007	30.3	20.3	5.0	5.0	9.0	16.3
2008	27.0	21.0	8.6	-2.6	17.2	4.8
2009	29.1	19.5	4.5	5.1	5.0	5.5
2010	41.1	21.4	13.5	6.2	6.7	52.6
2011	37.5	21.2	16.5	-0.2	51.5	7.4
CALCULATIONS						
	k = 0.95		k = 0.99		k = 0.999	
Water Year	dS (in)	Diff. (in)	dS (in)	Diff. (in)	dS (in)	Diff. (in)
2001	0.0	0.3	0.1	0.3	0.9	-0.6
2002	0.3	6.1	1.3	5.0	12.8	-6.5
2003	-0.3	-4.7	-1.5	-3.6	-14.7	9.7
2004	0.0	0.7	0.0	0.7	0.2	0.5
2005	0.0	1.4	0.1	1.4	0.8	0.6
2006	0.0	4.1	-0.1	4.1	-0.6	4.7
2007	0.0	5.0	0.1	4.9	1.2	3.8
2008	0.0	-2.5	-0.2	-2.4	-2.1	-0.5
2009	0.0	5.1	0.0	5.1	0.1	5.1
2010	0.2	6.0	0.8	5.4	7.8	-1.6
2011	-0.1	0.0	-0.7	0.6	-7.5	7.3
Cumulative Diff. =		21.5		21.6		22.4

The most striking result of this analysis is that the water budget based on the Water Census data indicates a net increase of about 22 inches of watershed storage over the 11-water-year period (2001-2011). This result is in sharp contrast to the estimates of net change in storage computed from baseflow, which ranges from -1 inch to zero inches. The latter result is clearly the most reasonable. We know of no mechanism that would explain a 22-inch increase in watershed storage over an 11-year period. We believe that this anomalous result is due to a consistent error in one or more water budget terms. It is highly unlikely that the streamflow data are biased. The precipitation data are likely biased **low**, by about 15% (estimated from information in Legates and DeLiberty (1993)), or 53 inches. Hence it appears that the evapotranspiration data for the period 2010 through 2011 are biased high by about 75 inches (33%).

Given these results, it is critical that similar water budget calculations be made at all watersheds that have been gaged during the period 2001-2011. A large bias in the Water Census ET data could create major problems, particularly in regions where there are major conflicts over water use. For example, in the Central Sands region of

Wisconsin, which is just southeast of the Eau Pleine watershed, steady increases in the area of irrigated agricultural lands since the 1950s are likely responsible for observed decreases in lake levels and stream baseflow. However, many of the agricultural producers have not accepted this explanation. Water Census ET values that are biased low could easily undermine efforts to resolve this water use conflict.

References

Chang, M. and L. A. Flannery, 1998, Evaluating the accuracy of rainfall catch by three different gages, Journal of the American Water Resources Association, 34(3), 559-564.

Doherty, J., and D. Welter, 2010, A short exploration of structural noise, Water Resources Research, 46, WR05525, doi:10.1029/2009WR008377.

Duchon, C. E. and G. R. Essenberg, 2001, Comparative rainfall observations from pit and above-ground rain gauges with and without wind shields, Water Resources Research, 37, 3253-3263.

Farmer, W. H., S. A. Archfield, T. M. Over, L. E. Hay, J. H. La Fontaine, and J. Kiang, 2014, A comparison of methods to predict daily historical streamflow time sries in the southeast United States, DRAFT MANUSCRIPT.

Goodrich, D.C., Faurès, J.M., Woolhiser, D.A., Lane, L.J., and Sorooshian, S., 1995, Measurement and analysis of small-scale convective storm rainfall variability. J. Hydrology 173:283-308.

Groisman, P.Ya. And D.R. Legates, 1994, The accuracy of United States precipitation data, Bulletin of the American Meteorological Society, 75: 215-227.

Joint Committee for Guides in Metrology, 2008, Evaluation of Measurement Data-Guide to the Expression of Uncertainty in Measurement, First Edition, 2008.

Krajewski, W. F., A. Kruger, J. A. Smith, R. Lawrence, C. Gunyon, R. Goska, B. C. Seo, P. Domaszczynski, M. L. Baeck, M. K. Ramamurthy, J. Weber, A. A. Bradley, S. A. DelGreco, and M. Steiner (2010b), Towards better utilization of NEXRAD data in hydrology: An overview of Hydro-NEXRAD, *Journal of Hydroinformatics*, *13.2*, 255–266, doi:10.2166/hydro.2010.056.

Legates, D. R. and T. L. DeLiberty, 1993, Precipitation measurement bias in the United States, Water Resources Bulletin, 29(5), 855-861.

Richter B., Baumgartner J.V., Powell J. & Braun D.P. 1996, A method for assessing hydrologic alteration within ecosystems, Conservation Biology, 10, 1163–1174.

Senay, G.B., S. Bohms, R.K. Singh, P.H. Gowda, N. M. Velpuri, H. Alemu, and J.P. Verding, 2013, Operational evapotranspiration mapping using remote sensing and weather datasets: A new paramteriization for the SSEN approach, Journal of the American Water Resources Association, 49(3), 577-591.

Sevruk, B., 1982, Methods of correcting for systematic error in point precipitation measurement for operational use, Operational Hydrology Report No. 21, World Meteorological Society (WMO-No. 589).

Thornton, P.E., S.W. Running, and M.A. White, 1997, Generating surfaces of daily meteorological variables over large regions of complex terrain, Journal of Hydrology, 190, 212-251.

Villarini, G., B.-C. Seo, F. Serinaldi, and W.F. Krajewski, 2014, Spatial and temporal modeling of radar rainfall uncertainties, *Atmospheric Research*, 135-136, 91-101.