ESTIMATING MONTHLY TIME SERIES OF STREAMFLOWS AT UNGAUGED LOCATIONS IN THE UNITED STATES

A thesis

submitted by

William Hastings Farmer

In partial fulfillment of the requirements
for the degree of

Master of Science

In

Civil and Environmental Engineering

TUFTS UNIVERSITY

May 2012

ADVISER: Richard M. Vogel

ABSTRACT:

Estimation of time series of monthly streamflows at ungauged locations is of paramount importance to many hydrologic projects and applications. This project uses leave-one-out experiments to assess the relative performance of a few traditional, regional, hydrostatistical prediction techniques at minimally-impacted streamflow gauges across the United States. This project considers four traditional flow-transfer techniques: drainage-area ratios (DA), standardization by mean (SM), maintenance of variance extension (MOVE) and the use of flow duration curves (QPPQ). Under idealized conditions, all methods significantly outperformed the drainage-area ratio (DA). However, when the flow-transfer techniques were combined with regional regression methods the relative performances were significantly degraded. A weighting scheme is introduced, combining the advantages of DA with the improved performance of MOVE or SM. This so-called weighted average (WAve) offers significant advantages over the traditional drainage area ratio techniques.

ACKNOWLEDGEMENTS:

Portions of this research were funded by the United Nations University's World Institute for Development Economics Research.

This project would not have been possible without the strong support of my thesis committee: Drs. Richard Vogel, Kenneth Strzepek and Stacey Archfield. All three provided fundamental criticism and invaluable advice. As my adviser, Dr. Vogel passed on both a passion for and appreciation of the subject matter at hand. He ignited a fire within me and has helped me to stoke that same fire; I am more than hopeful that we will continue our collaboration on a brilliant PhD project. Dr. Strzepek, as always, proved a priceless asset and dear friend. His passion for meaningful science has molded this project into a useful tool for decision-making in a real-world context. Thanks to Dr. Archfield for her vast understanding of the subject matter, insightful critiques and initial groundwork.

I would also like to thank all of my office colleagues and friends. I could not have completed this project without the unique mix of scholarly counsel, life advice and comic relief that infused our workspaces. I wish you all the luck in the world and hope that my contributions to your work have been as helpful as yours have been to mine.

Finally, I would like to thank my family – my parents Gregory and Mary, my brother Patrick and my sister Bridget – for their love and support throughout this process. Most especially, I would like to thank my fiancée, Laura, for her patience, support and tolerance through countless drafts and presentations.

Table of Contents

EXECUTIVE SUMMARY

Estimation of time series of monthly streamflows at ungauged locations is of paramount importance to many hydrologic projects and applications, ranging from hydropower and water supply to irrigation scheduling and development planning. In fact, prediction in ungauged basins (PUB) has been one of the leading initiatives of the International Association for the Hydrological Sciences (IAHS) over the past decade. The PUB initiative recognizes the pressing need to understand the availability of our water resources, especially in regions with developing populations and high vulnerability to climate change.

There exists a wide range of both process-based and regional, hydrostatistical approaches for estimating streamflows at ungauged locations, though few comparisons among such approaches exist. This project uses leave-one-out experiments to assess the relative performance of a few traditional, regional, hydrostatistical prediction techniques at over 1000 minimally-impacted Hydro-Climatic Data Network streamflow gages across the United States. The PUB problem can be broken into three distinct steps: (1) the selection of an index gauge, (2) characterization of the streamflow record at ungauged location and (3) transfer of streamflow information from the gauged to the ungauged site. This study disregards step one and almost exclusively considers step three.

Four different flow-transfer techniques were considered here: drainage-area ratios (DA), standardization by mean (SM), maintenance of variance extension (MOVE) and the use of flow duration curves (QPPQ). Under idealized conditions, when perfect information is assumed in step two, all methods significantly outperformed the drainage-area ratio (DA) method in terms of bias and Nash-Sutcliffe Efficiency. Across the US, a variant of MOVE, the top performer, outperformed DA at more than 86% of sites. However, when the flow-transfer techniques in step three were combined with regional regression methods for estimating streamflow moments in step two, very different results were obtained. In general, the application of steps two and three in tandem no longer led to superior performance, with the DA method performing best at over 50% of sites in the United States.

Evidence suggests that the performance of flow-transfer techniques is significantly influenced by the under- or over-lying hydroclimatology of the region. With this in mind, a weighting scheme based on long-term climate statistics and the relationship between the site and index site is introduced, combining the advantages of DA with the improved performance of MOVE or SM. This so-called weighted averaging (WAve) technique is found to outperform DA methods at 60% of sites across the United States. By considering overall and monthly performance, it is shown that WAve offers significant advantages over the traditional drainage area ratio methods in almost all regions of the United States.

These results are quite promising and may represent an important step forward in the PUB dilemma. Further research is suggested, allowing for the contribution of groundwater to streamflow, the application of streamflow estimates to real-world projects and the use of remote sensing data to augment hydrostatistical methods of time series estimation.

I. INTRODUCTION

Freshwater is the most important natural resource in the world. It is essential to each person on a day-to-day basis and is extremely influential in nearly all fields, from politics to science (Sivapalan 2003). Yet, despite a long history of research, there remains a wide range of uncertainty associated with estimating, predicting and forecasting the quantity and quality of this mighty resource across space and time. In truth, the scientific community has a rather poor understanding of the processes associated with runoff generated from rainfall: the question of where water goes when it rains has yet to be fully understood (Sivaplan *et al.* 2003).

With the importance of freshwater resources in mind, it becomes clear that understanding the quantity of water or water resources in space and time is perhaps the most important problem confronting the hydrological sciences today. Before one can begin to understand, predict and transform the quality and use of freshwater, one must first completely understand the amount of water present and available (Sivapalan *et al*. 2003). Accordingly, prediction of time series of streamflow at ungauged locations is one of the grand challenges facing hydrologic scientists today (Sivapalan 2003). Time series of flows are of supreme importance in a wide range of water-related projects, from irrigation scheduling and water supply planning to hydropower development and flood mapping. Beyond real-world applications, time series of flows are extremely important for the calibration and validation of rainfall-runoff models and more complicated models of streamflow.

Some regions of the world, like North America, have extensive streamflow gauge networks, while other, often poorer regions have much more scantily-gauged networks (Sivapalan *et al*. 2003). Additionally, in many of the well-gauged regions natural flow measurements are confounded by human development in the watershed (Sivapalan 2003). Therefore, within each network of stream gauges there exists a smaller subset of minimally-impacted gauges that can provide a time series of flows that can be used to transfer information to nearby ungauged locations. It is these naturalized flows that are most valuable to the water-related projects including understanding floods of given exceedance probabilities, mean annual water yields, reliability of water supply, crop yields, and soil moisture patterns needed for irrigation scheduling (Sivapalan *et al.* 2003).

Understanding natural streamflows is therefore essential to human development in ungauged basins (Sivapalan 2003). Furthermore, the natural streamflow regimes are being drastically affected by human development, land use change and global climate change; the stationarity of hydrologic science has dissolved (Wagener *et al.* 2004; Milly *et al*. 2008). It is further troubling that the places with the least-developed gauge networks are those same places where human development is having the greatest impact on natural flow regimes and the need for freshwater is most poignant (Sivapalan *et al.* 2003). All of these aspects sum up to a truly grand challenge.

It is the purpose of this thesis to increase our understanding and ability to estimate natural streamflow time series in ungauged basins. There are a wide range of models for predicting streamflows, from empirical models, to lumped models and on to distributed models and statistical regionalizations (Sivapalan *et al.* 2003). Yet, little attention has been paid to the relative performance of all of these techniques (Asquith *et al.* 2006). This report will examine a number of statistical regionalization techniques in the context of predicting monthly streamflow time series. This work, in the context of the greater hydrological sciences, sets the stage with a new methodology for assessing and evaluating predictive methods for streamflow.

In 2003, the International Association for the Hydrological Sciences (IAHS) launched an initiative focused on Prediction in Ungauged Basins (PUB). The PUB initiative is aimed at engaging the scientific community in a cohesive effort to advance the understanding and prediction capability of hydrologic parameters in ungauged basins (Sivapalan *et al*. 2003). By defining ungauged basins as those that lack sufficient length or quality of recorded data, prediction is understood to include reconstruction of past events, prediction of future and passed magnitudes and forecasting, the coupling of certain magnitudes with particular points in time (Sivapalan *et al*. 2003).

This project falls within the heart of the PUB initiative. This exercise will address Target One, which proposed to "examine and improve existing models in terms of their ability to predict in ungauged basins through appropriate measures of uncertainty" (Sivapalan *et al.* 2003). Furthermore, this project also directly addresses Theme Three, to "advance the [scientific understanding] from the application of existing models through uncertainty analyses and model diagnostics" (Sivapalan *et al*. 2003). In particular, this project will explore the performance and predictive capability of regional, statistical methods, much as Hirsch did in 1979.

As has been observed, there are many techniques for estimating streamflow at ungauged basins. Such techniques involve different levels of complexity, data availability and applicability. Generally, these techniques fall into two separate categories: process-based models, which use deterministic relationships to build a process-based model of an individual watershed and scale upwards; and hydrostatistical models, which use regional statistics of gauged streamflows to predict nearby ungauged streamflows. Within these groups exist several techniques: standardization, extrapolation, measurements by remote sensing and a whole range of climate-based models (Sivapalan *et al.* 2003).

The problem of estimating streamflow in ungauged basins with regional, hydrostatistical methods can be understood to consist of three distinct steps. In order, these steps are (1) the selection of an index gauge, (2) characterization of the streamflow record at the ungauged site and (3) the transfer of streamflow infromation from the gauged to the ungauged site. Books could be written about each of these steps, but this report focuses on both Steps Two and Three, with greatest attention given to Step Three.

Here Step One is ignored, and the implications of and the process by which an index gauge is selected are not addressed. In lieu of a more complete assessment of Step one, the most common approach, that of choosing the nearest gauge to a site of interest as the index gauge, is used in all experiments. Though this method is suggested by the work of Emerson *et al.* (2005), Asquith *et al.*

(2006) and Mohamoud (2008), and many others, recent publications have found other, more-valuable criteria for selecting index gauges. Achfield and Vogel (2010), for example, found that estimating the correlation between streamflow records is a strong metric for hydrologic similarity between an index gauge and a site of interest. Furthermore, they found that the closest gauge is often not the gauge that exhibits the strongest correlation with streamflows at the ungauged site.

This thesis is organized as follows: First, Chapter Two summarizes the general methods and datasets used. Then, in Chapter Three, considering Step Three, a number of flow-transfer techniques are assessed in an idealized sense, meaning that the moments of the streamflow record, usually estimated in Step Two, are instead assumed to be known. Jack-knife, or leave-one-out experiments are performed assuming the streamflow moments and other parameters are known a priori from Step Two, with perfect foresight. These idealized exercises allow one to assess the performance of flow-transfer techniques without confounding the analyses with uncertainty from techniques used to characterize the record at the ungauged site.

Chapter Four briefly considers Step Two, characterizing the streamflow record at the ungauged site. Here several methods are evaluated for their ability to estimate moments at ungauged sites. As the second step is not the focus of this report, only regional regression methods are considered as a possible technique for record characterization. Thus Chapter Four focuses on understanding the uncertainty associated with methods for record characterization alone.

In Chapter Five, the real-world application of the prediction of monthly streamflows in ungauged basins is assessed using leave-one-out experiments, combining regional characterization (Step Two) with the flow transfer techniques considered in Step Three. This chapter demonstrates that the uncertainty introduced by regional characterization severely hampers the performance of the streamflow transfer methods considered. Thus it is demonstrated, perhaps for the first time, why the drainage-area ratio method has become the default prediction technique.

Chapter Six, the crux of this project, introduces a weighting method that combines and improves upon the estimates of previous methods. It is shown that simple selection criteria based on hydroclimatic conditions can be used to weight the estimates of each prediction method appropriately and generally improve on the DA method with only the simplest of additional climatic inputs. In essence, this technique combines the advantages of the DA method with the added value of regional hydrostatistical methods.

This thesis concludes with a recommendation for predicting streamflows in ungauged basins based on the hydrostatistical weighting procedure. Of course, the door for further research is left wide open with suggestions for future promising ideas.

II. Data Sources and Generalized Methods

This chapter considers a few general topics that will be used in each of the subsequent chapters. These topics include a summary of the data being used and an overview of general method and terms. Outlining these terms now will reduce the need for repeating them in each chapter. After describing the sources for all of the data being used, a brief description is provided for each of several different methods that will be used on several occasions throughout this thesis. These methods represent, in a general form, the tools used to generate, analyze and evaluate all of the results and streamflow estimation techniques.

Data Sources

In order to reasonably evaluate several different techniques for estimating streamflow at ungauged sites, it is necessary to have some level of truth with which to compare the estimates. This was achieved by using gauged data from minimally-impacted sites and conducting leave-one-out or jack-knife experiments, which are described in more detail below. The streamflow data comes from the US Geological Survey's Hydroclimatic Data Network (HCDN; Slack *et al*. 1993). This dataset, which can be compiled monthly, is a collection of streamflow measurements from gauges that exhibit little or no anthropogenic impacts. The network contains more than 1500 gauges across the United States, though only around 1,300 were used here because only those sites could be coupled with reliable climate data for the same watershed over the period of record.

It is well known that long-term climate statistics are strongly connected with regional streamflow moments (Vogel *et al*. 1999). Accordingly, climate time series that are spatially and temporally representative of the watersheds defined by the HCDN streamflow gages are needed. Indexed by HCDN basins, such spatially aggregated climate data was compiled using automated GIS methods and made available by Vogel and Sankarasubramanian (2005). Their database includes a coincident time series of monthly climate variables, including temperature, precipitation and potential evapotranspiration, from 1951 through 1990 for about 1,300 sites from the HCDN streamflow database. The estimates of climate are based largely on the PRISM system defined by Daly *et al.* (1994 and 1997), which were spatially interpolated over each of the 1300 HCDN watershed boundaries. Further information about the hydrology of each site was garnered from the watershed characteristics database developed by Kroll *et al*. (2004).

The analysis and characterization of streamflow patterns and moments was conducted on a regional basis. Each basin of the HCDN is contained with one of the 18 2-digit HUCs of the United States (**Figure 2.1**). To enable an effective summary of the results three meta-regions of the US were employed: the East, Midwest and West. Respectively, these consist of two-digit HUCs one through six, seven through twelve and thirteen through eighteen. This analysis does not consider regions outside of the continental, conterminous United States of America.

A Compendium of Generalized Methods

Leave-One-Out or Jack-Knife Experiments

Leave-one-out or jack-knife experiments are a way to test different estimation procedures. This allows one to treat a gauged site as an ungauged site and quantify the performance of each prediction method. In practice, one site is blinded, meaning that the data from that site is ignored or removed for the moment so that the blinded site becomes the ungauged site of interest. The estimation and flow transfer techniques are then developed with the remaining data only. Once the methodology has been calibrated at the remaining sites, any input data from the blinded site can now be used to generate results at the ungauged or blinded site. The result is a fully estimated flow series at the blinded site that can be directly compared with observed values. This process can then be replicated for every site in the database, leading to a full sample of performance metrics across any region. The results mimic the application of the proposed methodology at an ungauged site and reflect the performance of the methodology under realistic conditions.

Nash-Sutcliffe Efficiency

The Nash-Sutcliffe efficiency is a commonly used performance metric for assessing the ability of models to accurately represent observations. This statistic is commonly used in the hydrological sciences and will be used here to assess the performance of each method in a leave-one-out exercise. The Nash-Sutcliffe model efficiency, *NSE*, is given as

$$NSE = 1 - \frac{\sum_{i=1}^{N}(X_i - \hat{X}_i)^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \qquad (2.1)$$

where $\hat{X}$ is the modeled estimate of X for a time series of length N. Better prediction methods will cause NSE to approach unity. The Nash-Sutcliffe efficiency, in terms of streamflow, can be calculated overall or on a monthly basis. The NSE is preferred over the correlation coefficient and other measures because it accounts for bias in addition to the correlation between the observations and predictions. For example, a negative value of NSE reflects situations in which the bias in the predictions is so great that the mean value would be a better prediction than the model, itself. For unbiased model estimates, NSE reduces to the product moment correlation coefficient.

Percent Bias

The percent bias of a leave-one-out method, calculated across all estimates or monthly, is a measure of how close, on average, a given estimation procedure is to the true values. The percent bias, B, of an estimate is given as

$$B = 100 * \left(\frac{m_X - m_{\hat{X}}}{m_X}\right) \tag{2.2}$$

where m is the arithmetic mean of the subscript variable and $\hat{X}$ is the modeled estimate of X. In this case, if a method overestimates the observation, the bias will be negative, while an underestimating method will exhibit positive bias. An unbiased method will exhibit both a narrow range of bias and a median value near zero across all of the leave-one-out experiments.

Relative Efficiency

Judging the relative efficiency of two methods is a technique borrowed from traditional statistics. Relative efficiency quantifies the relative performance of two estimators by assessing the ratio of their variances using

$$e(X_1, X_2) = \frac{E[(X_2 - \theta)^2]}{E[(X_1 - \theta)^2]} \qquad (2.3)$$

where X_1 and X_2 are estimators of θ. A relative efficiency greater than one indicates that the estimator X_1 has lower variance than the estimator X_2, in which case X_1 is said to be more efficient than X_2. In such a case, X_1 holds an advantage over X_2. Ratios similar to relative efficiency can also be used to compare one method over another in terms of NSE or bias. When considering the ratio in terms of NSE values, the true value would be one, while X_1 and X_2 would become the NSEs of two separate methods. Similarly, when considering the ratio in terms of bias the true value would be zero, while X_1 and X_2 would become the biases of two separate methods.

Handling of Outliers

When testing numerous methods over a wide range of sites and data it can be difficult to ensure the quality of data and the absence of discontinuities. Furthermore, extreme results can severely affect the performance measures indicated above, disrupting their distributions and thus their credibility. To avoid this, a decision was made to exclude outliers from the analysis so that each method could be judged equal. In all the exercises that follow the outliers that are beyond 1.5 times the inter-quartile range were ignored. This metric is common for removing outliers in elementary statistics. After outliers had been removed from the performance statistics of NSE and bias summarized across all the leave-one-out experiments, the relative efficiency and second-order performance metrics of the varying methods can be appropriately evaluated.

The Aridity Index

In this study, the aridity index is found to be an important explanatory hydroclimatic variable. The aridity ratio is defined as the ratio of the average precipitation to the average potential evapotranspiration. The aridity index is a common measure of the "wetness" of a particular site and can be defined on a monthly or annual basis. It combines the climate signals of precipitation, temperature and global position into a single measure. It will be shown that this measure becomes very important in classifying, understanding and improving the performance of the various methods considered here.

III. Traditional Streamflow Transfer Techniques

When estimating monthly streamflows at ungauged locations, uncertainty can be introduced at any of the three steps indicated above. As one of the motivations for this research is to understand the nature and origin of some of this error, this chapter attempts to consider Step Three independent of the choice of methods used in Steps One and Two. The goal is to evaluate the performance of various existing methods for the transfer of streamflows from a gauged to an ungauged site. By conducting leave-one-out experiments where Steps One and Two have no additional impact on the analysis because moments are known explicitly, one can understand the intrinsic uncertainty that is native to each of several flow-transfer techniques considered in Step Three.

Four types of flow transfer techniques are considered. These include standardizing flows by drainage area, the most common method, standardizing flows by mean flows, standardizing with a maintenance of variance extension (MOVE) and the use of flow duration curves (FDC). Each method is described below. Several variants are considered for each type, with the best being identified with an analysis of Nash-Sutcliffe and relative efficiencies. These four classes of methods are then compared against each other and each is than evaluated against the most common drainage area ratio technique.

Standardizing Flows by Drainage Area

Standardizing monthly flows by drainage area is one of the most commonly used hydrostatistical techniques for transferring streamflow information from one site to another (Asquith *et al.* 2006). It is mentioned in nearly every introductory hydrology textbook, as well as a variety of handbooks (Stedinger *et al*., 18.54). Commonly referred to as the drainage area ratio (DA) technique, it is perhaps the most widely used method for estimating streamflow time series in ungauged basins (Archfield and Vogel 2010). Surprisingly, this method has rarely been compared with other more rigorous methods that require more inputs. Such is the goal of this thesis.

The use of DA is appealing because it requires no additional information other than the streamflows at an index site and the drainage areas of the index and ungauged sites, making it the easiest possible method that one could consider. Thus the DA method does not require Step Two. The DA method is often selected simply because other methods are too complicated, require too much data or have yet to be developed for regions of interest (Emerson *et al.* 2005). In essence, the DA method is simply the method of choice by default.

Not surprisingly, the lack of streamflow characterization in Step Two implies that the selection of an index gauge becomes that much more important for DA methods (Asquith *et al.* 2006). In the DA method the index site transfers information about both the timing and the magnitude of the streamflows, while in other techniques that require streamflow characterization (Step Two), the index gauge transfers the timing of flows and relative, standardized flow magnitudes for DA techniques. It is for this reason that much attention has been given to the selection of an index gauge in the field of hydrology. Asquith *et al.* (2006) found that the performance of generalized DA methods was closely linked with the separation distance between two sites and the logarithm of the ratio of their drainage areas.

In general, the watershed of an index gauge should be hydrologically similar to the watershed of the ungauged site. It is thought that hydrologic similarity will ensure well-behaved estimates. For the DA method, many have argued that the ratio of drainage areas should approach unity between hydrologically similar basins (see the discussion in Emerson *et al.* 2005 and Asquith *et al.* 2006). Here, the optimal selection of an index gauge has been left for further research – see Archfield and Vogel (2010) for recent innovations in the selection of an index gage. Instead, the nearest streamflow gauge is always used, per the recommendations of Mohamoud (2008) and many others.

In order to understand the performance of the simplified DA technique, it is important to consider two variants of this technique. First is the more traditional approach, drainage area ratio in real space (DAR). DAR assumes that the flow per unit area, or unit discharge, in real space, is equal across hydrologically similar basins. That is, for any given month,

$$\frac{Q_X}{A_X} = \frac{Q_Y}{A_Y} \tag{3.1}$$

for two sites, X and Y, with monthly streamflow Q and drainage area A. Traditionally, site X is considered the gauged site and site Y is the ungauged site. Accordingly, the flow at the ungauged site can be approximated simply using

$$Q_Y = A_Y \frac{Q_X}{A_X} \tag{3.2}$$

which can be applied to each month to create a time series of monthly flows.

Often times the logarithms of streamflows are better behaved than the flows in real space. For this reason it is important to consider the log-space variation of the drainage area ratio (DAL). DAL recognizes that one could standardize the logarithms of flows by the logarithm of the drainage areas such that, for any given month,

$$\frac{\ln(Q_X)}{\ln(A_X)} = \frac{\ln(Q_Y)}{\ln(A_Y)} \tag{3.3}$$

for the two sites *X* and *Y*. This is an interesting conceptual framework because, with the change of base of logarithms, (3.3) can be rewritten as

$$\log_{A_X}(Q_X) = \log_{A_Y}(Q_Y) \tag{3.4}$$

Solving equation (3.3) yields,

$$Q_Y = Q_X^{\frac{\ln(A_Y)}{\ln(A_X)}} \tag{3.5}$$

which allows for the streamflow at the ungauged site to be estimated directly, yet with a different functional form than the common result in (3.2).

Having selected an index gauge, one may estimate the flows at a site and compare the resulting estimates with the known, gauged flows. At each site an overall Nash-Sutcliffe Efficiency (NSE) and average bias (B) can be calculated for each method. A summary of the range of NSEs and percent bias for DAR and DAL is presented in **Figure 3.1**. DAR exhibits generally higher efficiency and lower bias than DAL. For DAR, more than 75% of the sites yielded an NSE greater than 0.5, while, for DAL, many sites resulted in sub-zero NSEs. While both methods have a median bias of zero, the range of percent bias associated with DAR was narrower than that associated with DAL. Furthermore, the DAR's bias is generally symmetric about zero; DAL's is skewed slightly downward. These boxplots provide strong evidence for the use of DAR over DAL.

The previous figures represent the data nationally, but it is important to understand the performance of these methods across different hydroclimates. Instead of examining all 18 regions of the United States, it is useful to examine the results by three meta-regions: East, Midwest and West, as defined previously. The range of efficiency and bias for DAR and DAL in these meta-regions is presented in **Figure 3.2**. In all regions, DAR performs better than DAL. It is interesting to note that DAR performs very well in the East, but its performance degrades in the Midwest and West. This is a large weakness in the DAR method that will be explored later. The bias of DAR is better behaved than of DAL in all meta-regions. Poor performance in the West could be due to the scarcer gage network, aridity and heterogeneous climates of that meta-region.

While boxplots provide a very important understanding of the general behavior of each method, it is more important to understand the site-by-site behavior of each method. This can be done by calculating the relative efficiency of each method's NSE and bias. Here, a relative efficiency greater than one suggests that DAR is a better estimator than DAL. Nationally, DAR exhibited a relative efficiency of 4.5 for NSE and over 10 for bias. These results were of similar magnitude across all meta-regions. The advantage of DAR is most pronounced in the East, with a relative efficiency above 10 for both NSE and bias. DAR's advantage in terms of both metrics decreases towards the west, though the relative efficiency is still well above unity in the West.

Another valuable perspective is the comparison of the performance of each method at each site. This can be quantified by observing the percentage of sites at which one method outperforms the other in terms of bias and NSE. One would expect a stronger method to outperform another at more than 50% of sites. In the entire US, DAR yields a greater NSE than DAL at 77% of sites while resulting in a smaller bias at 83% of sites. Across the meta-regions, these percentages exhibit a similar trend to that seen in the relative efficiencies. In summary, DAR generally performs significantly better than DAL in all regions of the US, though especially in the East and less so in the Midwest and West.

One major drawback of the DA techniques is that they do not explicitly account for monthly seasonality. Some seasonality is transferred directly from the 'seasonal signal' contained within the index gauge, but this varies by selection of the index gauge. In order to begin to understand the seasonal impacts of the DA methods, it is useful to consider the monthly performance of each method rather than the annual performance summarized above.

Figure 3.3 shows the ranges of NSE and bias for DAR and DAL in each month. Both DAR and DAL exhibit an interesting trend in monthly NSEs: there is a slight decrease in NSE in the late summer and early fall months, which are generally the driest months. Investigation of this decrease is left for further research. On the whole, DAL performs poorly across almost all months, especially in terms of bias. The bias of DAR is much more interesting. There is a clear cyclical trend in the range. In the summer months, the range of bias for DAR is quite small. This range increases dramatically in the winter months. This variation in bias may be due to some hydroclimatic variation across months: differences in aridity indices or mean precipitation. This will be discussed in more detail below.

The monthly trend in the performance of DAR and DAL is repeated, to some degree, in each meta-region, as seen in **Figures 3.4-3.6**. Between all the meta-regions, DAR performs best in the East. The same cyclical trend of bias is seen in all three meta-regions. Again, the slight drop in NSE does not seem to be connected to the percent bias. Whereas the annual results suggested that annual bias was somehow linked to the hydroclimate of each meta-region, these results show that there exists a second trend in performance that is not tied to the meta-region, but rather is linked to the monthly changes in climate.

Considering the monthly relative efficiencies in the US and each meta-region, the advantages of DAR over DAL are quite remarkable. With a national average of 8.22, DAR has a relative efficiency well above unity for both NSE and bias in all months. Comparing their at-site performance: DAR outperforms DAL at an average of 75% and 79% of all sites in the US for NSE and bias across all months. This performance is the same in all meta-regions, except for a slight dip below 70% in the West. Clearly, DAR is a significantly more attractive flow transfer method than DAL.

While DAR appears to be an attractive method for standardizing flows by drainage area, there remains a concern that DAR does not account for differences in hydroclimatology. The meta-regional ranges of bias and efficiency of DAR shows that the performance degrades westward. This could be due to physical differences between the meta-regions. In an effort to understand this issue, it was hypothesized that the overall bias of DAR at a given site may be related to hydroclimatic characteristics of the index site and the ungauged site. A number of possible regressors were considered, including the ratio of drainage areas, mean, standard deviation and coefficient of variation of flows and the ratio of aridity indices. The correlation among these variables and the bias associated with the DAR method was measured with the non-parametric Kendall's Tau rank correlation. Kendall's Tau is employed because it is nonparametric and only measures monotonic correlation without assuming an explicit functional relationship.

Of all the regressors considered, only the ratio of aridity indices between the ungauged site and the index site yielded a significant ($\alpha = 0.05$), albeit weak, Kendall's Tau with a magnitude greater than 0.2. With a Kendall's Tau of 0.45, **Figure 3.7** shows the relationship between bias and the ratio of aridity indices. The relationship in Figure 3.7 is weak, but the percent bias does seem to approach zero as the ratio of aridity indices approaches unity. Breaking these aridity ratios apart by meta-region, one finds the range of the ratios expands westward. While, in all regions the ratio of aridity indices has a mean of one, the standard deviation of these ratios goes from 0.06 in the East, to 0.14 in the Midwest and 0.25 in the West. Clearly, the ratio of aridity indices is much more variable in the West than elsewhere. The sites in the West are therefore much more heterogeneous in terms of hydroclimatology. This pattern agrees with the pattern of bias seen earlier in Figure 3.2. Of further interest is that the ratio of drainage areas showed little correlation with either the Nash-Sutcliffe or bias of DAR. It can be concluded, then, that the difference in drainage areas between the index and gauged site does not play a key role in determining the performance of the DAR method, though the difference in aridity ratio between the two sites, does play a slight role, with performance weakening as hydroclimatology becomes more variable between the two sites.

Though much more research is needed, two conclusions can be drawn from this cursory examination of bias and hydroclimatology. First, DAR, with the selection of the nearest gauge as the index gauge, is not a particularly robust method in regions with highly variable climates. In addition, hydroclimatic variables may be useful in the selection of more appropriate index gauges, perhaps making the DAR a viable method in all regions. This is a result that is analogous to results presented below for other flow transfer techniques.

These results agree with the results of Emerson *et al.* (2005) in Minnesota and with those of Asquith *et al.* (2006) in Texas. Emerson *et al.* (2005) found a distinct seasonal bias in the ratio of the drainage areas in Minnesota. Both studies found that a regional parameterization of the generalized drainage area ratio could yield more promising results (Asquith *et al*. 2006). Regressing streamflow against a combination of area, precipitation and other climate ratios, Emersion *et al.* (2005) showed that the ratio of drainage areas was indeed the most significant explanatory variable considered. Though both of the above cited studies showed that regionalization and regression of streamflow and drainage area could produce coefficients of determination near 0.97 (Emerson *et al*. 2005), neither explored how this coefficient related to the practical application of the DAR method as is shown here. Left unanswered by those studies is what the bias and accuracy of a given method would be despite the high coefficient of determination.

In Texas, Asquith *et al*. (2006) generalized the form of the DA method. They argued that the common usage of the DA method assumes that the exponent of the ratio of the drainage areas is unity. They demonstrated that a more generalized DA technique would allow this parameter to vary. This, of course, introduces some calibration into the DA method. As the purpose of this study is to compare hydrostatistical methods against the most traditional technique, this project considers only the simplified and most commonly used version of the drainage area ratio. Considering these simplified methods, it is shown that the best method for standardizing by drainage area is to use a ratio of real-space flows to drainage area.

STANDARDIZING FLOWS BY MEAN STREAMFLOWS

Standardizing streamflows by drainage area using the DAR method is appealing because of both its reasonable accuracy and simple application. There is no need to characterize streamflows (Step Two) at the ungaged site and thus no further uncertainty associated with such characterization is introduced. A slightly more complex technique for standardizing flows is the use of a mean flow. If this mean flow can be estimated with some certainty, it may be that standardizing by the mean streamflow will capture the variability of flows and transfer additional hydrologic information that is not transferred by the ratio of drainage areas.

Standardization by the mean streamflow, which is common in hydrology and in flood frequency analysis, is termed the index flood method.

As with the DA methods, there are a number different variations for standardizing by mean (SM). First, one could consider either using a single annual mean to standardize flows or one could standardize streamflows using twelve individual monthly means. One of the most significant drawbacks of the drainage area ratio method is that there is no implicit correction for seasonality. That is, the relationship between the streamflow at the ungauged site and the streamflow at the index site is constant, regardless of the time of year. It was this fact that resulted in the seasonal trend associated with the bias of the DAR method. Standardization of streamflows using twelve monthly means may correct for this seasonality.

Just as the DAR technique considered the ratio of streamflow and drainage area in real space, the real-space SM methods (SMR) considers the ratio of streamflow and mean streamflow in real space. Mathematically,

$$\frac{Q_X}{\mu_{Q_X}} = \frac{Q_Y}{\mu_{Q_Y}} \qquad (3.6)$$

for the ungauged site, Y, and the index gauge, X, where Q is the monthly flow at the subscripted site and μ is the mean of the subscript. If the means are known, then the flow at the ungauged site can be estimated as

$$Q_Y = \frac{Q_X \mu_{Q_Y}}{\mu_{Q_X}} \qquad (3.7)$$

which is analogous to the estimation used in DAR, where now the ratio is of the mean streamflows instead of the drainage areas of the two sites. When only a single, annual value of μ is used at each site, this SM technique will be called annual, real-space standardization by mean (SM1R). When standardization by 12 monthly means is considered, the μ used for each site would change depending on which month is being estimated. This method is termed the monthly, real-space standardization by mean (SM12R).

The second consideration is whether to transfer flows in real or logarithmic space. Many studies have shown that monthly flows in the US are approximately lognormal. The logarithmic forms of SM (SML) consider the ratio of the logarithms of the monthly flows and their logarithmic means. This is written as

$$\frac{\ln(Q_X)}{\mu_{\ln(Q_X)}} = \frac{\ln(Q_Y)}{\mu_{\ln(Q_Y)}} \tag{3.8}$$

with the same definitions from above. Taking the logarithm of a monthly flow of zero can be avoided by solving the above equation as

$$Q_Y = Q_X^{\mu_{\ln(Q_Y)}/\mu_{\ln(Q_X)}} \tag{3.9}$$

which enables one to estimate the flow at the ungauged site *Y*. Again, if a single annual mean is used, then this method can be called the annual, log-space standardization by mean (SM1L). If one uses twelve monthly means to correct for seasonality, this method will be called the monthly, log-space standardization by mean (SM12L).

This method requires some record characterization at the ungauged site because, in practice, there is no way to know the mean streamflows at the ungauged site without such an augmentation procedure. However, this analysis focuses primarily on flow transfer techniques (Step Three) with little attention given to record characterization (Step Two) at this point. Accordingly, maximum-likelihood estimators of mean and standard deviation of the monthly flows at the ungauged sites are used for the required moments, minimizing error introduced by the characterization procedure. Thus it is assumed for the moment, that the first moment of the monthly streamflows are known at the ungauged sites; this assumption is relaxed later on. In the logarithmic case, the theory of the two-parameter lognormal distribution is used to calculate the logarithmic moments from the moments in real space. Again, the nearest gauge is used as the index gauge. In this manner, minimal error is introduced through record characterization, thus the analysis focuses solely on the performance of the flow transfer techniques.

The range of overall Nash-Sutcliffe Efficiency (NSE) and bias (B) associated with the SMR and SML methods is presented for the entire US in **Figure 3.8**. The difference between the SMR and SML methods is quite dramatic: The SMR methods are much better behaved than the lognormal analogs. The NSEs of the SMR techniques are well above zero for almost all sites, while a large fraction of the sites exhibit an NSE less than zero for the SML methods. There is very little bias introduced in the SMR method, while the bias from SML is quite large, larger than the bias seen in DAR. From this figure it is fair to conclude that the SML methods perform poorly and may be dropped from further consideration. Accordingly, this analysis will continue by examining only the SMR techniques. It may be that SML is in need of bias correction, but this will be considered later.

The SM12R method appears slightly more competitive than SM1R in terms of NSE, but the methods appear quite similar in terms of bias. Both are much less biased than DAR. The overall performance of each method can be considered by meta-region in **Figure 3.9**. Here there is no westward trend like that seen in the DA methods, but it is clear than the SMR techniques are more accurate in the East and West. Further research should explore the reason for poorer performance of SMR in the Midwest. Again, the NSEs for SM12R appear to be slightly higher than those associated with SM1R. The overall bias is nearly identical, regardless of method. It is promising that the bias associated with SMR is less than half of the bias associated with DAR. With DAR, bias was +/-50%, while here the bias is only +/- 10%. This seems to indicate that, in terms of bias, there is little distinction between the two SMR methods considered here.

The relative efficiency of SM1R to SM12R in terms of bias and NSE enables one to evaluate the relative performance of these methods, with a value greater than one indicating that SM12R is the more accurate technique, on average. With a relative efficiency of 1.68 nationally, SM12R was the more efficient technique in terms of NSE. Across all meta-regions, the relative efficiency did not drop below 1.5 for NSE. The added value of SM12R over SM1R is greatest in the West, with a relative efficiency of 2.23. This could be due to the extreme variations in hydroclimatology associated with Western basins, in which case the SM12R method corrects for more seasonality in hydroclimatology. In terms of bias, nationally the relative efficiency was 1.00. Strictly speaking, SM1R is relatively less biased in all but the West, but all values of relative bias were extremely close to unity. Accordingly, it may be worthwhile to use SM12R because of its added value to NSE and to disregard the marginal impact on bias.

The conclusions are identical when comparing the methods site by site. Overall, SM12R exhibits a higher NSE than SM1R at 79% of sites in the US. The percentage decreases slightly in the Midwest and increases in the West. Again, this could be due to the seasonality of the West, but this must be examined further. In terms of bias, SM12R outperforms SM1R at only 55% of sites. This means that the bias is similar across techniques. Overall, SM12R is therefore the preferred SMR technique.

One of the main drawbacks of the DA technique was that there was no implicit correction for seasonality. Because of this, the monthly bias of the DA methods displayed a strong seasonal trend. The monthly performance of both SMR methods can be seen in **Figure 3.10**. Only by looking at the monthly performance does the added value of the SM12R methods become apparent, in that it reduces the impact of seasonality on performance. Both SMR techniques exhibit the same drop of efficiency in the late summer and early fall that was seen in the DA methods, but the efficiency remains high for all months. The analysis of bias is much more interesting: SM1R, while the magnitude of bias is much smaller than DAR, still exhibits dramatic seasonality in the range of bias. The SM12R method, by using twelve means correct for this trend in bias. While the monthly bias of SM1R goes beyond +/- 50% at times, the monthly bias for SM12R rarely exceeds +/- 25%.

The conclusions are generally the same as above when compared by meta-region, as shown in **Figures 3.11-3.13**. What is interesting is the seasonal decrease in the NSE associated with SM1R and SM12R for each meta-region. In the East and Midwest the decrease occurs in August and September, which is traditionally the driest time of the year. In the West, this decrease occurs earlier in the summer, around July. This may indicate that the decline in monthly efficiency is due in some part to the hydroclimatic seasonality.

The monthly relative efficiencies of SM12R compared to SM1R give strong insight into their relative performance. In all months, SM12R is relatively more efficient than SM1R, averaging a relative efficiency of 1.56 for NSE. Again, the added value was greatest in the West, where the relative efficiency was over 2.00 on average. The added value of SM12R is especially remarkable for bias, with a national relative efficiency over 8.00 on average. Correcting for seasonality vastly improves the overall performance of SMR, a fact that was not readily apparent when considering the overall bias of each method. Because the aim is to create an accurate time series of monthly flows, this seasonal bias becomes extremely important.

When considering at-site performance, SM12R continues to outperform SM1R. In terms of NSE, SM12R outperforms SM1R at about 64% of sites in the US for each month. In the West, this percentage increases to about 72% for each month, which highlights the effect of seasonality in the West. In terms of bias, the comparison is even more definitive. In the US, SM12R is less biased than SM1R at about 76% of sites in each month; 82% in the West. This analysis shows that of the SM techniques, SM12R is more attractive than SM1R when the monthly means are known a priori.

As would be expected from the low level of overall bias, there was little correlation between bias and annual hydroclimatology. The same regressors used for the DA methods were considered here, namely the ratios of aridity, drainage area and streamflow statistics. Of all the significant correlations ($\alpha = 0.05$), none had a Kendall's Tau greater than 0.1. When considering correlation with overall NSE, the overall NSE of SM12R exhibited a small but significant correlation ($\tau = -0.31$) with the distance between the site and index gauge. This relationship, seen in **Figure 3.14**, is fairly weak, but it seems to suggest that as the distance grows the NSE becomes more variable. This may offer an attractive approach for selecting an index gauge for SM12R or other methods, similar to what Asquith *et al*. (2006) found for the drainage area ratio.

It is clear from this initial exploration that standardizing flows by mean flows could be a reliable method of flow prediction if those means are known a priori. While standardizing by logarithmic means showed little promise, normalizing flows in real-space showed high values of NSE with minimal bias. In conclusion, of all the SM techniques, SM12R proved the most promising, though the distinction was small. Still, SM1R may be useful in real-world applications because less record characterization is needed: estimating one mean is easier than estimating twelve. This issue is considered in Chapter Five.

Maintenance of Variance Extension Standardization

So far the flow transfer techniques discussed have attempted to standardize streamflows with a single parameter, namely the drainage area or the mean. Another commonly used method for transfer of streamflows is the Maintenance of Variance Extension (MOVE) method introduced by Hirsch (1979). MOVE standardizes streamflows with two parameters: mean and standard deviation of flows.

In 1979, Hirsch introduced a streamflow reconstruction technique he called a regional statistics method for estimating streamflow records at ungauged and shortly-gauged sites in Virginia. For this method, Hirsch hypothesized that for each month, the standardized flows at a site of interest and an index site are approximately equal. Here he uses a traditional standardization approach:

$$\frac{Q_X-\mu_X}{\sigma_X}=\frac{Q_Y-\mu_Y}{\sigma_Y} \tag{3.10}$$

where μ and σ are the mean and standard deviation of the flows at the subscripted site. Note that this standardization produces a new standardized variable with mean zero and variance one, regardless of the probability distribution of the original flows. This formulation allows for the estimation of monthly streamflows at the ungauged site, Y, as

$$Q_Y=\mu_Y+\sigma_Y\frac{Q_X-\mu_X}{\sigma_X} \tag{3.11}$$

which is an algebraic manipulation of equation (3.10). In testing this approach at two sites in Virginia, Hirsch concluded that this method was “distinctly superior to the drainage area ratio method” (1979).

Though Hirsch developed this method as a flow reconstruction technique for ungauged sites, it is rarely used as such. Soon after it was introduced, Hirsch applied it a streamflow record extension technique (Hirsch 1982), in which the method is used to extend a short record, using information from a nearby longer streamflow record. As such, there was no discussion of reconstruction at ungauged sites in the 1982 paper. As in Hirsch (1979), Hirsch (1982) found that MOVE is a more attractive method than other techniques. It is this second publication – where the technique is dubbed Maintenance of Variance Extension (MOVE) – that is cited in the literature most often. It is perhaps an accident of publication order that MOVE has been so extensively used as a record extension technique and so rarely as a record reconstruction technique.

In order to truly understand the power of the MOVE standardization as a record reconstruction technique it is important to consider four variants of the method analogous to those considered for the SM methods. Firstly, MOVE can be considered in real space, as is shown in (3.11). When this technique is applied with a single, annual mean and standard deviation, it will be called annual, real-space MOVE (MOVE1R). Additionally, it can be applied as Hirsch (1979) applied it: with twelve monthly means and standard deviations, a method that will be called the monthly, real-space MOVE (MOVE12R) here.

In contrast to the real-space approach, the logarithmic approaches may prove more accurate because of the lognormal behavior of monthly streamflows. For this case one must work with the means and standard deviations of flows in log-space, which can be estimated from real-space statistics with the theory of the two-parameter lognormal distribution. The normalization of flows then becomes

$$\frac{\ln(Q_X) - \mu_{\ln(Q_X)}}{\sigma_{\ln(Q_X)}} = \frac{\ln(Q_Y) - \mu_{\ln(Q_Y)}}{\sigma_{\ln(Q_Y)}} \tag{3.12}$$

with the same definitions as the equations above. This can be solved to predict the ungauged flows at site Y as

$$Q_Y = Q_X^{\sigma_{\ln(Q_Y)}/\sigma_{\ln(Q_X)}} \cdot \exp\left(\mu_{\ln(Q_Y)} - \mu_{\ln(Q_X)} \frac{\sigma_{\ln(Q_Y)}}{\sigma_{\ln(Q_X)}}\right) \tag{3.13}$$

which avoids the need to take a logarithm of any zero flows. As before, this approach can be used with a single mean and standard deviation in an annual, log-space MOVE (MOVE1L) approach or with twelve means and standard deviations in a monthly, log-space MOVE (MOVE12L).

The reader will recall that this examination is only concerned with the third step of the ungauged-site problem: flow transfer techniques. Again, the nearest site is used as an index gauge (Step One). In practice this method would certainly require some record characterization (Step Two), but, as was done for SM, the at-site maximum likelihood estimators of the moments will be used to characterize the record. In this way, the MOVE variants can be compared with each other and later compared against the other classes of flow transfer techniques considered earlier.

As Hirsch found (1979), MOVE performs very well. One will notice that the ranges of overall NSE and bias for the MOVE variants in **Figure 3.15** are quite different than those same ranges for DA and SM. Qualitatively, the monthly methods exhibit a better-behaved range of NSE than the annual methods. The difference between real- and log-space is much less apparent; the methods appear almost identical. In terms of bias, all of the methods exhibit a percent bias of +/- 9% and there is little distinction between the methods. None of the methods have a skewed percent bias, but the logarithmic techniques appear to have a slightly higher distribution of bias.

The meta-regional range of NSEs in **Figure 3.16** tells a similar story to the national NSE. There is little distinction between real and logarithmic techniques, but the monthly techniques perform slightly better than the annual techniques. Looking meta-regionally enables one to evaluate the behavior of MOVE across hydroclimates: as with SM, there is a slight drop in performance in the Midwest. Furthermore, MOVE seems to perform best in the West. In terms of bias, the pattern is the same in **Figure 3.17** for MOVE. The range of bias is slightly elongated in the Midwest. The use of streamflow parameters, like mean and standard deviation in SM and MOVE, appears to favor the more heterogeneous hydroclimate of the West. Where regions are very homogeneous, like in the East, all methods perform similarly.

The relative efficiencies of all the MOVE methods in terms of NSE are presented in **Table 3.1**. The entire US and each meta-region are represented by a single pane of the table. In the table, a value greater than one indicates that the method on the vertical axis is more efficient than the one on the horizontal axis. If a row in a pane has three values greater than one, then the method indicated by that row on the vertical axis is relatively more efficient than all other methods. **Table 3.2** is the same table for relative efficiency in terms of bias.

These relative efficiencies show that, for the entire US and all meta-regions, MOVE12L dominates all other methods in terms of NSE. These numbers confirm the conclusions of the boxplots, namely that the monthly techniques are far superior to the annual techniques, but there is little added-value between the logarithmic and real-space variants. If there is any added-value, the logarithmic techniques hold a slight advantage over the real-space techniques, though the relative efficiency is close to unity. Comparing the annual techniques, the relative efficiency is around 1.02, while for the monthly it is only 1.04. When comparing monthly to annual, this relative efficiency increases to about 1.85 and the results are similar in all meta-regions. The added-value of MOVE12L is most dramatic in the West, where the relative efficiencies are all greater than one.

The comparison in terms of bias is much more nuanced. Here, the most efficient method in terms of bias is the MOVE12R technique, but the advantage over MOVE12L is very slight. This may be the result of a need for a bias correction in the logarithmic transformation. The advantage of MOVE12R over MOVE12L is most apparent in the Midwest, which is to be expected, given the relatively wide bias distribution seen in the boxplots. Still, all of the relative efficiencies are very near unity. This seems to suggest that all methods are viable. It may be that the parsimony of the annual methods outweighs the added-value of the monthly techniques. This issue will be addressed in Chapter Five.

Another way to think about the at-site comparisons is presented in **Table 3.3**, which shows the percentage of sites where each method out performs all others. NSE and bias are included with the US and all meta-regions. If all methods are equal, this percentage should be about 25% for all methods. That is clearly not the case for NSE. MOVE12L shows a much higher percentage performance. In terms of NSE, MOVE12L outperforms all others at 59% of sites in the US. Not surprisingly, MOVE12R has the second highest percentage. For bias, the picture is somewhat different. The annual methods come out ahead slightly more often, but the percentages are all relatively close to 25%. This indicates that all MOVE methods are approximately equally competitive in terms of bias.

The monthly NSEs of each MOVE variant are shown in **Figure 3.18**. All of the variants show the characteristic decline in NSE that was seen in late summer and early autumn. Overall, the monthly methods show a higher range of NSE and less of a decline in late summer. The range of bias is shown in **Figure 3.19**. Not surprisingly, the annual methods of MOVE exhibit a high degree of seasonality in the bias. Bias increases in the winter months and decreases in the summer months. This trend, while still present, is much smaller and less dramatic in the monthly techniques. The bias of the monthly methods remains within +/- 25%, which may make the trend in this small degree of bias irrelevant.

The trends of monthly NSEs in each meta-region, which can be seen in **Figures 3.20-3.22**, are very similar to those seen with the SM methods. The monthly methods outperform the annual methods. The best performance is in the West, while the Midwest has the worst. What is most interesting, though, is the timing of the decrease in NSE. In the Midwest and East this decline occurs in August and September. In the West, the drop occurs earlier in the summer, mainly in July. Why this happens could be linked to the seasonality or monthly hydroclimate of the region. It may be that this decrease coincides with the driest time of the year and the beginning or end of the water year. Further research may discover a reason for this deficiency.

The trend in monthly bias, as seen in **Figures 3.23-3.25**, is similar across all meta-regions. The range of monthly bias is widest in the annual techniques of the Midwest and West, while the seasonal trend is surprisingly dramatic in the East. In all cases, the bias of the monthly techniques shows very little seasonality. Again, it may be useful to examine the changes in monthly climate that may affect this behavior.

The monthly relative efficiencies of these methods document the added-value of MOVE12L. In order not to inundate the reader with tables and figures, these monthly results have been condensed into **Table 3.4**. This shows, for each statistic and meta-region, the average monthly relative efficiency of the other methods compared to MOVE12L. A value greater than one indicates the MOVE12L is more efficient. This table also includes the number of months where the indicated method has a relative efficiency less than one, or the number of months where MOVE12L is not relatively more efficient. The average monthly relative efficiency is greater than one for all regions and methods in terms of NSE. The relative efficiency drops to 1.05 in comparison with MOVE12R, which indicates the MOVE12R is nearly as efficient as MOVE12L. MOVE12L is more efficient in all months in the East, though 12R become more competitive as one moves westward; MOVE12R is more competitive than

MOVE12L half of the year in the West. For bias, the comparison is more muddled: MOVE12L is significantly better than the annual techniques, but MOVE12R holds a distinct advantage over MOVE12L in ten months of the year. Still, the distinction is small as the relative efficiency is very close to unity.

The percentage of sites at which each method outperforms all others in terms of NSE and bias in all meta-regions can be seen in **Table 3.5**. These values are an average of the monthly percentages. For NSE, MOVE12L outperforms all other methods at 39% of sites in the US. This percentage is similar in all meta-regions. In all months, in all meta-regions, MVOE12L had a fraction greater than all others. For bias, though, MOVE12L is less competitive. Not surprisingly, the monthly techniques outperform the annual techniques in terms of monthly bias. The distinction between MOVE12R and MOVE12L is less fine. By the numbers, MOVE12R is the least biased method most often. Here, MOVE12L is not nearly as promising; MOVE12R categorically outperforms all other methods in all months. This is misleading simply because the difference between the two is so small; consider again the boxplots.

All of these comparisons involving the MOVE methods arrive at a single conclusion: monthly, log-space MOVE (MOVE12L) is, overall, the most attractive technique for transferring monthly streamflow information from an index gauge to an ungauged site. If one is more concerned with the bias of an estimated series, then it might be useful to consider MOVE12R, though all monthly MOVE techniques are similarly biased overall. The annual methods are plagued by seasonality in the monthly bias, but this seasonality is corrected in monthly methods.

Further research on MOVE should focus on the causes of bias. Using a bias correction may improve the accuracy of monthly estimates. As an initial study, the bias and efficiency of these methods was regressed with a number of hydroclimatic variables. The distance, ratios of aridity, mean, standard deviation and coefficient of variation of streamflow, and analogous ratios in logarithmic space were considered as predictors of overall bias and efficiency. Of all the significant (α = 0. 05) Kendall's Tau, few exceeded 0.10.

TRANSFERRING INFORMATION WITH FLOW DURATION CURVES

A fourth class of flow transfer techniques involves the use of flow duration curves (FDC) or streamflow distributions to transfer relative flow timing. This technique was first developed by Fennessey (1994), who dubbed it QPPQ. It was first published in the serial literature by Hughes and Smakhtin (1996). It has since been used by a number of others in a number of different applications: See Archfield and Vogel (2010), Archfield *et al.* (2010) and Mohamoud (2008) among many others. This method has the advantage of standardizing flows relative to an entire distribution rather than using only one or two parameters, as the previous three classes of methods did.

The QPPQ method assumes that the relative level of a flow occurs in the same month between two hydrologically-similar watersheds. That is, if, in month *m*, the index gauge experiences a flow that is exceeded 10% of the time, then, in that same month, the site of interest will experience a flow of a similar exceedence probability. Because the flow record is known at the index site, one can easily characterize the distribution of flows and sequence of exceedence probabilities. At the site of interest, one must estimate the probability distribution. Once this is done, for every month one recognizes the flow (Q) at an index site, determines the exceedence probability (P) at that site, transfers that probability to the distribution at the site of interest (P) and then calculates the estimated flow (Q) associated with that exceedence.

The performance of this flow transfer method is highly contingent on the selection of an index gauge. Of course, in this experiment, since Step One is held constant: the nearest gauge will be used as the index gauge. Furthermore, this method requires that the flow duration curve (FDC) be estimated at the site of interest. This process would fall under Step Two, record characterization, or estimation of an FDC at an ungauged site. In an effort to introduce little error from Step Two, the “blinded” or Jack-knifed streamflow record from the ungauged site and the flow record from the index site are both used to create empirical FDCs for each site. Thus the following analyses assume that an empirical FDC is available at both the index and the ungauged site. This assumption allows for the evaluation of the QPPQ transfer method, alone and apart from the ability to estimate an FDC at an ungauged site.

The overall performance of QPPQ in the United States is summarized in **Figure 3.26**, and is quite good. The Nash-Sutcliffe efficiencies are quite high, with more than 50% above 0.8 and 755 above 0.7. Similarly, the percent bias is well behaved, being both symmetric around zero and narrowly banded between

+/- 8%. The meta-regional performance is highlighted in **Figure 3.27**. The pattern is similar to SM and MOVE: there is a distinct drop in NSE and expansion of bias in the Midwest. Still, the method continues to perform well, with high NSE and well-behaved bias.

The monthly performance of QPPQ in the US is shown in **Figure 3.28**. Again, there is a distinct seasonal trend in NSE and bias. The NSE displays a characteristic drop off in August and September, while the bias takes on a wider range in winter months. Considering these performances by meta-region, as in **Figures 3.29-3.31**, shows the same pattern that was seen with SM and MOVE. The seasonality of bias is strong in all regions. The late-summer decline in NSE is greatest in the Midwest, while the West exhibits a characteristically earlier drop in NSE.

It is difficult to judge the performance of QPPQ here because no variants are considered for comparison. For now it is sufficient to observe that QPPQ method performs reasonably well. Later, this method will be considered against the other classes and judged accordingly.

A Comparison of Traditional Flow Transfer Techniques

Having explored four different classes of flow transfer techniques, it is important to ask which method performs the best. The above analysis found the best method from each class included the drainage area ratio in real-space (DAR), the monthly, real-space standardization by mean (SM12R), the monthly, lognormal MOVE (MOVE12L) and QPPQ. Considering these four methods together will show that one of them is generally preferred over the others.

The reader will recall that this is merely a comparison of flow transfer techniques. As such, this comparison is an idealized experiment that does not require streamflow characterization procedures (Step Two). In practice, the record characterization process will introduce additional uncertainty that could change the relative performance of these methods. This issue is addressed later in Chapter Five.

The overall performance of the four methods in the United States is summarized in **Figure 3.32**. This analysis assumes that one can characterize the streamflow record with a high degree of certainty in which case the SM, MOVE and QPPQ all behave more favorably than DAR in terms of NSE and bias. Compared to the other three methods, DAR yields a much lower range of NSE and a range of bias that is almost five times as large. The meta-regional comparison of NSE and bias can be seen in **Figures 3.33** and **3.34**. In all of the methods that require record characterization (SM, MOVE and QPPQ), performance degrades in the Midwest. For DAR, it is the West that performs most poorly. While all of the methods show high levels of NSE and well-behaved bias, it is interesting that the performance DAR is significantly below that of the others. This is especially concerning, as DA methods are considered the most common flow transfer methods (Archfield and Vogel 2010).

The previous figures provide some evidence that DAR is not the best flow transfer technique, but little can be said of the distinction between the other three. **Table 3.6** summarizes the relative efficiency of NSE for the entire US and each meta-region. (Recall that, within a pane, a row that contains three values greater than one demonstrates that the method indicated by that row out performs all others.) In all regions, it is the MOVE technique that provides some added-value over the other three methods. On average, MOVE has a relative efficiency over 2.00. The advantage of MOVE is particularly strong in the West. The distinction is much less fine in the consideration of bias, shown in **Table 3.7**. Here, MOVE continues to hold a slight edge, but the comparison with SM and QPPQ is much closer. For SM and MOVE, the distinction is almost irrelevant because the relative efficiency is so close to unity.

When the methods are compared site-by-site, MOVE remains the best performer in terms of NSE, but SM appears most competitive in terms of bias. The percentage of sites where each method outperforms all others is presented in **Table 3.8**. The MOVE method has the highest NSE at 59% of sites across the US. The distinction between MOVE and the other methods is particularly strong in the West, where MOVE had the highest NSE at 68% of sites. For bias, SM out performs MOVE, but the advantage is much smaller. In the entire US, SM is less biased than all other methods at 38% of sites, while MOVE is the least-biased method at 31% of sites. This narrow performance of bias can be understood by returning to the boxplots above. One can see a clear distinction between SM, MOVE and QPPQ when it comes to NSE, but for bias the difference between the boxplots is much harder to see. So, while the MOVE technique is the best method for overall NSE, any method other than DAR exhibits very low bias.

As was mentioned above, it is important to consider the monthly performance of these methods in addition to the overall performance. The monthly performance of the four methods can be seen in **Figures 3.35** and **3.36**. Not surprisingly, the first figure, considering NSE, highlights the relative poor performance of DAR, but all methods show the late-summer drop-off in NSE. For bias, all of the methods have some degrees of cyclical seasonality, where the variability of bias increases in the winter months. This seasonality is most dramatic for DAR and QPPQ, which do not explicitly correct for monthly variability. Relatively, it appears that MOVE and SM exhibit the least monthly bias. The meta-regional range of NSE for each method can be seen in **Figures 3.37-3.39**. It is extremely interesting that, for all methods, the decline in NSE is shifted earlier in the year for the West. Regardless, it appears that the MOVE method is the strongest competitor in all meta-regions. For bias, in **Figures 3.40-3.42**, it is clear the DAR and QPPQ are not the best techniques.

As the MOVE technique appears to be the most advantageous, **Table 3.9** summarizes the mean relative efficiency of NSE and bias across all months. A value greater than one indicates the MOVE maintains some advantage over the other method. Additionally, this table provides the number of months where MOVE is not the top performer. In terms of NSE, MOVE outperforms all other methods in every month. For bias, MOVE and SM are about on par. The relative efficiency between MOVE and SM approaches unity in all regions, and SM is actually the better performer in half of the months.

The average fraction of sites where each method is the least biased in a month is displayed in **Table 3.10**. Again, MOVE holds a clear advantage in terms of NSE, but SM is the least-biased method. For the entire US, SM outperforms MOVE in all months. This is strong evidence of for the use of MOVE, unless one is extremely concerned about precise bias. Still, it should be noted that the difference between the bias of SM and MOVE is quite marginal.

When all is said and done, the most important comparison is an observation of how well each method outperforms the most common technique, the drainage area ratio. Nationally, MOVE, which led all methods, had a greater NSE than DAR at 86% of sites and was less biased at 88% of sites. **Figure 3.43** shows the fraction of sites where each method outperforms DAR by meta-region. When consider by meta-region, MOVE continued to outperform all other methods. On average, MOVE had a greater monthly efficiency than DAR at 78% of sites, being less biased at 83%. This trend was replicated across each meta-region, though SM proved competitive in terms of bias, as in **Figure 3.44**.

Of all the traditional techniques considered here, the best overall flow transfer technique is the monthly, lognormal variation of MOVE. Of course, this is only true if minimal uncertainty is introduced through record characterization. That is to say, this analysis has only examined the third part of this problem of estimating monthly time series at ungauged sites. In the next chapter, some consideration will be given to the methodology behind record characterization (Step Two) and then record characterization will be used to simulate a real-world application of the flow transfer techniques.

IV. Regional Regression and Record Characterization

In the previous analysis of flow transfer techniques, there was little consideration given to the characterization of streamflow records, or estimation of streamflow moments at the ungauged sites. In an effort to reduce the introduction of uncertainty, at-site maximum likelihood estimators of the required streamflow moments were used to characterize the streamflow records. This allowed for an equitable comparison of the various techniques. In practice, however, it is not possible to use maximum-likelihood estimators at an ungauged site. In this chapter, the issue of record characterization at the ungauged site is considered.

Regional regression has often been used to estimate streamflow parameters for use in flow transfer techniques (Archfield *et al.* 2010; Hirsch 1979 and Vogel *et al.* 1999). In general these studies have focused on the prediction of annual streamflow moments analogous to those required in the 1R and 1L approaches discussed above. Hirsch (1979) uses regional regressions that depend on a wide range of explanatory variables, some of which are not readily available in many regions across the globe. In an effort to develop a method that may be applicable beyond the US, this effort extends the works of Vogel *et al*. (1999), hypothesizing that regressions able to predict monthly streamflow moments, can be developed from easily-accessible climate variables.

Vogel *et al.* (1999) showed that the real-space mean and variance of streamflow could be estimated regionally based on drainage area combined with simple climate variables using multivariate regression. Their regression equations were remarkably precise with R^2 values usually well above 90% for most regions of the US for estimation of both the mean and variance at ungauged sites. This same approach can be extended to estimate monthly streamflow moments, whether they are real-space or lognormal moments.

In the prediction of monthly streamflow moments there are at least two avenues to consider: predicting each monthly moment independently or predicting all means with a single equation. The first approach develops a single, recursive equation for each month and initializes them with a thirteenth non-recursive equation. The second recursive approach requires a single, equation to predict each moment. This single, recursive relationship can then be initialized with the same non-recursive equation used above.

The number of regression equations required for each will distinguish these two techniques. The independent monthly regression approach uses thirteen equations, while the other uses only two. In addition, each of these methods can be applied to estimate moments in real-space or moments in log-space; thus overall, there are four methods. The ability of these methods to accurately predict streamflow moments will be summarized below. Finally, a brief summary of the performance of the regressions from Vogel *et al.* (1999) will be shown. Selecting the best techniques for record augmentation will allow for a true real-world coupling of streamflow record augmentation and flow transfer techniques which follows in Chapter Five. Of course, the best technique is largely dependent on the flow transfer technique used: SM1R only requires equations from Vogel *et al.* (1999), while MOVE12L requires a set of 12 new regressions for estimation of each of the monthly mean and variances at the ungauged sites.

A general methodology will be used for comparing regression estimators, as was discussed earlier. This assessment depends largely on leave-one-out or jack-knife experiments. To summarize, a site of interest is selected and regional regressions are developed without any data from that site. The regressions are then used to predict the streamflow parameters at that site. The accuracy of those predictions is then assessed in terms of Nash-Sutcliffe efficiency (NSE) and bias. This process is then repeated for every site being considered.

THE DEVELOPMENT OF REGIONAL REGRESSIONS WITH TWO EQUATIONS

The parsimony of the two-equation approach (Reg2) is attractive. All else being equal, one would much prefer a solution of two-equations rather than thirteen. For this reason, the two-equation approach will be developed first for real-space means. The same techniques can be used for estimating real-space variances by changing the left-hand side of the equations.

The main thought for this approach is simply that for each two-digit HUC region, *r*, there exists a function, *f*, such that the monthly mean flow, $\mu_{Q_{m,r,i}}$, for month *m* and site *i* can be related to watershed characteristics via

$$\mu_{Q_{m,r,i}} = f\left(A_i, \mu_{P_{m,r,i}}, \mu_{PET_{m,r,i}}, \mu_{T_{m,r,i}}, \sigma^2_{P_{m,r,i}}, \sigma^2_{PET_{m,r,i}}, \sigma^2_{T_{m,r,i}}, AI^{-1}_{m,r,i}\right) \qquad (4.1)$$

where $\mu_{X_{m,r,i}}$ and $\sigma^2_{X_{m,r,i}}$ are the mean value and variance of *X*, respectively, in month *m*, region *r* and at site *i,* and A_i is the drainage area associated with site *i*; *P* is precipitation, *PET* is potential evapotranspiration, *T* is temperature and *AI* is the aridity ratio. The aridity ratio is defined as the ratio of average precipitation to the average potential evapotranspiration. Some the variables included here may not lead to significant improvements to the model and in such instances they will be excluded through the diagnostic model-building process outlined below.

While the equation (4.2) may be plausible, not all hydrologic information can be transferred by climatic variables alone. Here, it is argued that the inclusion of a flow lag of one month would carry over the majority of basin characteristics that cannot be captured by climate variables alone. This causes (4.1) to become a recursive equation, *R*, such that

$$\mu_{Q_{m,r,i}} = R(\mu_{Q_{m-1,r,i}}, A_i, \mu_{P_{m,r,i}}, \mu_{PET_{m,r,i}}, \mu_{T_{m,r,i}}, \sigma^2_{P_{m,r,i}}, \sigma^2_{PET_{m,r,i}}, \sigma^2_{T_{m,r,i}}, AI^{-1}_{m,r,i}) \tag{4.2}$$

where all variables and notations are similar to (4.1). It is possible to compute all monthly means with *R* alone, but the task becomes one of solving a system of recursive linear equations that may or may not be convergent. However, to start the recursion in (4.2) one can consider an initializing equation, I^*, such that

$$\mu_{Q_{r,i}} = I^*(A_i, \mu_{P_{r,i}}, \mu_{PET_{r,i}}, \mu_{T_{r,i}}, \sigma^2_{P_{r,i}}, \sigma^2_{PET_{r,i}}, \sigma^2_{T_{r,i}}, AI^{-1}_{m,r,i}) \tag{4.3}$$

where I^* is simply the best regression obtained for predicting the moment of streamflows across all months and m^* is the month with the best 'at-site' regression. The equation I^* is thus an initializing equation for *R*, allowing for the computation of all monthly means without the solving of a recursive equation.

Analogous to the work of Vogel *et al.* (1999), the functional form of both *R* and I^* is the multivariate "power law" model

$$\hat{\mu}_{Q_{m,r,i}} = e^{\beta_0}\mu^{\beta_1}_{Q_{m-1,r,i}} A_i^{\beta_2} \mu^{\beta_3}_{P_{m,r,i}} \mu^{\beta_4}_{PET_{m,r,i}} \mu^{\beta_5}_{T_{m,r,i}} \left(\sigma^2_{P_{m,r,i}}\right)^{\beta_6} \left(\sigma^2_{PET_{m,r,i}}\right)^{\beta_7} \left(\sigma^2_{T_{m,r,i}}\right)^{\beta_8} e^{\beta_9 AI^{-1}_{m,r,i}} \tag{4.4}$$

where the values of β are the coefficients fit for each region. Similarly, I^* can be written as

$$\hat{\mu}_{Q_{m^*,r,i}} = e^{\alpha_0} A_i^{\alpha_1} \mu_{P_{m^*,r,i}}^{\alpha_2} \mu_{PET_{m^*,r,i}}^{\alpha_3} \mu_{T_{m^*,r,i}}^{\alpha_4} (\sigma_{P_{m^*,r,i}}^2)^{\alpha_5} (\sigma_{PET_{m^*,r,i}}^2)^{\alpha_6} (\sigma_{T_{m^*,r,i}}^2)^{\alpha_7} e^{\alpha_8 AI_{m,r,i}^{-1}} \tag{4.5}$$

where the values of α are the coefficients for the regression of month m^*, the identification of which will be explained in detail below. When predicting a real-space moment with a power-law model, it is necessary to correct for the bias introduced by the logarithmic transformation. This bias is corrected by multiplying the (4.4) and (4.5) by the bias correction factor, *BCF*,

$$BCF = e^{\frac{\sigma_p^2}{2}} \tag{4.6}$$

where σ_p^2 is the variance of the prediction errors associated with the regression developed. The variance of prediction errors is the quotient of the prediction sum of squares and the degrees of freedom of the regression.

Equations (4.4) and (4.5) present the functional relationship of monthly mean flows being considered here. It remains only to illuminate how the coefficients of each function are selected. In principle, simple linear regression techniques can be used to estimate the parameters of these functions with a logarithmic transformation. In order to ensure that only significant variables are considered, the data will undergo a series of screenings before the parameters are appropriately estimated. Here, this regression technique will be referred to as tri-step regression.

As its name implies, tri-step regression uses a three-step regression diagnostic technique: two filters to remove insignificant coefficients or variables and one to estimate the values of the significant coefficients. For all steps, a 5% significance level is used to evaluate the significance of estimated coefficients. First, all possible explanatory variables are considered using traditional stepwise regression. This process identifies the insignificant variables, which are removed from the analysis at this stage. Multivariate linear regression is then applied to the remaining variables and the most insignificant coefficient, is removed. This step is repeated until all coefficients are significant. Finally, with the final set of variables, weighted least-squares regression is used to estimate the values of each coefficient. Each dependent observation is a sample moment, each with different record length, thus the length of the record used to estimate its moment weights the residuals associated with each ungauged site.

Tri-step regression ensures that only significant variables are considered for estimating the monthly moments in a region. The residuals of the final regression from this process are assessed for normality and homoscedasticity. With a calibrated recursive and initializing equation, R and I^* respectively, for each region it is possible to estimate the streamflow moments in each month at an ungagged site from a number of elementary climate statistics. The same methodology that was developed above for the estimation of streamflow means can be used to develop regressions for estimating monthly variances.

For predicting the moments of the logarithms of streamflow at an ungauged site, the methodology is the same but the functional form of the equations is slightly different. For real-space statistics, the functional forms of (4.2) and (4.3) were power-law models. With lognormal moments, the functional forms of (4.2) and (4.3) become linear summations of the logarithms of the independent variables, except for the aridity index, which remains in real-space. Despite this change in functional form, the coefficients of the equations can be estimated in the same fashion. In this manner, the monthly lognormal streamflow moments can be reasonably estimated from elementary climate variables.

It would prove extremely elegant if all the monthly moments could be estimated with only two equations per moment. But before this fact is tested, it is important to develop the alternative method: that in which each month is regressed individually, resulting in thirteen equations. Only once both approaches have been developed will one be able to determine the best approach for predicting streamflow moments.

THE DEVELOPMENT OF REGIONAL REGRESSIONS WITH THIRTEEN EQUATIONS

The second approach for regional regression of streamflow moments requires a few more regression equations per variable than Reg2. Here, a recursive regression equation is developed for each month and then the initializing equation, which will prove identical to that used in Reg2, will be used to estimate the moment in the first month. This results in thirteen equations; for the sake of clarity, this approach will be called Reg13. While less parsimonious than Reg2, it may be that a single recursive equation cannot capture the true seasonality of streamflow moments and thus Reg13 may prove more accurate. In such a case, Reg13 would capture this seasonality explicitly, analogous to the distinction between annual and monthly flow transfer techniques, i.e. SM1R and SM12R.

The development of a suite of thirteen equations to be used to estimate real-space means of streamflow is presented below. As with Reg2, the process can be replicated for either mean or variance by changing that left-hand side of the equations. For lognormal moments, the functional form of the equations will also be slightly altered, but the parameterization of the functions is the same. The regression equations for real-space moments follow a power-law, as with equations (4.4) and (4.5); for lognormal moments the regressions take the form of a linear summation of the logarithms of the independent variables, except for the aridity indices, which remain in real-space. It is also important to recognize that, when predicting real-space moments, each of thirteen equations has its own bias correction factor identical to the form presented in (4.6).

Several explanatory climate variables were considered in each regression: the mean and variances of monthly precipitation, temperature and potential evapotranspiration and the monthly aridity ratio. This means that for each month, *m*, in a region, *r*, there exists a unique equation

$$\mu_{Q_{m,r,i}} = R_m(\mu_{Q_{m-1,r,i}}, A_i, \mu_{P_{m,r,i}}, \mu_{PET_{m,r,i}}, \mu_{T_{m,r,i}}, \sigma^2_{P_{m,r,i}}, \sigma^2_{PET_{m,r,i}}, \sigma^2_{T_{m,r,i}}, AI^{-1}_{m,r,i}) \quad (4.7)$$

for the mean at a given site, *i*, where *A* is the drainage area, *P, PET,* and *T* are precipitation, potential evapotranspiration and temperature and *AI* is the aridity ratio. The distinction between (4.7) and (4.2) is the index of (4.7). Equation (4.7) is exclusive to a given month, meaning that there are 12 versions of (4.7) per moment, while there is only one version of (4.2).

Use of (4.7) assumes that an estimate of the mean streamflow is available from the previous month, which will not be true for the initial month, whichever month that may be. A system of 12 recursive equations could be solved simultaneously, but this assumes that the solution is convergent, which may not be the case. Instead, the initializing month and equation, as was introduced earlier, can be used to provide the moment of the first month. Because the process is the same, the initializing equation for Reg13 will be identical to that developed for Reg2. Using this equation to accurately estimate the mean for the initial month, m^*, makes it possible to use the recursive equation (4.7) to estimate the means of all the subsequent months, including a re-estimate of the initialized month. To allow for the greatest level of accuracy, the process of estimating the means with recursive equations was conducted iteratively by repeatedly using the 12 recursive equations until the final estimates were less than 0.01% different than the previous estimates.

The coefficients of these thirteen equations can be fitted using the exact same technique that was used for Reg2: tri-step regression. Recall, tri-step regression ensures that only significant variables are considered for estimating the monthly means in a region. The residuals of the resulting regressions are assessed for normality and homoscedasticity, and if those assumptions are not violated, the resulting model can be advocated for use.

Relative Performance of Regional Regression Methods

The performance of each of the regional regression techniques presented above can be evaluated with a set of leave-one-out experiments. As was explained earlier, a leave-one-out experiment blinds a single site, develops the regressions from data that does not include the blinded site and then uses those regressions to estimate the moments at the blinded site. These estimated moments can then be compared against the maximum-likelihood moments calculated from the blinded flow record. This process can be replicated for each individual site.

From above, there are four methods to consider: Reg2 for predicting real-space moments, Reg2 for predicting log-space moments, Reg13 for predicting real-space moments and Reg13 for predicting log-space moments. Respectively, these can be abbreviated as Reg2R, Reg2L, Reg13R and Reg13L. The overall performance, including NSE and bias, of these methods for predicting means and variances is presented in **Figure 4.1**. From this presentation, it is clear that Reg13R provides the best overall estimates of mean, though Reg13L is less biased. For both estimating the mean and variance, the Reg13 methods perform better than the Reg2 methods. The choice between Reg2 and Reg13 can be solely based on the performance, but the choice between real-space and log-space techniques is dictated by the flow transfer technique.

When using a 12R flow transfer technique like SM12R or MOVE12R, then the choice of which regression reduces to Reg2R versus Reg13R. In all cases, the relative efficiency of Reg13R over Reg2R was well over one, indicating that Reg13R is a preferred to Reg2R for estimation of the real-space moments.

Reg13R predicted means more accurately than Reg2R at 86% of sites and was less biased at 60% of sites. In terms of bias, Reg13R yielded a higher NSE at 82% of sites and was less biased at 65% of sites in the US. As could be surmised from the boxplots, Reg13R is the more attractive real-space regression technique.

When using a 12L flow transfer technique, one cannot use a real-space regression technique. Instead, one must distinguish between Reg2L and Reg13L. In terms of predicting log-space means, Reg13L had a relative efficiency of 3.1 for NSE and 1.4 in terms of bias. For variances the relative efficiencies are 1.4 and 1.5, respectively. This shows that Reg13L has a slight advantage over Reg2L. At 83% of sites, Reg13L predicted the mean more accurately than Reg2L. Reg13L was less biased at 59% of sites. For variance, Reg13L outperformed at 74% and 61% of sites respectively. Again, this is strong evidence that the Reg13 techniques are favored over the Reg2 techniques. It should be noted, though, that the ability of either method to predict variances is quite weak.

As will be discussed later, the error introduced by a monthly regression technique may outweigh the performance of the given flow transfer technique. In such a case, a 1R flow transfer technique may outperform a 12R flow transfer technique. If only a single, annual mean and variance is required, then the regressions of Vogel *et al*. (1999) can be used to predict real-space moments and, with the theory of a two-parameter lognormal distribution, the moments of the logarithms are easily estimated. Vogel *et al*. (1999) provides the full development of these regressions. Across the entire US, the Vogel *et al*. (1999) regressions predicted annual means with an NSE of 0.9752 and a bias of 11%. With an NSE of 0.6254 and a bias of 71%, the results are less accurate for variance. Still, these regressions may prove useful in the next chapter.

V. ESTIMATING STREAMFLOW IN UNGAUGED WATERSHEDS

When one wishes to estimate time-series of monthly streamflows at ungauged sites in the real world, one must consider very carefully each of the three steps outlined above: (1) The selection of an index gauge, (2) the characterization of the streamflow record at the ungauged site and (3) flow transfer method. Previous chapters have considered idealized experiments that were designed to shed light on Steps Two and Three individually. Having documented the performance of several flow transfer techniques and explored possible methods for record characterization, it is appropriate to ask how these methods would truly perform at an ungauged site.

The following experiments will be conducted so as to reflect real-world applications. For each site, the three steps will be executed as if there is no information on the streamflow at the ungauged location. The nearest gauge will continue to be used as the index gauge. Record characterization will be executed with use of leave-one-out applications of monthly regression with thirteen equations (Reg13) or with the regressions for the mean and variance of streamflows from Vogel *et al.* (1999), as applicable. Finally, the moments estimated in Step Two will be used with the different streamflow transfer methods. The length of the estimated time series was the length of the overlapping records between the index gauge and the site of interest. The accuracy of these estimates will then be estimated with Nash-Sutcliffe efficiencies (NSE) and overall bias (B). All methods are compared with traditional drainage area approaches.

Standardizing by Mean Streamflow with Regional Regression

Standardizing by mean streamflows was considered to be an attractive flow transfer technique because it required minimal record augmentation at the ungauged site and accounted some measure of seasonality. It was previously shown that the real-space, monthly standardization (SM12R) held a slight advantage over the annual equivalent (SM1R). Here, because the uncertainty of record characterization has been introduced, it is appropriate to reconsider both methods. It may be that the added uncertainty characterizing the records with regional regression degrades the performance of SM12R.

The national, overall performance of the SM methods with regional regression can be seen in **Figure 5.1**. In general, the distributions of Nash-Sutcliffe efficiencies are quite similar. SM1R appears to exhibit slightly more upward bias, meaning the estimates are, on average, slightly less than the observed values. As would be expected, the NSE associated with these estimates is less than the NSEs associated with the idealized application of SM in previous chapters. Similarly, the bias here is dramatically more apparent here than with the idealized experiments.

When the overall performance is broken up by meta-region, as in **Figure 5.2**, the degradation of performance and bias becomes immediately apparent. As with the idealized experiment, the NSE is lowest in the Midwest, though the Midwest and West are quite similar here. Not surprisingly, both methods perform well in the East. The bias, on the other hand, tells a different story. In both methods, the bias increases westward. In the idealized example, bias was always symmetric about zero and ranged from +/- 10%. When the uncertainty of record characterization is introduced, the bias increases to a range of +/- 80%. Furthermore, the bias is no longer symmetric about zero. Relatively, the breakdown by meta-region confirms that SM1R is outperforming SM12R in terms of NSE, though the bias is concerning.

Overall and in all meta-regions, the relative efficiency of SM1R was greater than one in terms of NSE and bias. For NSE, the relative efficiency was greatest in the East (1.45) and smallest in the Midwest (1.23). Considering bias, the relative efficiency decreased westward from 1.78 in the East to 1.10 in the West. Overall, SM1R had greater Nash-Sutcliffe efficiency than SM12R at 64% percent of sites in the US. It was less biased at 54% of sites. Interestingly, SM1R showed a greater NSE than SM12R at only 52% of sites in the West, compared to only 67% and 68% in the East and Midwest. This shows that SM1R shows only a marginal advantage over SM12R in the West. The comparison is less striking for bias: While both methods showed an inflated bias westward, the percentage of sites where SM1R was less biased remained at 55% in the East and Midwest. In the West, SM1R only outperformed at 48% of sites.

The monthly performance of these two methods, as seen in **Figure 5.3**, shows that both methods are plagued by seasonal variation in NSE and the range of bias. As can be seen in **Figures 5.4-5.6**, this inaccuracy is minimal in the East and increases westward. The seasonal trend was evident in the idealized consideration of both methods. Interestingly, only SM1R showed a pronounced seasonal trend in the range of bias; here, SM12R also shows some seasonal variation. The seasonal variation arises from the error introduced by the record characterization method. Interestingly, this uncertainty leads the range of bias in SM12R to exceed that seen in SM1R. This would lead one to advocate for the use of SM1R, though one must also consider the median bias as well. For SM1R the bias is slightly positive in all months in the Midwest and West; on the other hand, SM12R shows a bias that is symmetric about zero.

When compared site-by-site, SM1R emerges as the most competitive method. Across all the months the relative efficiency of SM1R over SM12R averaged about 1.53 for NSE. From East to West this advantage increased from 1.43 to 1.49 and 1.91. The relative efficiency only dropped below unity in June nationally (0.9997) and in the Midwest (0.9868). By month, SM1R showed a higher NSE than SM12R at more than 50% of the sites for all months, averaging 59% nationally. In the East this percentage dipped to 58% and increased elsewise to 61% in the Midwest and 57% in the West. This reflects the increased advantage of SM1R in the Midwest and West.

Site-by-site, monthly bias showed a similar trend to NSE. The national, average monthly efficiency of SM1R over SM12R in terms of bias was 1.37. The advantage was greatest in the East (1.62) and decreased westward to 1.44 and 1.22. This relative efficiency dropped below unity in June (0.92) nationally, once in the Midwest (June, 0.994), and twice in the West (May, 0.956; June, 0.795). Excepting June in the West, these values are all nearly unity, meaning that SM12R rarely provides an advantage over SM1R. As with NSE, SM1R was less biased than SM12R at more than 50% of sites in all months, nationally averaging

57% of sites. This average percentage fluctuated, but was generally constant across the three meta-regions. The West was the only region where the percentage dipped below 50% in any months. In May, June, July and September in the West this percentage dropped to an average of 49.5%. These results show that, in terms of bias, both methods perform nearly equally, though SM1R holds a slight advantage.

When records are characterized with regional regression, this analysis has shown that real-space annual standardizing by mean is a better flow transfer technique than the monthly parallel. In fact, SM1R had a greater NSE than DAR at 48% of sites, while SM12R outperformed DAR at only 36% of sites nationally. There is some positive bias introduced monthly, but the bias remained generally smaller for SM1R compared to SM12R. Both methods showed a clear relative deficiency in summer months. The idealized results showed that this was a result of the flow transfer technique itself and not the coupling with regional regression. In future work it may be necessary to better understand the seasonal trend in NSE that spanned all flow transfer techniques.

Maintenance of Variance Extension with Regional Regression

Previously it was shown that using Hirsch's (1979) maintenance of variance extension (MOVE) as a standardization and flow transfer technique was the most competitive idealized technique. The difference between real-space and log-space MOVE methods was minimal, as was the difference between annual and monthly MOVE methods. In the idealized examples, this comparison was conducted with maximum likelihood moment estimators. Here, the record characterization will be conducted in a real-world, practical sense with regional regression, as was done with SM above. It is important to recognize that because MOVE uses two parameters for standardization, uncertainty is introduced through both the regression of means and the regression of variances. Because of the uncertainty of regional regression, all MOVE methods will be considered anew. Those four methods are annual and monthly real-space methods (1R and 12R) and annual and monthly lognormal methods (1L and 12L).

The overall Nash-Sutcliffe efficiencies of the four variants are presented in **Figure 5.7** for the US and each meta-region. From the first pane, it is clear that the monthly methods exhibit a higher trend in NSE. Between the monthly methods, MOVE12R may have a greater magnitude, though MOVE12L has a slightly smaller range. In both the East and West, MOVE12R maintains a slight advantage over MOVE12L, but in the advantages of MOVE12L are greatest in the Midwest. There may be some regional difference associated with the Midwest that favors MOVE12L. In all cases, these results are far inferior to those seen in the idealized experiments of Chapter Three, where the lower extremity of NSE was 0.5.

The overall bias of each variant is presented in **Figure 5.8**. For the entire US, the annual methods show a slight positive bias, while the bias of monthly method is nearly symmetrical about zero. In general, MOVE12R exhibits the best behavior in terms of bias. This result, though, varies across the meta-regions. In the East, all the variants perform well, with a small range of bias that is symmetrical about zero. In the Midwest, MOVE12R continues to yield well-behaved bias, while all other variants show a slight positive bias. All variants are positively biased in the West. As with NSE, these results reflect the uncertainty of regional regression, being strikingly different than the results of the idealized experiments. In the idealized case, the bias ranged to +/- 8%, which is a nearly an order of magnitude lower than the +/- 70% seen here.

Similar to Table 3.1, **Table 5.1** reports the relative efficiencies of NSE for the four MOVE methods with regional regression. Recall that, within a panel, a row with three values greater than zero indicates that the method on that row is superior to all others. In terms of NSE, MOVE12L holds a significant advantage over all other methods. Nationally, the relative advantage averages to 1.65. The advantage is greatest in the East (2.09) and West (1.75), while it dips to 1.33 in the Midwest.

The relative efficiencies of bias are presented in **Table 5.2**. For bias, it is actually MOVE1R that provides the greatest advantage. Nationally, the advantage of MOVE1R over all others averages 1.21 compared to MOVE12L's advantage of 0.84. MOVE12R is more competitive than MOVE12L, averaging 0.97 nationally. In general, the annual methods are relatively more efficient, in terms of bias, than the monthly methods. Head-to-head, MOVE1R has a 1.19 advantage over MOVE12R and a 1.32 advantage over MOVE12L. MOVE1R provides the greatest advantage in the East and decreases westward. These results show that MOVE1R is actually significantly less biased than other methods, on the average.

When considered site-by-site the advantage of MOVE12R in terms of NSE becomes quite apparent. The percentage of sites where each method outperforms the others is presented in **Table 5.3**. If all methods were performing equally, one would expect the percentage to be 25%. Across the US, MOVE12R exhibited a greater NSE than all other methods at a 48% of sites. In terms of bias, MOVE12R only outperformed at 31% of sites while MOVE1R outperformed at only 32% of sites. While this is strong evidence for the equality of the two methods, it is useful to consider the methods compared against each other: MOVE1R was less biased than MOVE12R at 56% of sites, nationally. When considering bias, the annual and monthly real-space methods perform similarly. MOVE12R, on the other hand, provides a significant advantage over MOVE12L in terms of overall bias.

The monthly performance of each method highlights the varying affect of seasonality on each of the methods. The NSEs are summarized in **Figure 5.9**. These NSEs are starkly inferior to those seen in the idealized experiments with MOVE. The annual methods are characterized by lower NSEs and a steady decline in efficiency from January through to an increase from August onward. This trend is much different than the late-winter and late-summer deficiencies that characterized the idealized results and presents their selves more dramatically in these monthly results. The range of NSEs is smaller for the annual methods, but the median efficiencies are greater for monthly methods.

Considering monthly bias, as in **Figure 5.10**, shows the reason for the new trend in annual results. While both the monthly and annual results show some distinct seasonality of bias, the two are quite different. As would be expected when using a single mean instead of twelve, the annual methods tend to overestimate flows in the fall and winter, while under estimating flows in the summer. The monthly methods, on the other hand, show a varying range of bias, but the bias remains symmetrical about zero. On average, there are no months where the monthly methods over- or under-estimate the streamflows. For this reason, it appears the monthly methods are the best technique for accurately predicting streamflow time series.

The same patterns observed nationally replicated themselves in varying degrees across the three meta-regions. The NSEs are generally highest in the East and degrade westward, as can be seen in **Figures 5.11-5.13**. As shown in **Figures 5.14-5.16**, bias is most extreme in the West and Midwest. Interestingly, in the Midwest, even MOVE12L starts to show some positive bias. This may support the monthly, real-space technique over the lognormal one.

From an analysis of the monthly NSE and bias, the choice for the best MOVE method is between the monthly methods. From Tables 5.1 and 5.2, one may recall that MOVE12L showed a relative advantage of 1.11 for NSE and 0.90 for bias against MOVE12R. Monthly, MOVE12L showed an average advantage of 1.07 in terms of NSE. Nationally, MOVE12L showed a relative efficiency less than one in only four months. In the East the advantage was 1.10 and was less than one in only two months. In the West, though, the relative efficiency dropped below one in eight months and averaged 0.98. But, when the methods are compared on a site-by-site basis, MOVE12L shows a greater NSE at more than

50% of sites, nationally, for six months, averaging 50%. Across all meta-regions the average remained above 50%, except in the West where is dipped to 47%. MOVE12L did not exceed 50% in three in half of the years in the East and Midwest. These results show that MOVE12L is a strong flow-transfer method, though it may not be the strongest in the Midwest. In future research, it may be useful to consider any hydroclimatic reasons for the varying performance of each method.

The distinction between MOVE12R and MOVE12L is less clear in terms of bias. The average monthly advantage of MOVE12L is 1.04 nationally and dips below one in five months of the year. The advantage is relatively constant across meta-regions, dipping to 1.00 in the West. Site-by-site, MOVE12L is less biased than MOVE12L at more than 50% of sites in only two months. Nationally, MOVE12L outperforms MOVE12R at 48% of sites on average. This percentage ranges from 50% in the East to 48% in the Midwest and 43% in the West. Though there is little competitive advantage, the site-by-site results show that MOVE12R remains the most competitive method for providing unbiased flow time series.

After all the methods have been considered, it is actually the monthly, lognormal variant of MOVE that, when coupled with regional regression, provides the best MOVE flow-transfer technique. In the idealized results, it was argued that MOVE12L was the best method as it significantly outperformed MOVE12R. Here, though the uncertainty introduced by regional regression degraded the accuracy of MOVE12L, MOVE12L outperforms DAR at 36% of sites while MOVE12R outperforms at only 33%. By this metric, MOVE12L is the best regionally-characterized MOVE flow transfer technique.

One may also wonder that the annual results faired so poorly, while they were quite successful when standardizing by mean. Looking back at the results from the analysis of regional regression, the source of uncertainty comes from the estimates of variance. Vogel *et al.* (1999) estimated means with a high degree of certainty, but the estimation of bias, despite the high regression statistics they reported, showed an NSE of 0.62 with a 71% bias. Furthermore, the regional regression methods predicted variances with a much lower efficiency than means across the board, though the real-space regressions predicted variance more accurately than the lognormal regressions. This uncertainty of variance is clearly propagated through the flow-transfer technique. It is no surprise then, that the monthly, lognormal method was degraded to a point that the real-space method was nearly as competitive. Further study into how to better predict variances of streamflow may improve the performance of other methods.

Comparing Real-World Applications of Estimating Streamflow in Ungauged Basins

When regional regression is fused with flow transfer techniques the performance of those methods is considerably degraded. Uncertainty from regional regressions is carried through the flow transfer techniques and results in a less accurate time series of estimated flows. However, one may wonder if these estimation techniques, though inferior to the idealized results, outperform simpler traditional techniques. In this section, the two best methods, SM1R and

MOVE12L will be compared against the drainage area ratio. The results document that the additional uncertainty introduced by the regional regression techniques may not be worth the effort.

The range of overall Nash-Sutcliffe Efficiency for the entire United States and each meta-region shows that these three methods are very similar. As can be seen in **Figure 5.17**, SM1R and DAR perform almost identically, with MOVE12L falling just below those, when considering the entire US. In the East, SM1R shows a slightly smaller range of NSE than DAR. This is true across all meta-regions, though the medians remain fairly similar. From this single figure, it is hard to draw any definite conclusion. SM1R seems to offer some advantage, though the advantage is slight.

Considering the relative efficiencies of SM1R and MOVE12L to DAR, it is clear that there is some advantage to SM1R. SM1R showed a relative efficiency of 1.35 for the entire US in terms of NSE. This advantage was lowest in the Midwest (1.16) and greatest in the West (1.66). On the other hand, MOVE12L showed little added-value with a relative efficiency of only 1.05. This suggests that SM1R is the stronger of the two methods. Although relative efficiencies are useful for understanding the general picture, one should focus more on the site-by-site comparison. SM1R exhibited a greater NSE than DAR at 48% of sites in the US, compared to only 36% for MOVE12L. For SM1R this percentage reached a high of 51% in the East. MOVE12L maximized at 43% in the West. Clearly, SM1R holds a strong advantage over MOVE12L, but neither technique provides significant value beyond DAR.

When considering bias, as in **Figure 5.18**, the deficiencies of both regionally-characterized methods become immediately apparent. In the US on the whole and in the Midwest and West, both methods show a slight positive bias. MOVE12L only shows a positive bias in the East. However, the DAR method shows symmetrical bias across all meta-regions. This is an important indicator that regional regression methods may not be a substantial improvement over the DAR method.

Not surprisingly, neither SM1R nor MOVE12L shows a significant advantage in terms of the relative efficiency for bias. For the entire US those efficiencies were 0.96 and 0.72, respectively. Interestingly, both regional regression methods showed an advantage in the East, with relative efficiencies of 1.91 and 1.03. Again, these relative efficiencies only report an average advantage. Site-by-site SM1R was less biased than DAR at 48% of sites, while MOVE12L was less biased at only 43% of sites in the US. Outside of the East, these percentages drop to an average of 39% for each method. Again, these results do not reflect a vast, systematic improvement over DAR. In fact, neither method is able to outperform DAR.

The meta-regional percentage of sites where each method outperforms DAR is shown in **Figure 5.19**. From this figure, it is clear the SM1R holds an advantage over MOVE12L in relation to DAR, but neither method significantly improves on DAR. In the East and West, SM1R achieves a better NSE than DAR at just barely 50% of sites nationally. Considering overall performance metrics, neither regionally-characterized method is worth the added effort.

As for the monthly performance, the range of national, monthly NSE for each method is shown in **Figure 5.20**. It is promising that SM1R is almost indistinguishable from DAR. MOVE12L is clearly inferior to DAR in most months. Across all months, SM1R showed an average relative efficiency of 1.29 with no months having a relative efficiency below one. Site-by-site, SM1R had a greater NSE at 49% of sites on average, a fraction that dropped below 50% in seven months. MOVE12L had an average relative efficiency of only 0.79 with a greater NSE at only 41%.

When one considers the bias of these methods in **Figure 5.21**, a concerning trend arises in the bias of the regionally parameterized methods. In all months, both demonstrate a clearly positive median bias, a trend that increases westward. DAR maintains symmetric bias, but there is a seasonal trend to the range. Interestingly, SM1R maintained an average monthly relative efficiency of 1.09 while MOVE12L demonstrated an average of 0.81. In four months, the relative efficiency of SM1R dropped below unity. Site-by-site, SM1R was less biased than DAR at 47% of sites in the US, while MOVE12L was less biased at only 44% of sites.

Figure 5.22 shows the average monthly percentage of sites where each method outperforms DAR in each meta-region. This presentation reaches the same conclusion as the overall results: namely that neither method results in a significant improvement over DAR. The relative performance is worst in the Midwest, where both methods displayed a percentage of sites uniformly below

50% for all months. MOVE12L was always under 50%, regardless of month or meta-region. SM1R performed well in the East, but only marginally.

When all analyses are taken into account, neither SM1R nor MOVE12R, coupled with regional regression, provides any sizeable advantage over DAR. In both cases, one would estimate flows more accurately by using DAR rather than using regional regression with a more complicated flow-transfer technique. It may be that another technique for record characterization could improve the performance of the regionally-characterized flow transfer methods. For now it is sufficient to recognize that these methods provide no advantage over the simple and traditional DAR technique. Compared to each other, it is SM1R that is more promising, but neither is more attractive than the parsimonious drainage area ratio.

VI. WEIGHTED FLOW TRANSFER TECHNIQUES

The previous chapter showed that the combination of regional regression and traditional flow-transfer techniques did not provide a significant advantage over the drainage-area ratio (DAR) method. The use of regional regression for flow characterization at the ungauged site combined with flow-transfer methods did not generally outperform DAR because of the uncertainty introduced by the regional regression methods. Still, the methods using regional record characterization performed best at some sites, while DAR performed best at others. Additionally, the idealized experiments showed that the performance of some flow transfer methods could be linked to hydroclimatic conditions in the region and at the site of interest. Combining these two discoveries, it may be possible to predict which method will perform best at a given site based on the hydroclimatic conditions of that site.

If one can estimate the relative performance of DAR and a method using regional characterization, it would be possible to develop a weighted estimate that combines the benefits of both methods. Such a weighted estimator combines the relative stability of the drainage-area approach, which relies on no record characterization, with the added accuracy of regionally-characterized methods like standardizing by mean (SM) or the maintenance of variance extension (MOVE). This can be achieved by considering the weighted average of the two competing techniques such that the estimated streamflow, Q, is

$$\hat{Q} = \omega\hat{Q}_{SM} + (1-\omega)\hat{Q}_{DAR} \quad (6.1)$$

where the subscripts dictate the method used to estimate each flow and ω denotes a weight between zero and one that is based on the relative performance of the two methods and the hydroclimatic conditions at the site of interest. Equation (6.1) can be replicated for a weighted averaging of MOVE by replacing $\hat{Q}_{SM}$ with $\hat{Q}_{MOVE}$. This technique can provide a favorable weighting of the two techniques that maximizes the advantages of each method while minimizing the disadvantages of each technique.

This chapter will explore the development, application and performance of such a weighting technique. First, the development of and estimation of the weight, ω, will be presented. Then, this approach will be applied in both an idealized and real-world context to both the SM and MOVE methods of estimating flow. Finally, it will be shown that a weighted averaging of SM and MOVE provides the best technique for estimating monthly time series at ungauged sites, among all the methods tested.

Developing a Weighted Estimator for Streamflow Estimation

The weight, ω, presented in equation (6.1) is based on the relative efficiency of the two competing methods. Thus the weight can only be known explicitly in an idealized experiment analogous to those presented in Chapter Three. In a real-world application, it would be necessary to estimate that relationship. Here, all notations will be presented as if combining DAR and SM; the process is the same with MOVE.

The basis of the weight is the relative efficiency of DAR and SM methods. Earlier, the overall relative efficiency of the Nash-Sutcliffe efficiencies (NSEs) was used as a performance metric. As can be seen in Chapter Two, this relied on an average across all sites in the numerator and denominator. As the concern is now only with a single site, the average is no longer relevant and the relative efficiency, *r*, is given as

$$r = \frac{(NSE_{DA}-1)^2}{(NSE_{SM}-1)^2} \qquad (6.2).$$

The relative efficiency, *r*, in (6.2) will be greater than one if SM provides a greater NSE than DA.

The relative efficiency *r* can be converted into a weight bounded by 0 and 1 by considering the form of Langmuir equation. This approach is similar to that used by McGarity (2008) to develop a different weighting scheme in optimizing for best-management practices. With this formulation, the optimizing weight is

$$\omega = \frac{r}{1+r} \qquad (6.3).$$

Small values of ω indicate that DAR performed better than SM at the given site. Accordingly, the streamflow estimates in (6.1) would favor the estimated from DAR. As the relative efficiency approaches unity, indicating that both methods perform similarly, this weight approaches 0.5, weighting both techniques equally.

The commonly-used logistic link function can be applied to estimate the weight, *w,* as a function of the distance between the ungauged and index sites (*d*) and the drainage area, average annual aridity index, mean precipitation, potential evapotranspiration and temperature of the ungauged and index site. This function ensures that the weight, *w*, can only vary between zero and one. Here, the logistic link function is given by

$$w = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)}} \qquad (6.4),$$

where $X_1, X_2, \dots, X_n$ are the prediction variables listed above. This equation can be transformed into a linear regression through the transformation

$$\ln\left(\frac{w}{1-w}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{6.5}$$

With some algebraic manipulation, it can be shown that

$$ln\left(\frac{w}{1-w}\right) = \ln{(r)} \tag{6.6}$$

Thus, tri-step regression (see Chapter Four) can be used with the calibration data to develop an equation for estimating the weights within a certain region. Initial results showed that the most-promising functional form of (6.5) was a linear sum of logarithms rather than real-space predictors.

In the next couple of sections, this weighted method is applied to the SM and MOVE techniques. In each case, the results with a known weight from the leave-one-out application of the regionally-characterized methods will be contrasted with a leave-one-out estimated weight. In the case that no definite equation can be calibrated to estimate a weight, an arithmetic mean will be used instead, weighting each estimate equally and setting the weight equal to 0.5. These weights can then be used in (6.1) to estimate flows at an ungauged site; this approach will be abbreviated as Weighted Averaging (WAve).

A Weighted Combination of Drainage-Area Ratios and Standardization by Mean

The first implementation of WAve couples the SM and DA methods. Recall that the real-space annual (SM1R) and monthly (SM12R) methods were the most competitive, though neither provided much advantage over the drainage area ratio in real space (DAR) when combined with regional regression. The better of the two techniques, SM1R, exhibited a greater overall Nash-Sutcliffe efficiency (NSE) at 48% of sites in the US, being less biased at only 48% of sites. Weighting the performance of either SM method in combination with DA should improve overall performance relative to the strict use of the DA method.

Considering the estimates from DAR and SM with a known, idealized weight based on the calculated relative efficiency of the two methods will demonstrate that this method will indeed provide a significant added value over its component parts. That is, rather than estimating the weight, one can calculate it directly as if it were known a priori. The ranges of overall NSE and bias for DAR, SM1R, SM12R, WAve(SM1R) and WAve(SM12R) with idealized weights are presented in **Figure 6.1**.

As can be seen by comparing the boxplots, the weighted estimate based on DAR and SM1R demonstrates significant advantages over both DAR and SM1R, alone. For NSE, WAve(SM1R) demonstrated a relative efficiency to DAR of 1.90 and 1.41 to SM1R, overall. Site-by-site, WAve(SM1R) exhibited a greater NSE than DAR and SM1R at 63% and 67% if all sites in the US. In terms of bias, the respective relative efficiencies were even more dramatic – 2.51 and 2.63 – with WAve(SM1R) producing less overall bias at 63% and 69% of sites in the US. In this idealized context, WAve is a much better flow transfer technique than the standalone version of SM1R.

The story is similar for the weighted average of the real-space, monthly implementation of SM, WAve(SM12R). Again, the boxplots show that WAve(SM12R) has greater NSEs and a smaller range of bias than both DAR and

SM12R. In terms of NSE and bias, the relative efficiency of WAve(SM12R) was greater than 2.00 against DAR and SM12R. WAve(SM12R) had a greater NSE than than DAR at 65% of sites in the US and was less-biased at 63%. Against SM12R, WAve(SM12R) was less biased at 64% of sites and held a greater NSE at 81% of sites.

These results demonstrate the advantage of WAve over DA. If the optimal weights are known explicitly, the weighted-average approach intelligently favors the strongest competitor. Therefore, WAve significantly outperforms the traditional drainage-area technique as well as a regionally-characterized flow transfer technique. Below, the performance of WAve with a weight estimated from regional hydroclimatology is evaluated. The WAve(SM1R) and WAve(SM12R) are then contrasted so as to select the most advantageous method.

The overall performance of WAve(SM) with estimated weights is presented in **Figure 6.2**. WAve(SM1R), with an estimated weight, had a greater overall NSE than DAR at 58% of sites in the US and was less biased at 59% of sites. Versus SM1R, WAve(SM1R) performed better at 64% and 67% of sites in the US. As is shown in **Figure 6.3**, these percentages are above 50% in all meta-regions. The fact that WAve(SM1R) has a greater relative advantage against DAR in the East and SM1R in the Midwest and West shows the effectiveness of the weighted technique. Here WAve(SM1R) vastly improves over DAR in the East while significantly improving over SM1R in the other regions. The relative efficiencies of WAve(SM1R) against DAR and SM1R were all greater than 1.10

for NSE and bias in all meta-regions except for one. In the East, the relative efficiency against SM1R was only 0.98 for NSE.

With an average monthly relative efficiency of 1.44, WAve(SM1R) has a greater NSE than DAR at 57% of sites in the United States. In all months, the relative efficiency was greater than one and WAve(SM1R) outperformed at more than 50% of sites. The same is true in terms of bias, where the relative efficiency was 1.43 and WAve(SM1R) was less biased at 55% of sites on average. This is strong evidence that WAve(SM1R) is a significantly better flow-estimating technique. Versus SM1R, WAve(SM1R) resulted in a greater NSE at 59% of sites on average with an average monthly relative efficiency of 1.13. While the fraction of sites in the US where the weighted technique had a greater NSE than SM1R never dropped below 50%, the relative efficiency was below one in three months. In terms of bias, WAve(SM1R) outperformed SM1R at 61% of sites on average, with an average monthly relative efficiency of 1.35.

The average monthly percentage of sites where WAve(SM1R) outperforms DAR and SM1R can be seen by meta-region in **Figure 6.4**. From this figure it is clear that WAve(SM1R) holds a slight advantage over both of its component parts in the three meta-regions. Still, against DAR, this percentage dropped below 50% in one month in the Midwest for NSE and bias. Against SM1R, the East was the most competitive meta-region: the percentage of sites fell below 50% in three months for NSE and four for bias. Indeed, the relative efficiency of NSE against SM1R was less than one in five months in the East and the relative efficiency of bias was less than one in six months. This is because

SM1R was relatively strong technique in the Eastern region. Therefore, the added value is small in the East, though it is significant overall.

WAve(SM12R) was also able to significantly outperform its component parts. Overall this technique had a greater NSE than DAR at 58% of sites, with a relative efficiency of 1.58. In terms of bias the percentage was 60%, with a relative efficiency of 1.40. Against SM12R, WAve(SM12R) had a greater NSE at 74% of sites and was less biased at 61% of sites. The performance of WAve(SM12R) varied widely by meta-region, as can be seen in **Figure 6.5**. In the left panel, the fraction of sites where WAve(SM12R) outperforms DAR is just over 51% in the Midwest. This is because of coupling of SM12R's poor performance and DAR's relatively stable performance. As a result, weight was able to vastly improve upon SM12R, but the gap between the two components was too large to significantly surpass DAR as well. In the right panel, the relative performance against bias is fairly stable.

On average, WAve(SM12R) had a greater monthly NSE than DAR and SM12R at 55% and 58% of sites nationally. Nationally, this percentage was never less than 50% against DAR or SM12R. In terms of bias, WAve(SM12R) was less biased at 64% and 63% of sites, a percentage that was again uniformly greater than 50%. **Figure 6.6** shows the average monthly site-by-site performance in the three meta-regions. The pattern is similar to that seen overall. The relative performance of WAve(SM12R) and DAR is most degraded in the Midwest. This indicates that WAve(SM12R) behaves very similar to DAR in the

Midwest but still holds a slight advantage. Taken as a whole, WAve(SM12R) does appear to succeed beyond its component parts.

Both WAve methods presented above significantly outperform their component parts, but, looking at the right-hand side of each panel in Figure 6.2, both perform quite similarly overall. In the idealized sense, WAve(SM12R) was the better, but, when weights are estimated, it is actually WAve(SM1R) that appears superior. Overall, WAve(SM1R) had a greater NSE at 52% of sites and was less biased at the same percentage. While these percentages favor WAve(SM1R), the relative efficiency is about 1.03 for NSE, indicating that both methods are about on par, while it is merely 0.82 for bias. This shows that WAve(SM1R) while performing slightly better is somewhat plagued by bias.

Considering monthly performance, the range for both WAve(SMR) methods is shown in **Figure 6.7**. The first stark change from the SMR methods alone is the dampening of the positive bias seen in SM1R. Here, the influence of DAR causes the monthly bias to be slightly more symmetrical about zero. The seasonal variation of range in bias remains with WAve(SM1R) and is introduced by DAR into WAve(SM12R). The trend in NSE is similar to the trend seen in the idealized implementation of flow transfer methods in Chapter Three. Site-by-site, WAve(SM1R) had a greater monthly NSE at 51% of sites and was less biased at only 48% of sites on average. The percentage was below 50% nationally in three months for NSE and ten for bias. While this number is large, the percentage of outperformers indicates how similar the methods truly are.

These percentages were fairly consistent across all meta-regions, though **Figure 6.8** shows the introduction of positive bias for WAve(SM1R) in the West. Still, WAve(SM12R) shows a season trend in the median bias, but is always less than the median of WAve(SM1R). It seems that if one is concerned with bias over NSE, then one should be cautious in the application of WAve(SM1R). The hazard of underestimating flows could easily be imagined by considering the development of flood-frequency maps with erroneous data.

This analysis reinforces the similarity of the performance of both methods. Both Wave(SMR) weighted estimators are improvements over the individual estimators upon which they are based, but it is WAve(SM12R) that offers the strongest competitive advantage overall. WAve(SM12R) avoids the systematic positive bias that appears with Wave(SM1R) in the West. Still, both methods are on par with each other. Site-by-site the advantage against each other is only 50-50; thus both methods are comparable.

A Weighted Combination of Drainage-Area Ratios and Maintenance of Variance Extension

The same technique that was used to develop a weighted estimator based on DAR and SMR can be used to merge the performance of DAR and monthly standardizations with the maintenance of variance extension. Recall that the two most attractive regionally-characterized variants of MOVE were the real-space and log-space, monthly variants MOVE12R and MOVE12L. On a site-by-site basis, MOVE12R outperformed MOVE12L, but, compared to DAR, MOVE12L had a greater NSE at 36% of sites, with MOVE12R was greater at only 33% of sites. In either case, regional-parameterization degraded the performance of MOVE such that DAR was the most competitive method. Weighting each approach should intelligently select the better performing method; this can be judged in the same manner as the combination of DAR and SMR.

The idealized performance of the WAve(MOVE) can be assessed by considering the combination of DAR and MOVE with the optimal weight calculated directly from the component Nash-Sutcliffe efficiencies (NSEs) as if they were known a priori. The range of performance of the components and the weighted techniques are presented in **Figure 6.9**. Clearly the weighted techniques provide an improved range of NSE, but the affect on bias is most remarkable. The range and magnitude of bias is greatly reduced in the weighted techniques.

With an idealized weight, WAve(MOVE12R) had a greater overall NSE than DAR and MOVE12R at 66% and 85% of sites, respectively. For bias, WAve(MOVE12R) outperformed at 63% and 68% of sites. This is a marked improvement that clearly demonstrates the added-value of intelligently weighting the methods. WAve(MOVE12L) had a greater overall NSE at 68% and 83% of sites and was generally less biased at 62% and 66%. From these numbers and the boxplot, one can conclude that WAve(MOVE12L) is the better weighted technique if the weight is known explicitly. In the real world, this relative performance will depend on largely on one's ability to accurately estimate the weight.

Figure 6.10 displays the overall performance of WAve(MOVE) when the weights are estimated from regional hydroclimatic data. WAve(MOVE) continues to display an added advantage over the components, but the advantage is not as striking as that seen in the idealized case. This is to be expected due to the uncertainty introduced by estimating the optimizing weight. Below, WAve(MOVE12R) will be rigorously compared with its component parts. The WAve(MOVE12L) will be considered analogously. Finally, WAve(MOVE12R) and WAve(MOVE12L) will be contrasted directly.

With estimated weights, WAve(MOVE12R) continued to showed a marked improvement over DAR and MOVE12R with regional regression. WAve(SM12R) had a greater overall NSE than DAR and MOVE12R at 56% and 78% of sites respectively. Similarly, it was less biased at 60% and 65% of sites nationally. In terms of relative efficiency, WAve(MOVE12R) always exhibited an efficiency greater than 1.28, indicating significant added value. The percentage of sites where WAve(MOVE12R) outperformed varied only slightly across the three meta-regions, as can be seen in **Figure 6.11**. The relative performance of WAve(MOVE12R) against DAR dipped dramatically in the Midwest, but remained above 50%. As with WAve(SMR), this is due to the extreme difference between regionally-parameterized methods and DAR in the Midwest.

The average monthly relative performance of WAve(MOVE12R) is presented in **Figure 6.12**. These results are quite similar to the overall relative performance. On average, WAve(MOVE12R) had a greater monthly NSE than DAR at 55% of sites nationally and was less biased at 58%. Against MOVE12R, WAve improved at 68% and 63% of sites. In no months were any of these national percentages below 50%. On average, this performance was consistently above 50% in all meta-regions. As would be expected from the initial results, the relative performance of WAve(MOVE12R) against DAR in terms of NSE was slightly degraded in the Midwest. There, the percentage of outperforming sites was below 50% in five months. Though remarkable, closer examination revealed that the deficient months remained competitive at about 48%.

This comparison shows that WAve(MOVE12R) behaves significantly better than its component estimators, even when the weight must be estimated from hydroclimatic variables. Some weakness remains in the Midwest, but WAve(MOVE12R) maintains a slight advantage there. One will note that WAve(MOVE12R) exhibits a slightly positive median bias overall (recall Figure 6.10), but both the median and range is smaller than that associated with the regionally-parameterized edition of MOVE12R.

WAve(MOVE12L) also demonstrated a significant advantage over DAR and MOVE12L. WAve(MOVE12L) had a greater overall NSE than DAR at 60% of sites nationally and was less biased at 58%. Compared to MOVE12L, those percentages were 77% and 64%. This relative performance was significant across the three meta-regions, as is shown in **Figure 6.13**. It should be noted that the deficiency against DAR in the Midwest is less drastic here than as was the case with WAve(MOVE12R), though the general trends were quite similar.

When considered monthly, WAve(MOVE12L) had a greater NSE than DAR and MOVE12L at 56% and 67% of sites on average. The bias was smaller at 58% and 64% of sites nationally, on average. As **Figure 6.14** demonstrates, this relative performance was again robust across the three meta-regions. Of all the comparisons, the average percentage fell below 50% in only a single month and was only for the comparison of bias against DAR. This is strong evidence that WAve(MOVE12L) corrects for the monthly deficiency seen in the comparison between DAR and WAve(MOVE12R). The percentage of sites where WAve(MOVE12L) showed a smaller NSE never dropped below 50%, but for WAve(MOVE12R) it was below 50% in five months in the Midwest.

Both WAve(MOVE) methods are strongly competitive, but one must wonder which is the strongest method. Tete-a-tete, WAve(MOVE12L) had a greater overall NSE at 52% of sites, while it was less biased at only 49% of sites. Indeed, it was only in the West that WAve(MOVE12R) had a greater percentage of higher NSEs and lower bias. These comparisons are so close to 50% that it is clear that both methods are extremely competitive. Again, it may be a coin toss. **Figure 6.15** summarizes the monthly performance of each WAve(MOVE) technique. On average, WAve(MOVE12L) had a greater monthly NSE at 51% of sites while being less biased at only 49%. In all cases the statistics were extremely similar. It should be noted, though, that both methods exhibited a tendency towards underestimation in the West, as can be seen in the comparison of monthly bias in **Figure 6.16**.

Though WAve(MOVE12R) and WAve(MOVE12L) are both strong flow transfer techniques, WAve(MOVE12L) seems to have the greatest advantage over the traditional DAR technique. WAve(MOVE12L) had a greater overall NSE at 60% of sites compared to only 59% for WAve(MOVE12R). Yes, the difference is small, but it is consistent across all meta-regions and months. As with WAve(SM) one should seriously consider the application and caveats before choosing one method over another. As this analysis is concerned with the strongest overall method, WAve(MOVE12L) must be advocated for any general usage.

Comparing Two Weighted Combination Methods

Having considered a number of variants for weighting results, it is important to now consider which one is superior to the other. WAve(SM12R), which will be referred to as WAve1 here, offered significant improvements over DAR but was hampered by a slight seasonal trend of median bias in the West. WAve(MOVE12L), or WAve2, was also very strong, if not stronger, but was limited by even larger bias in the West. As it is important to understand the relative performance of these methods, they will be compared with each other and with the standard drainage area ratio technique here.

First, it is interesting to take note of the strongest WAve techniques relative to the idealized flow-transfer techniques of Chapter Three. In the idealized experiments, it was shown that SM12R was the strongest flow-transfer technique. When regional regression was incorporated, SM1R took the advantage, but WAve returned SM12R to prominence. This gives strong evidence to the ability of intelligently-weight averaging to correct for some level of uncertainty introduced by the regional parameterization. While MOVE12L was consistently the best technique, the story is similar in terms of relative performance: WAve(MOVE12L) corrects for some uncertainty, bringing the results nearer to the idealized case than with the regionally-parameterized MOVE12L.

The range of overall NSE for DAR, Wave1 and WAve2 are shown nationally and by meta-region in **Figure 6.17**. WAve1 and WAve2 are almost identical nationally. Furthermore, both weighted techniques are superior to DAR. In the East, WAve2 holds a slight edge, but WAve1 has the advantage in other regions. **Figure 6.18** shows the range of bias for all of the methods. Here, both methods are even less distinguishable. Overall, WAve1 seems to have a smaller range. In the East, WAve2 shows a slightly smaller range. All of the methods are nearly identical in the Midwest, but both weighted techniques show a slightly positive median bias in the West.

Site-by-site, WAve1 had greater NSE than DAR at 58% of sites and was less biased at 60%. WAve2, on the other hand, had a greater NSE than DAR at 60% of sites and was less biased at 58% of sites nationally. Clearly this is a very tight comparison. **Figure 6.19** shows that WAve1 holds an advantage in the West, but WAve2 has the edge in terms of NSE in the East and Midwest. Site-by-site, WAve2 has a greater NSE than WAve1 at 52% of sites, a percentage that is above 50% in all meta-regions except the West. WAve1 holds the advantage in terms of bias in all meta-regions except the East.

When considered monthly, WAve1 has a greater monthly NSE than DAR at 55% of sites on average, while WAve2 outperforms DAR at 56% of sites on average. In terms of bias, both outperform DAR at 58% on average. Meta-regionally, both WAve1 and WAve2 similarly outperform DAR, as is shown in

Figure 6.20. It is nearly impossible to select which method is the better weighted technique. The two methods are equally valuable.

The monthly range of NSE is shown nationally for all three methods in **Figure 6.21**; **Figure 6.22** shows the range of bias. Again, both methods are nearly identical, but consider the bias associated with WAve2 in Figure 6.18. It may be that WAve1 offers slightly less bias. Furthermore, **Figure 6.23** shows the range of monthly bias in the West for each method. The positive median bias is apparent in all methods, but is most dramatic in WAve2. In the future, a bias correction may be able to correct for this, but for now it is enough to slightly favor WAve1 over WAve2.

The weighted averaging of the real-space drainage area ratio (DAR) and the real-space, monthly standardization (SM12R) with regional regression is a robust method for estimating monthly time series of streamflows at ungauged sites in the United States. This analysis has shown that this technique has a competitive advantage over the simple, traditional approach of using a drainage area ratio. As with all methods, there are a number of caveats associated with this technique. Without summarizing the entire study, it is sufficient to say that one should always look at the regional and monthly performance of a method in the region of interest before advocating any single technique.

VII. An Extension: Hybrid Flow Transfer Techniques

Until this point, this exploration has only considered traditional techniques for estimating monthly streamflow time-series and linear combinations of those methods. It was shown that, though regionally-characterized, traditional methods were unable to significantly outperform the simplest flow-transfer tool, the real-space drainage area ratio (DAR), a weighted combination of those methods provided significant added value. Of course, linear combinations are only a first order exploration of combining the benefits of traditional methods. In this chapter, a brief extension of this work will be considered for further research.

As none of the traditional methods, when combined with regional parameterization, provided a significant advantage over the DAR method, it may be that these methods do not capture the true functional form of the relationship between streamflow time-series in hydrologically similar basins. If the functional form was known explicitly, an idealized application of that form would achieve extremely high Nash-Sutcliffe values across all sites; as in an earlier chapter, where it was seen that the idealized standardization by mean and the standardization with maintenance of variance were both closer to the functional form than DAR.

The exploration of functional forms is worth an entire report of its own and is not the true focus of this report, but some preliminary results are presented here. A number of combinations of DA and SM were tested in an idealized case and only the combination of the logarithmic drainage area ratio (DAL) and the real-space standardization by mean (SMR). Similarly, the combination of DAL and the lognormal standardization with the maintenance of variance extension (MOVEL) was the only combination of DA and MOVE that was found to be viable. In the following sections the potential of these methods will be presented and some additional functional forms will be theorized. Though this analysis is incomplete, it provides an interesting point from which to move forward.

FUNCTIONAL-FORM COMBINATIONS OF DRAINAGE AREA RATIOS AND STANDARDIZATION BY MEAN

One can easily imagine a wide range of functional forms that combine the benefits of drainage area ratios and standardization by mean. The first part of the Appendix shows a number of equations for the fusion of DA and SM. Idealized, initial results showed that, of these methods, only the combination of DAL and SMR was promising in any way.

The combination of DAL and SMR theorizes that, between two hydrologically similar basins, the relationship of streamflows can be described as

$$\frac{\ln\left(\frac{Q_X}{\mu_X}\right)}{\ln\left(A_X\right)} = \frac{\ln\left(\frac{Q_Y}{\mu_Y}\right)}{\ln\left(A_Y\right)} \tag{7.1}$$

where Q is flow and A is drainage area at the subscripted sites X and Y. Clearly, this method requires some parameterization of streamflows at the ungauged site. Here, this was accomplished with regional regression, as outlined above. The functional form in (7.1) leads to two variants of DALSMR: an annual and a monthly variant.

The ranges of overall Nash-Sutcliffe efficiencies (NSEs) for both the annual (DALSM1R) and monthly variants (DALSM12R) are presented with DAR in **Figure 7.1**. Nationally, both methods are very similar, yet both appear inferior to DAR. In the Midwest, DALSM12L distinguishes itself above the others, but the difference is only marginal. A consideration of bias in presented in **Figure 7.2**. As would be expected, the annual technique displays some non-zero bias nationally. This bias increases westward. DALSM12R, on the other hand, displays non-zero median bias only in the West. In all regions, neither range outperforms DAR.

Compared to DAR, neither the DALSM1R nor DALSM12R offered a significant site-by-site improvement. With a relative efficiency of 0.82, DALSM1R exhibited a greater overall NSE at only37% of sites. Still worse, DALSM12R had a relative efficiency of 0.85, outperforming DAR at only 31% of sites. In terms of bias, both methods outperformed DAR at about 44% of sites, though DALSM1R had a much great relative efficiency (0.85) compared to DALSM12R (0.69). This poor relative performance replicates itself meta-regionally, as is shown in **Figure 7.3**. Only in the East does one method seem to offer some marginal advantage over DAR. Elsewise, the functional-form combinations are not an improvement on DAR.

The monthly performance of each method is briefly summarized in terms of NSE in **Figure 7.4** and bias in **Figure 7.5**. For the NSE, neither method displays an alarming trend dissimilar form DAR, but DALSM12R appears slightly inferior in most months. For bias, it is DALSM1R that is most worrisome. The annual method continues to display a distinct positive bias, though it is not huge. The monthly method is more symmetrical about zero, but the range is still greater than that seen with DAR.

On average, DALSM1R had a greater monthly NSE than DAR at 41% of sites, while DALSM12R outperformed at only 38% of sites. In all the months, this percentage was never greater than 50%. At 44% outperformance for each method, the story for bias was similar except in the East and West. These results can be seen in **Figure 7.6**. In the East, DALSM1R was less biased at nearly 50% of sites on average. In the West, it was DALSM12R that thrived, with less monthly bias at half of the sites on average. In both case, the percentage of sites with less monthly bias exceeded 50% for about half of the year.

Compared to each other, DALSM1R actually provided greater overall NSEs than DALSM12R at 52% of sites. It was less biased at 49%. These results show that the two methods are extremely similar, but, on the basis of NSE, DALSM1R holds a slight advantage. Still, DALSM1R is plagued by some positive bias. Furthermore, neither was a significant improvement over DAR. This suggests that this functional-form combination of DA and SM is not a promising combination. Further work along this path may lead to a more promising result.

Functional-Form Combinations of Drainage Area Ratios and Maintenance of Variance Extension

Another functional-form combination to consider is the fusion of drainage area ratios and the maintenance of variance extension standardization. Three potential combinations of these methods are presented in the second part of the

Appendix. Of those approaches, only one, the combination of DAL and MOVEL was found to have any potential.

The combination of DAL and MOVEL can be described in two parts. First consider the standardization with MOVEL, *S*,

$$S_{\ln(Q_X)} = \frac{\ln(Q_X) - \mu_{\ln(Q_X)}}{\sigma_{\ln(Q_X)}} \tag{7.2}$$

where *Q* is flow at the subscripted site. The combination of DAL and MOVEL can be summarized as,

$$\frac{\ln(S_{\ln(Q_X)})}{\ln(A_X)} = \frac{\ln(S_{\ln(Q_Y)})}{\ln(A_Y)} \tag{7.3}$$

where *A* is the drainage area of the subscripted site. Again, this method requires regional parameterization, which means that one can also consider both the annual and monthly variants. These will both be explored below.

From the overall range of NSEs for both methods presented in **Figure 7.7** it is clear that the annual technique, DALMOVE1L does not perform well relative to DAR. This is true across all meta-regions. DALMOVE12L, on the other hand, seems somewhat competitive, especially in the West. When examining bias in **Figure 7.8**, the deficiency of the monthly method becomes more apparent, though the annual method is also far from reassuring. The annual technique displays a clear, uniform positive bias. The monthly variant only displays positivity westward, but also exhibits a much greater range of bias.

Against DAR, DALMOVE1R exhibited a greater overall NSE at only 18% of sites, with a relative efficiency of 0.57. DALMOVE12L was slightly more competitive with a relative efficiency of 0.72 and an outperformance rate of 29%. Clearly there is little advantage for NSE in either of these methods. For bias the relative performance improves slightly: DALMOVE1L was less biased at 47% of sites nationally, and DALMOVE12L was less biased at 39% of sites. This same trend can be seen across the meta-regions in **Figure 7.9**. Again, this is strong evidence that neither method is incredibly useful.

Figures 7.10 and **7.11** show the monthly range of overall NSE and bias for each combination method. DALMOVE12L appears in line with DAR, though it is slightly inferior. The performance of DALMOVE1L is alarmingly poor. The trend of NSE is vastly different than the trend seen in the other methods and the ranges are far lower. For bias, DALMOVE1L exhibits a seasonal trend in the median and the range of bias.

On average, DALMOVE1L had a greater monthly NSE than DAR at 24% of sites, while DALMOVE12L outperformed at more than 35% of sites on average. Still, these percentages never exceeded 50% in any month. The story is nearly identical for monthly bias. Additionally, this pattern does not change from the East to Midwest and West, as is seen in **Figure 7.12**.

It almost goes without saying that DALMOVE12L is significantly better than DALMOVE1L. Still, the monthly method had a greater overall NSE than the annual method at over 75% of sites in the US. It was less biased at 45% of sites. Still, DALMOVE12L was not an improvement over DAR. As before, some additional exploration here may find a more valuable functional form.

Comparing Two Functional-Form Combination Methods

Taking the two functional-form combinations, DALSM1R and DALMOVE12L, it is clear that neither offers a significant improvement over

DAR. As such, neither is really worth much further consideration. Still, in the interest of seeding further research, it is useful to consider which technique is better than the other. For ease, these methods will be referred to as DASM and DAMOVE here.

Combining the range of overall NSE and bias for each method on several figures, as in **Figures 7.13** and **7.14**, it is clear the DASM appears to have a slight advantage over DAMOVE. The NSE of DAMOVE is more widely distributed than the others nationally and in the Midwest. It becomes slightly more competitive in the East and West. The range of bias for DAMOVE is always greater than the others. Furthermore, both methods exhibit a strong positive bias westward. In all cases, neither vastly improves upon DAR.

Recall that DAMOVE had a greater NSE than DAR at only 28% and DASM improved at 37% of sites. This alone suggests that DASM is the more promising method. For bias, DASM outperforms at 40%, while DAMOVE outperformed at only 35%. This relative comparison was replicated across each meta-region, as in **Figure 7.15**. In all meta-regions and statistics, DASM has an edge over DAMOVE, yet it is only for bias in the East that DASM is an improvement on DAR.

The monthly range of performance, in **Figures 7.16** and **7.17**, the distinction between the two methods is less apparent. For NSE, both are strong, though DASM is closer to the performance of DAR. But, in terms of bias, DASM is strongly positive while DAMOVE remains relatively symmetric about zero. Considering the average monthly percentage of sites where each method outperforms DAR, DASM continues to hold a slight edge, both nationally and in each meta-region (**Figure 7.18**).

After all, the need for further research is strikingly clear in the case of new functional forms. Of the two contrasted here, it is DASM that performs better than DAMOVE on the merit of NSEs. Still, this technique has yet to be developed to the point that it provides a significant advantage over traditional, more parsimonious efforts such as the use of DAR. From here, future research should delve into bias correction and the development of novel functional forms of the relationship between streamflows in hydrologically similar basins.

VIII. Conclusions and Recommendations

Through a myriad of different experiments, this thesis has demonstrated a number of important conclusions regarding the methods for estimation of time series of monthly streamflows at ungauged sites. Perhaps most importantly, this work introduces a unique methodology for assessing and evaluating techniques for hydrologic prediction in ungauged basins that can be extended to assess many other new ideas for transferring streamflows. This thesis only examined a small set of regional, hydrostatistical methods; one could easily imagine continuing the research by applying a similar methodology to evaluate a host of other promising transfer methods.

In summary, this thesis first considered four different classes of flow-transfer techniques: the drainage-area ratio (DA), standardization by means (SM), maintenance of variance extension (MOVE) and the use of flow duration curves (QPPQ). In an idealized sense, when streamflow moments at the ungauged site are assumed known a priori, it was shown that DA is not favored over the other methods. In such instances, all other methods demonstrated a significant site-by-site improvement over DA, but among the many methods evaluated, it was the monthly, log-space variant of MOVE that was the most successful. MOVE12L had a greater Nash-Sutcliffe efficiency than DAR at over 86% of sites in the United States. All methods were also significantly less biased at a large percentage of sites as well.

When regional regression was introduced to characterize ungauged streamflows by estimating moments and to simulate the real-world prediction in ungauged basins (PUB), the results of the hydrostatistical techniques were markedly degraded. In that case, neither SM nor MOVE offered a greater NSE at more than 50% of sites in the US. Furthermore, both methods exhibited a much greater amount of bias than the traditional DA method. This degraded performance was related to the uncertainty introduced through regional characterization. Alternative methods of record characterization may improve future comparisons.

It was noted that the performance of different methods was closely related to a wide range of hydroclimatic variables. It was hypothesized and demonstrated that a weighted estimator could be developed to blend the advantages of hydrostatistical and traditional methods. These weighted or WAve methods were able to outperform DA at about 60% of sites in the United States. This method was generally invariant to the region being considered, showing WAve to be a generally robust approach. Though WAve was unable to match the results of the idealized cases, it does show that blending of methods may be an attractive approach for estimation at ungauged sites,

Overall, the results of this thesis are promising, but this research is far from complete. While this thesis does advance the scientific understanding of PUB, it also provides a methodology for evaluating other techniques. There are many new techniques to be considered, as was hinted at in the functional-form extensions considered. By following the methods of this paper, one can reasonably evaluate any method. In addition, it is important to consider the drawbacks of current techniques.

RECOMMENDATIONS:

Current modeling techniques rely on two basic assumptions: the past is a reasonable guide to future and that models developed in gauged basins can be applied in ungauged basins (Sivapalan *et al*. 2003). Here, the second point is rather intractable, but, on the first point, this is analogous to observing that most methods rely on the stationarity of hydrologic statistics (Sivapalan *et al*. 2003). If the world is no longer stationary, this dependence becomes increasingly troubling (Milly *et al.* 2008). As hydrostatistical methods rely on the past so heavily, it remains to be seen if they can accurately describe the future. Hirsch (1982) showed that MOVE could be used for extending records, which is an idealized case of predicting the future. Further research should consider a more extensive testing of this very question.

One possible extension of this current work is an exploration into improvements associated with the traditional hydrostatistical techniques. For example, it may be possible to derive a bias correction factor that significantly improves the performance of the logarithmic DA or the log-space SM methods. Similarly, Emerson *et al*. (2005) and Asquith *et al.* (2006) showed that a more-advanced calibration of DA might prove quite valuable. Furthermore, it may be possible to improve Step Two, regional characterization of streamflow records, to improve these hydrostatistical methods. While this may prove useful, it gives rise to a whole range of improvements and new methods.

The logical next step would be to consider how hydrostatistical methods perform relative to other techniques, from generalized DA techniques to statistical regression and process-based models. In this pursuit one could followed the methods outlined here: consider the idealized case and then consider a leave-one-out rendition of the real world. This will advance the cause of PUB in that it will give a quantifiable demonstration of the relative performance of these methods.

Another area of research would be to consider the implication of using estimated streamflow data in real-world decision making processes. That is, how does a syndicated streamflow record affect the decision making process in the development of hydropower, irrigation scheduling or even flood mapping. Throughout the paper, results were quoted in terms of Nash-Sutcliffe efficiencies and Emerson *et al.* (2005) cite coefficients of determination, but what does an NSE of 0.70 really mean? This could be examined by using leave-one-out experiments: one could develop a flood frequency map with the observed flow record at an HCDN site and then repeat that exercise with an estimated record with a given NSE. What would be the difference between those maps? What are the policy implications?

The final and most-promising realm of further research involves the use of remote-sensing data. Lakshmi (2004) observed that remote sensing data could be used to develop hydroclimatic variables across a wide range of ungauged sites. While it has often been thought that remote sensing could be used to directly estimate hydrologic responses such as streamflow, this is merely in development and has not been extensively vetted. Alternatively, it may be possible to use remote sensing to more accurately develop information about other hydroclimatic variables.

In one case, the lengthy remote sensing record of precipitation, vegetation or even temperature may be useful tool for understanding the wetness of an ungauged site. One could then develop a sort of “wetness indicator” for the ungauged basin. Then, similar to the work of Smakhtin and Masse (2000), one could map the wetness indicator back to the streamflow values. This technique would limit the need for an index gage, which could reduce uncertainty.

Another use of remote sensing would be to estimate long-term streamflow statistics. While it may prove difficult to estimate the time series of streamflows at an ungauged site via remote sensing, one may be able to more accurately understand the moments of that streamflow through long-term remote sensing. If those moments could be known with extreme accuracy, then the case of streamflow transfer devolves into the idealized case presented above, which had quite promising results.

Furthermore, the use of remote sensing may be able to give some insight into the contribution of groundwater flows to surface water flows. For this project, little attention was given to groundwater. In essence, the groundwater basin was assumed to be identical to the topographic drainage basin. This may not always be the case. It may be that the hydrostatistical methods presented above perform poorly in basins dominated by groundwater flows. Remotely sensed data may a better understanding of the range of groundwater flow around an ungauged site so that one could adjust climatic inputs to more accurately estimate long-term flow parameters and behavior.

After all has been postulated and examined, the most important consideration is the time scale of the streamflow record desired. This project examined only monthly time series of streamflows. More and more studies are demanding daily data and daily records. The evaluation process presented here may arrive at a different conclusion when considering daily flows.

This project has shown one hydrostatistical technique for estimating streamflow more reliably than with the drainage area ratio. While it is a significant contribution to a project like IAHS's PUB, it has revealed a number of pressing concerns and avenues of further research. As the risk of sounding cliché, this project has merely opened a floodgate of PUB. One can be sure that, as other methods are assessed in a similar fashion, attractive methods for predicting streamflow time series at ungauged sites, whether on a monthly or daily scale, will begin to emerge. With sound estimates of streamflow, development in ungauged watershed can proceed intelligently, addressing the needs of both the natural and human populations.

REFERENCES

Asquith, WH, MC Roussel and J Vrabel. 2006. Statewide analysis of the drainage-area ratio method for 34 streamflow percentile ranges in Texas. *US Geological Survey Scientific Investigative Report,* 2006-5286: 34 pp.

Archfield, SA, and RM Vogel. 2010. Map correlation method: Selection of a reference streamgauge to estimate ungauged streamflow at ungauged catchments. *Water Resources Research* 46: W10513.

Archfield, SA, RM Vogel, P Steeves, S Brandt, P Weiskel and S Garabedian. 2010. The Massachusetts sustainable-yield estimator: a decision-support tool to assess water availability at ungauged sites in Massachusetts. *US Geological Survey Scientific Investigative Report,* 2009-5227: 41 pp.

Daly, C, RP Neilson and DL Phillips. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33.2: 140-158.

Daly, C, G Taylor and W Gibson. 1997. The PRISM approach to mapping precipitation and temperature. *Proceedings of the America Meteorological Society's 10th Conference on Applied Climatology*, American Meteorological Society, Boston, MA.

Emerson, DG, AV Vecchia and AL Dahl. 2005. Evaluation of drainage-area ratio method used to estimate streamflow for the Red River of the North Basin, North Dakota and Minnesota. *US Geological Survey Scientific Investigative Report,* 2005-5017: 13 pp.

Fennessey, NM. 1994. A hydro-climatological model of daily streamflow for the northeast United States. Ph.D. dissertation, Department of Civil and Environmental Engineering, Tufts University, Medford, MA.

Hirsh, RM. 1979. An evaluation of some record reconstruction techniques. *Water Resources Research* 15.6: 1781-1790.

Hirsch, RM. 1982. A comparison of four streamflow record extension techniques. *Water Resources Research* 18.4: 1081-1088.

Hughes, DA, and V Smakhtin. 1996. Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. *Journal of Hydrologic Sciences* 41.6: 851-871.

Kroll, CN, JG Luz, TB Allen, and RM Vogel. 2004. Developing a watershed characteristics database to improve low streamflow prediction. *Journal of Hydrologic Engineering*, ASCE, 9.2: 116-125.

Lakshmi, V. 2004. The role of satellite remote sensing in the prediction of ungauged basins. *Hydrological Processes* 18: 1029-1034.

McGarity, AE. 2008. Optimization of Riparian Easements and BMP Technologies: A Multiobjective Tradeoff Model. AWRA 2008 Summer Specialty Conference, Virginia Beach, Virginia.

Milly, PCD, J Betancourt, M Falkenmark, RM Hirsch, ZW Kundzewicz, DP Lettenmaier and RJ Stouffer. 2008. Stationarity is dead: Whither water management? *Science* 319.5863: 573-574.

Mohamoud, YM. 2008. Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves. *Journal of Hydrologic Sciences* 53.4: 706-724.

Sankarasubramanian, A, and RM Vogel. 2002. Annual hydroclimatology of the United States. *Water Resources Research* 38.6: WR000619.

Sivapalan, M. 2003. Prediction in ungauged basins: a grand challenge for theoretical hydrology. *Hydrological Processes* 17: 3163-3170.

Sivapalan, M, K Takeuchi, SW Franks, VK Gupta, H Karambiri, V Lakshmi, X Liang, JJ McDonnell, EM Mendiondo, PE O'Connel, T Oki, JW Pomeroy, D Schertzer, S Uhlenbrook, E Zehe. 2003. IAHS decade on prediction in ungauged basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal* 48.6: 857-880.

Slack, JR, AM Lumb and JM Landwehr. 1993. Hydroclimatic data network (HCDN): A US Geological Survey streamflow data set for the United States for the study of climate variation, 1874-1988. *US Geological Survey Water Resources Investigations Report 93-4076*. < http://pubs.usgs.gov/wri/wri934076/>.

Smakhtin, VY, and B Masse. 2000. Continuous daily hydrograph simulation using duration curves of a precipitation index. *Hydrological Processes* 14: 1083-1100.

Stedinger, JR, and RM Vogel, E Foufoula-Geogiou. "Frequency Analysis of Extreme Events." *The Handbook of Hydrology*. Ed. DR Maidment. New York: McGraw-Hill, 1993. 18.1-18.66.

Vogel, RM, I Wilson and C Daly. 1999. Regional regression models of annual streamflow for the United States. *Journal of Irrigation and Drainage Engineering* 125.3: 148-157.

Wagener, T, M Sivapalan, J McDonnell, R Hooper, V Lakshmi, X Liang and P Kumar. 2004. Predictions in ungauged basins as a catalyst for multidisciplinary hydrology. *Eos Transactions AGU*, 85.44: 451, 457.

APPENDIX

A Collection of Functional Form Combinations:

Variables:

X	(An arbitrary index site)
Y	(An arbitrary ungauged site)
A_X	(Drainage area of the subscripted site)
Q_X	(Monthly streamflow of the subscripted site)
μ_{Q_X}	(Mean of the subscripted variable)
σ_{Q_X}	(Standard deviation of the subscripted variable)

PART ONE:

Combinations of Drainage Area Ratios and Standardizations by Mean

i. Combining real-space drainage area ratios and real-space standardizations by mean:

$$\frac{Q_X}{A_X \mu_{Q_X}} = \frac{Q_Y}{A_Y \mu_{Q_Y}} \tag{A.1}$$

ii. Combining real-space drainage area ratios and lognormal standardizations by mean:

$$\frac{\ln(Q_X)}{A_X \mu_{\ln(Q_X)}} = \frac{\ln(Q_Y)}{A_Y \mu_{\ln(Q_Y)}} \tag{A.2}$$

iii. Combining logarithmic drainage area ratios and real-space standardizations by mean:

$$\frac{\ln\left(\frac{Q_X}{\mu_{Q_X}}\right)}{\ln\left(A_X\right)} = \frac{\ln\left(\frac{Q_Y}{\mu_{Q_Y}}\right)}{\ln\left(A_Y\right)} \tag{A.3}$$

iv. Combining logarithmic drainage area ratios and lognormal standardizations by mean:

Two Variations:

$$\frac{\ln\left(Q_X\right)}{\ln\left(A_X\right)\mu_{\ln\left(Q_X\right)}} = \frac{\ln\left(Q_Y\right)}{\ln\left(A_Y\right)\mu_{\ln\left(Q_Y\right)}} \tag{A.4}$$

$$\frac{\ln\left(\frac{\ln\left(Q_X\right)}{\mu_{\ln\left(Q_X\right)}}\right)}{\ln\left(A_X\right)} = \frac{\ln\left(\frac{\ln\left(Q_Y\right)}{\mu_{\ln\left(Q_Y\right)}}\right)}{\ln\left(A_Y\right)} \tag{A.5}$$

PART TWO:

Combinations of Drainage Area Ratios and Maintenance of Variance Extension Standardization

Standardizations:

Real-space MOVE:

$$S_X = \frac{Q_X - \mu_{Q_X}}{\sigma_{Q_X}} \tag{A.6}$$

Lognormal MOVE:

$$S_{\ln\left(X\right)} = \frac{\ln\left(Q_X\right) - \mu_{\ln\left(Q_X\right)}}{\sigma_{\ln\left(Q_X\right)}} \tag{A.7}$$

i. Combining real-space drainage area ratios and real-space MOVE:

$$\frac{s_X}{A_X} = \frac{s_Y}{A_Y} \tag{A.8}$$

ii. Combining real-space drainage area ratios and lognormal MOVE:

$$\frac{s_{\ln(X)}}{A_X} = \frac{s_{\ln(Y)}}{A_Y} \tag{A.9}$$

iii. Combining logarithmic drainage area ratios and lognormal MOVE:

$$\frac{s_{\ln(X)}}{\ln(A_X)} = \frac{s_{\ln(Y)}}{\ln(A_Y)} \tag{A.10}$$

TABLES

CHAPTER THREE:

Table 3. 1. The relative efficiencies of four variants of MOVE in terms of NSE across all meta-regions.

	US				East			
	1R	1L	12R	12L	1R	1L	12R	12L
1R		0.9806	0.5469	0.5249		0.9854	0.7144	0.6772
1L	1.0198		0.5578	0.5353	1.0149		0.7250	0.6873
12R	1.8284	1.7929		0.9597	1.3997	1.3792		0.9479
12L	1.9053	1.8682	1.0420		1.4767	1.4551	1.0550	
1R		0.9920	0.5415	0.5250		0.9476	0.4008	0.3800
1L	1.0080		0.5459	0.5293	1.0553		0.4230	0.4010
12R	1.8466	1.8319		0.9695	2.4948	2.3641		0.9481
12L	1.9046	1.8895	1.0314		2.6314	2.4935	1.0548	
	1R	1L	12R	12L	1R	1L	12R	12L
	Midwest				West			

Table 3. 2. The relative efficiencies of four variants of MOVE in terms of bias across all meta-regions.

	US				East			
	1R	1L	12R	12L	1R	1L	12R	12L
1R		1.0556	0.8759	0.9445		1.0846	0.9926	1.0448
1L	0.9473		0.8297	0.8948	0.9220		0.9152	0.9632
12R	1.1417	1.2052		1.0784	1.0074	1.0927		1.0525
12L	1.0587	1.1176	0.9273		0.9572	1.0382	0.9501	
1R		0.9955	0.7780	0.8729		1.1231	0.9118	0.9540
1L	1.0046		0.7815	0.8769	0.8904		0.8119	0.8495
12R	1.2854	1.2795		1.1220	1.0967	1.2317		1.0463
12L	1.1456	1.1404	0.8913		1.0482	1.1772	0.9558	
	1R	1L	12R	12L	1R	1L	12R	12L
	Midwest				West			

Table 3. 3. Fraction of sites where each method outperforms all others.

	NSE				Bias			
	1R	1L	12R	12L	1R	1L	12R	12L
US	14.75	9.89	19.63	58.95	26.71	32.60	21.61	23.36
East	16.22	10.18	15.42	59.32	25.92	31.96	20.28	22.43
Midwest	17.38	13.14	22.39	53.94	25.69	32.17	25.63	24.65
West	9.57	5.67	22.64	63.64	28.92	34.05	19.72	23.51

Table 3. 4. Average relative monthly performance of three MOVE variants compared to MOVE12L.

Meta-Region	Variant	Efficiency: Months where MOVE12L is not dominant	Efficiency: Average	Bias: Months where MOVE12L is not dominant	Bias: Average
US	1R	0	1.8182	0	5.1555
	1L	0	1.7728	0	5.5986
	12R	2	1.0450	10	0.9643
East	1R	0	1.5305	0	3.9274
	1L	0	1.5145	0	4.0077
	12R	0	1.0483	10	0.9553
Midwest	1R	0	1.8063	0	4.8353
	1L	0	1.7662	0	5.1644
	12R	3	1.0437	9	0.9599
West	1R	0	2.6933	0	7.8882
	1L	0	2.5593	0	8.9333
	12R	6	1.0266	9	0.9788

Table 3. 5. Average monthly fraction of sites where each variants outperforms all others.

	NSE				Bias			
	1R	1L	12R	12L	1R	1L	12R	12L
US	18.31	18.04	28.97	39.43	19.70	19.12	39.29	33.01
East	20.03	19.28	25.52	37.91	19.63	17.81	36.52	30.35
Midwest	21.11	20.10	29.48	37.20	21.53	22.70	41.64	32.34
West	12.71	13.92	33.14	43.80	17.54	16.91	40.87	37.51

Table 3. 6. The relative efficiencies of four flow-transfer techniques in terms of NSE across all meta-regions.

	US				East			
	DA	SM	MOVE	QPPQ	DA	SM	MOVE	QPPQ
DA		0.2820	0.1949	0.3548		0.3851	0.3253	0.4828
SM	3.5462		0.6913	1.2583	2.5967		0.8446	1.2538
MOVE	5.1301	1.4466		1.8203	3.0746	1.1840		1.4845
QPPQ	2.8183	0.7947	0.5494		2.0711	0.7976	0.6736	
DA		0.3694	0.2349	0.4134		0.1278	0.0817	0.2092
SM	2.7072		0.6359	1.1192	7.8275		0.6395	1.6376
MOVE	4.2572	1.5725		1.7600	12.2402	1.5637		2.5608
QPPQ	2.4188	0.8935	0.5682		4.7797	0.6106	0.3905	
	DA	SM	MOVE	QPPQ	DA	SM	MOVE	QPPQ
	Midwest				West			

Table 3. 7. The relative efficiencies of four flow-transfer techniques in terms of bias across all meta-regions.

	US				East			
	DA	SM	MOVE	QPPQ	DA	SM	MOVE	QPPQ
DA		0.0266	0.0255	0.0279		0.0358	0.0346	0.0358
SM	37.6431		0.9602	1.0518	27.9288		0.9665	0.9994
MOVE	39.2041	1.0415		1.0955	28.8981	1.0347		1.0341
QPPQ	35.7878	0.9507	0.9129		27.9446	1.0006	0.9670	
DA		0.0343	0.0320	0.0375		0.0145	0.0144	0.0154
SM	29.1175		0.9331	1.0925	68.9768		0.9936	1.0606
MOVE	31.2050	1.0717		1.1709	69.4208	1.0064		1.0674
QPPQ	26.6514	0.9153	0.8541		65.0343	0.9428	0.9368	
	DA	SM	MOVE	QPPQ	DA	SM	MOVE	QPPQ
	Midwest				West			

Table 3. 8. Percentage of sites where each method outperforms all others.

	NSE				Bias			
	DA	SM	MOVE	QPPQ	DA	SM	MOVE	QPPQ
US	9.34	20.95	58.77	17.95	18.23	37.50	30.72	26.99
East	5.45	22.38	56.59	18.65	9.62	35.47	29.75	28.54
Midwest	19.68	21.32	53.03	18.79	32.80	40.00	31.34	21.99
West	0.95	18.57	67.71	15.95	12.23	38.11	31.58	29.58

Table 3. 9. Average monthly relative efficiency of three methods compared to MOVE.

Meta-Region	Variant	Efficiency: Months where MOVE12L is not dominant	Efficiency: Average	Bias: Months where MOVE12L is not dominant	Bias: Average
US	DA	0	4.1292	0	19.5135
	SM	0	1.4857	6	1.0257
	QPPQ	0	1.6360	0	4.0154
East	DA	0	3.0413	0	14.4525
	SM	0	1.4015	6	1.0074
	QPPQ	0	1.4148	0	3.2431
Midwest	DA	0	3.6024	0	15.4185
	SM	0	1.4960	4	1.0338
	QPPQ	0	1.6192	0	3.9498
West	DA	0	9.6046	0	37.0189
	SM	0	1.7251	6	1.0288
	QPPQ	0	2.2801	0	5.3466

Table 3. 10. Average monthly fraction of sites where each method outperforms all others.

	NSE				Bias			
	DA	SM	MOVE	QPPQ	DA	SM	MOVE	QPPQ
US	12.66	24.48	43.44	28.08	15.78	38.66	31.43	27.59
East	11.31	23.50	40.91	29.37	14.38	36.37	28.20	27.04
Midwest	18.03	24.01	40.70	29.82	21.48	40.34	30.61	28.14
West	7.24	26.30	49.88	24.22	9.62	40.26	36.86	27.89

CHAPTER FIVE:

Table 5. 1. Relative efficiencies in terms of NSE for four variants of MOVE with regional regression.

	US				East			
	1R	1L	12R	12L	1R	1L	12R	12L
1R		0.9300	0.5552	0.5034		0.8912	0.3681	0.3611
1L	1.0753		0.5970	0.5413	1.1221		0.4131	0.4052
12R	1.8011	1.6750		0.9067	2.7165	2.4210		0.9810
12L	1.9864	1.8473	1.1029		2.7690	2.4677	1.0193	
1R		0.9443	0.8493	0.7005		0.9479	0.4645	0.4584
1L	1.0589		0.8993	0.7418	1.0550		0.4900	0.4836
12R	1.1774	1.1119		0.8248	2.1531	2.0409		0.9869
12L	1.4275	1.3480	1.2124		2.1816	2.0679	1.0132	
	1R	1L	12R	12L	1R	1L	12R	12L
	Midwest				West			

Table 5. 2. Relative efficiencies in terms of bias for four variants of MOVE with regional regression.

	US				East			
	1R	1L	12R	12L	1R	1L	12R	12L
1R		1.1310	1.1850	1.3155		1.1424	1.7854	1.8460
1L	0.8842		1.0478	1.1631	0.8753		1.5628	1.6158
12R	0.8439	0.9544		1.1101	0.5601	0.6399		1.0340
12L	0.7602	0.8597	0.9008		0.5417	0.6189	0.9672	
1R		1.1778	1.2167	1.3853		1.0916	1.0618	1.1837
1L	0.8491		1.0331	1.1762	0.9161		0.9727	1.0844
12R	0.8219	0.9680		1.1386	0.9418	1.0281		1.1148
12L	0.7218	0.8502	0.8783		0.8448	0.9222	0.8970	
	1R	1L	12R	12L	1R	1L	12R	12L
	Midwest				West			

Table 5. 3. Percentage of sites where each variant of MOVE with regional regression outperforms all others.

	NSE				Bias			
	1R	1L	12R	12L	1R	1L	12R	12L
US	5.50	12.60	47.73	43.98	31.55	17.91	31.32	25.35
East	3.54	6.08	47.56	45.79	28.46	21.91	26.32	23.80
Midwest	8.38	18.44	42.52	43.42	32.21	15.81	33.54	26.17
West	4.58	14.88	55.40	41.46	35.93	13.97	37.44	27.05

FIGURES

CHAPTER TWO

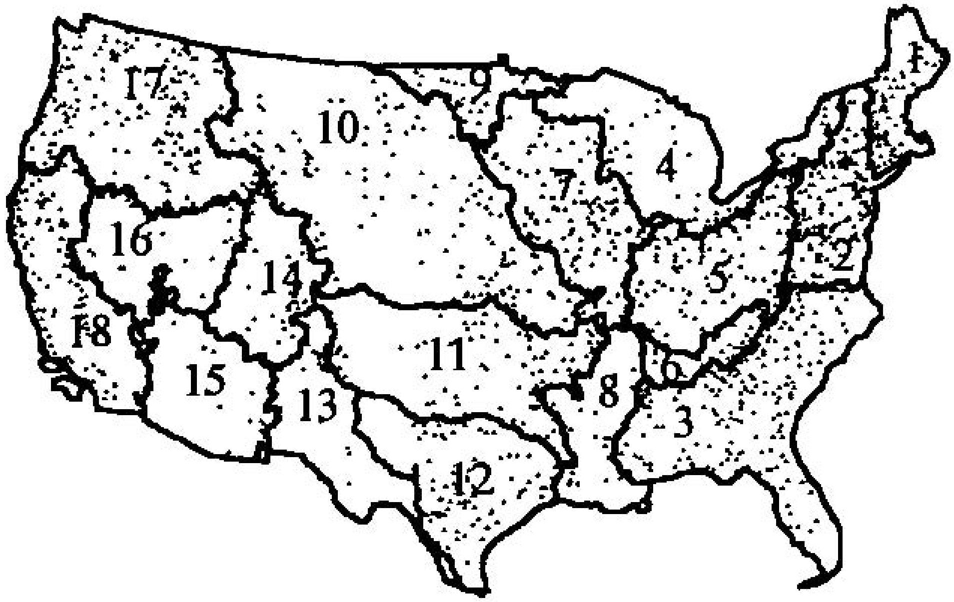

Figure 2. 1. Two-digit HUCs of the continental United States.

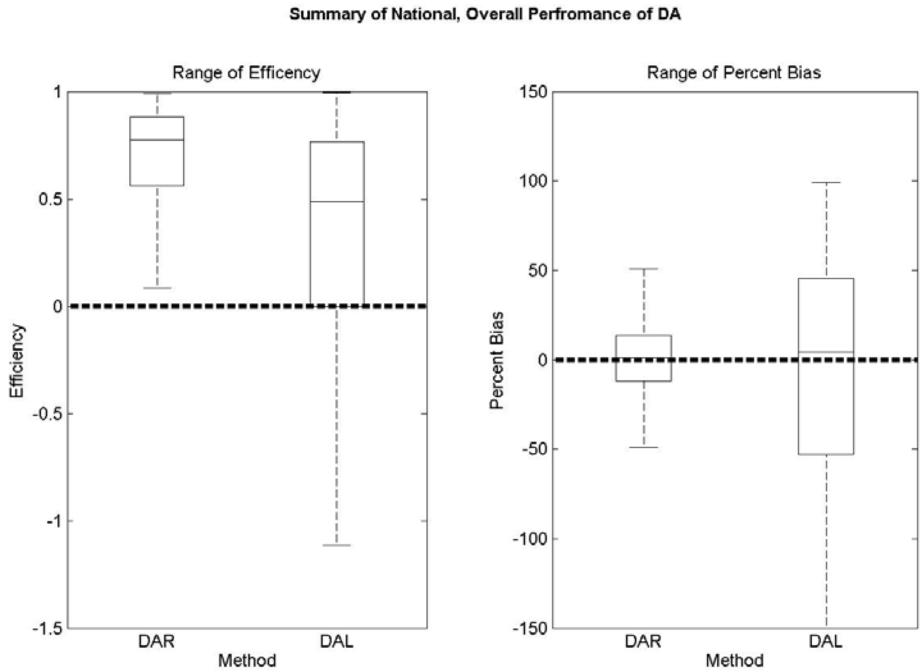

Figure 3. 1: Summary of National, Overall Performance of Drainage Area Methods.

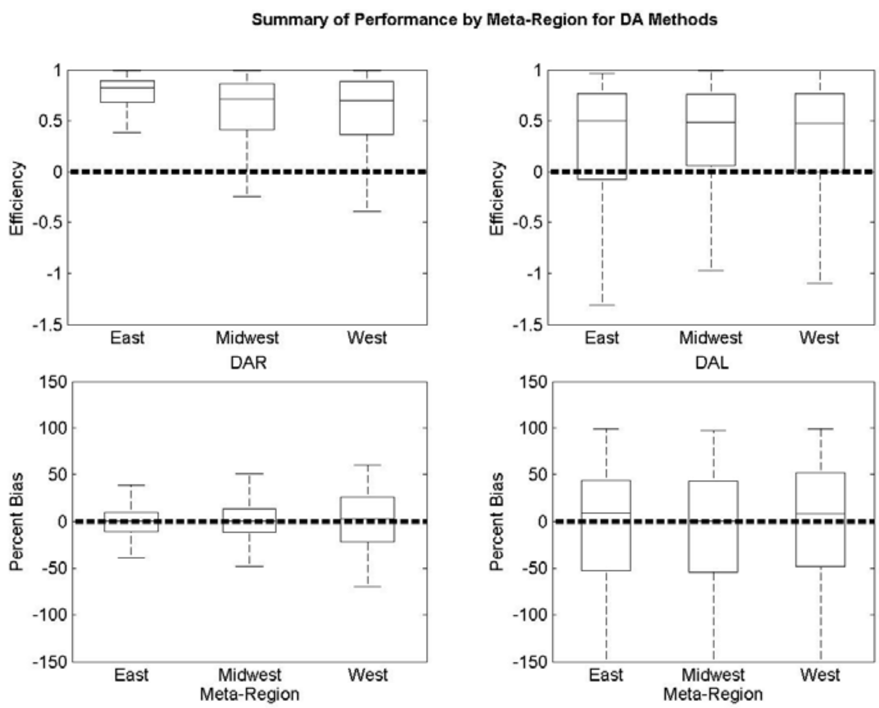

Figure 3. 2: Summary of Performance of DA Methods by Meta-Region.

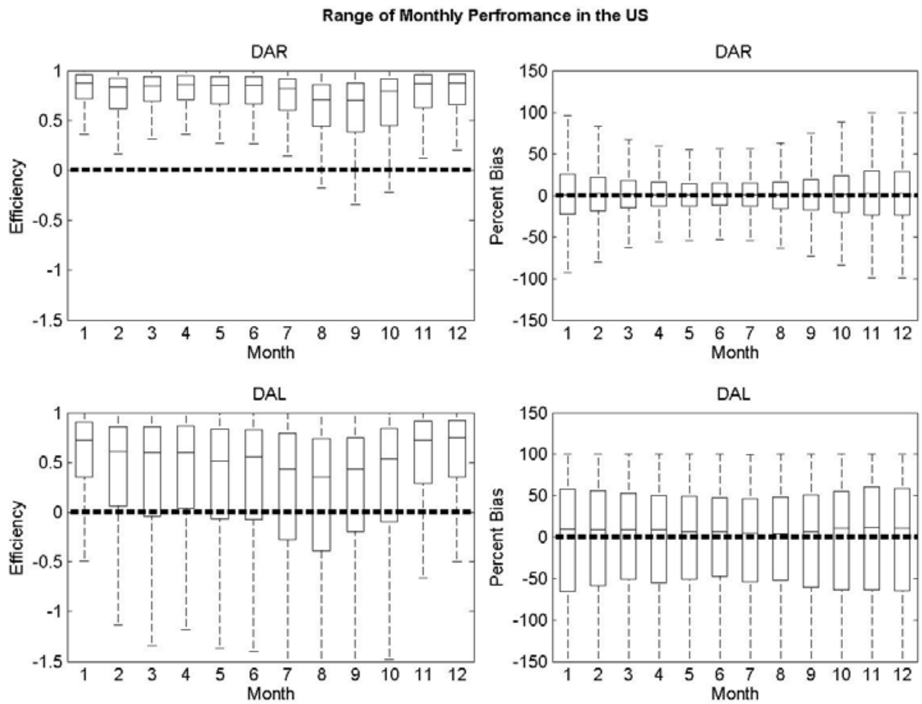

Figure 3. 3. Range of national monthly performance of DAR and DAL.

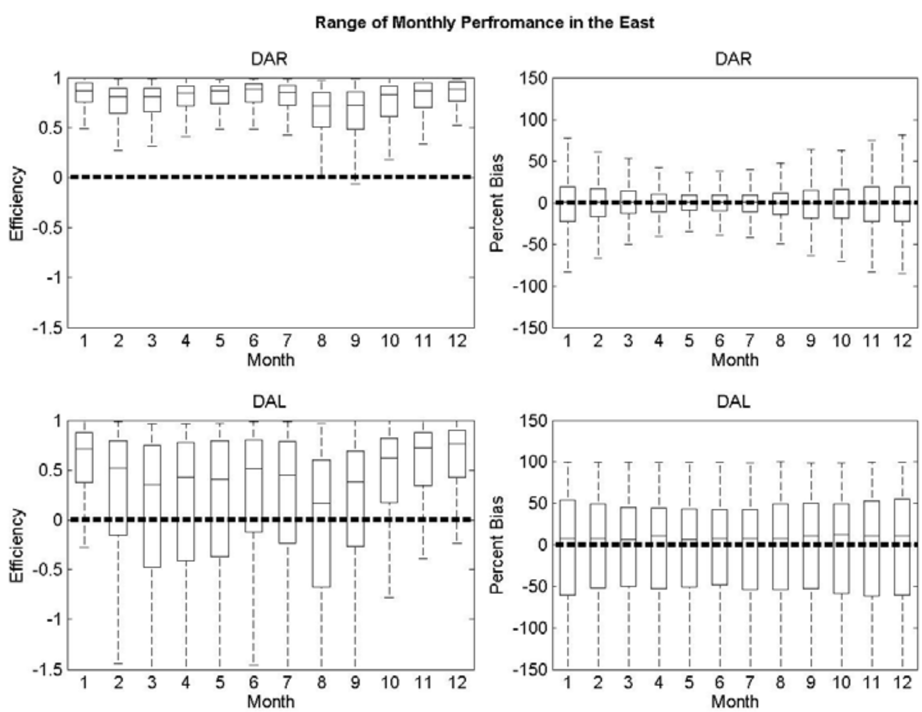

Figure 3. 4. Range of monthly performance of DAR and DAL in the East.

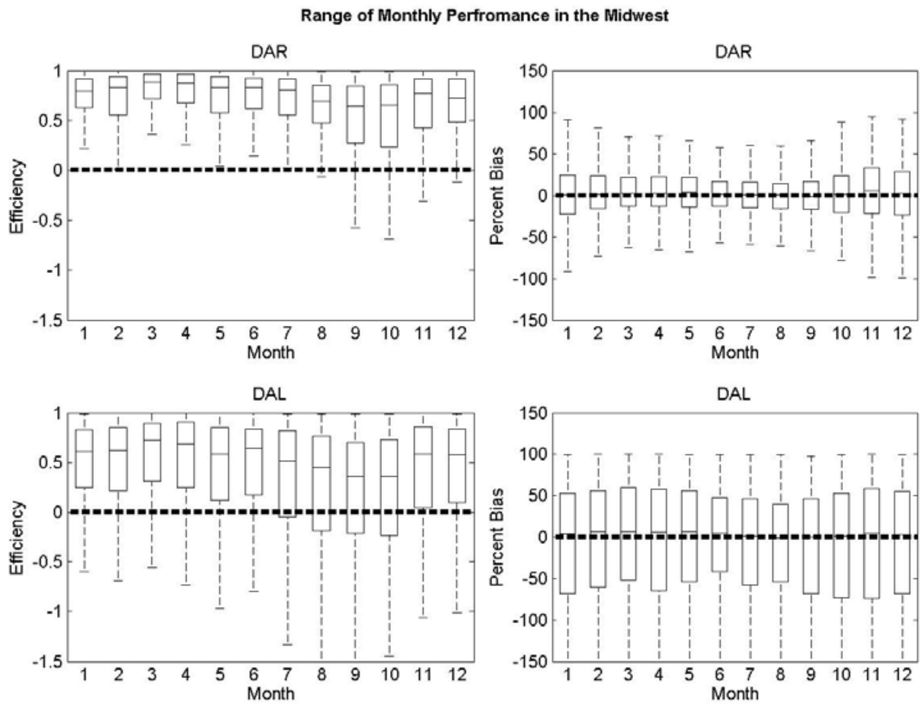

Figure 3. 5. Range of monthly performance of DAR and DAL in the Midwest.

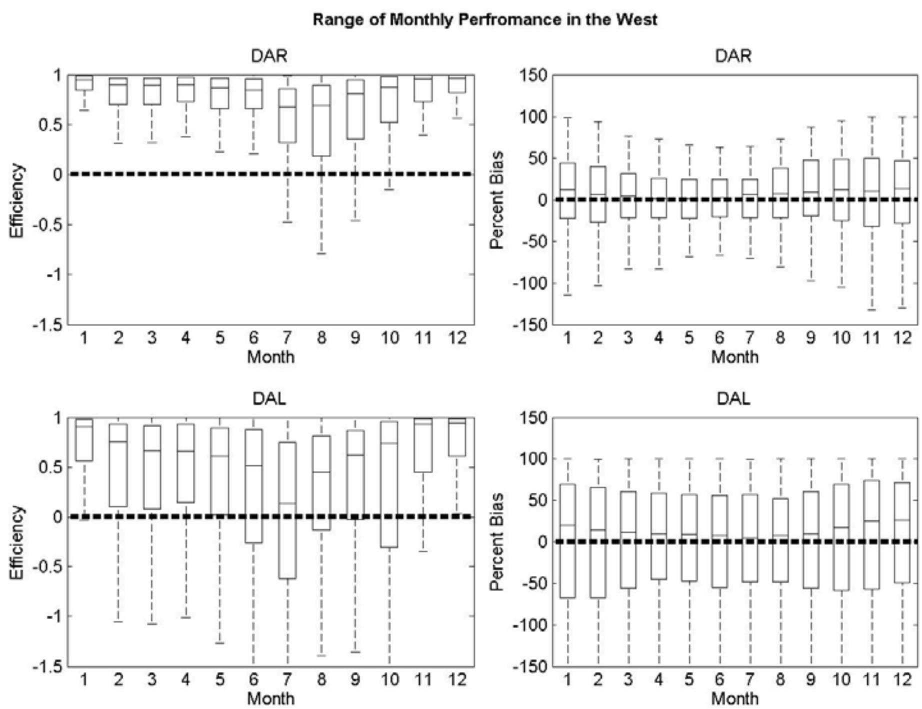

Figure 3. 6. Range of monthly performance of DAR and DAL in the West.

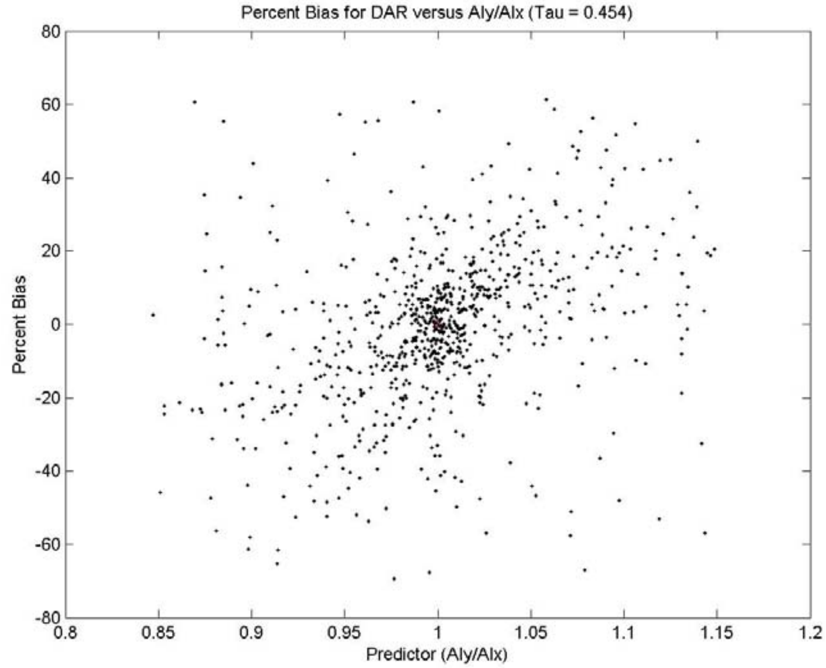

Figure 3. 7. Correlation of the ratio of aridity to bias for DAR.

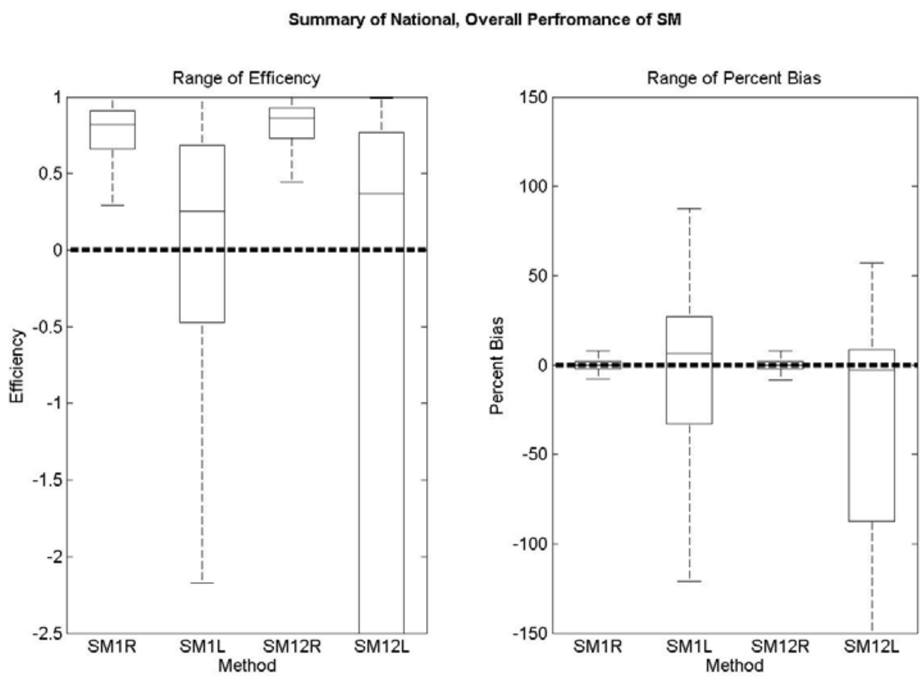

Figure 3. 8. National, overall performance of SM methods.

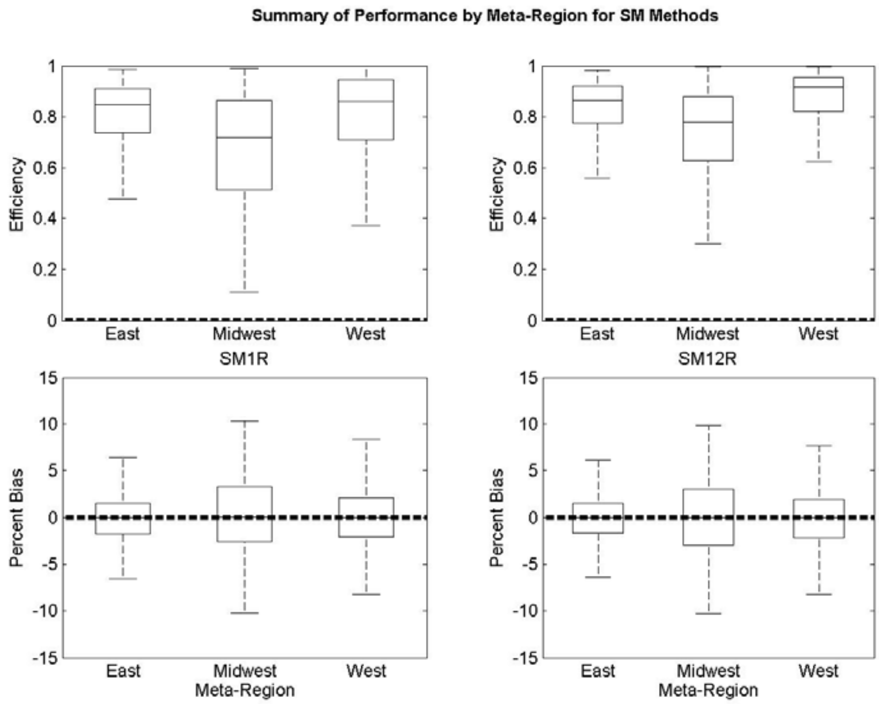

Figure 3. 9. Overall performance of SM1R and SM12R by meta-region.

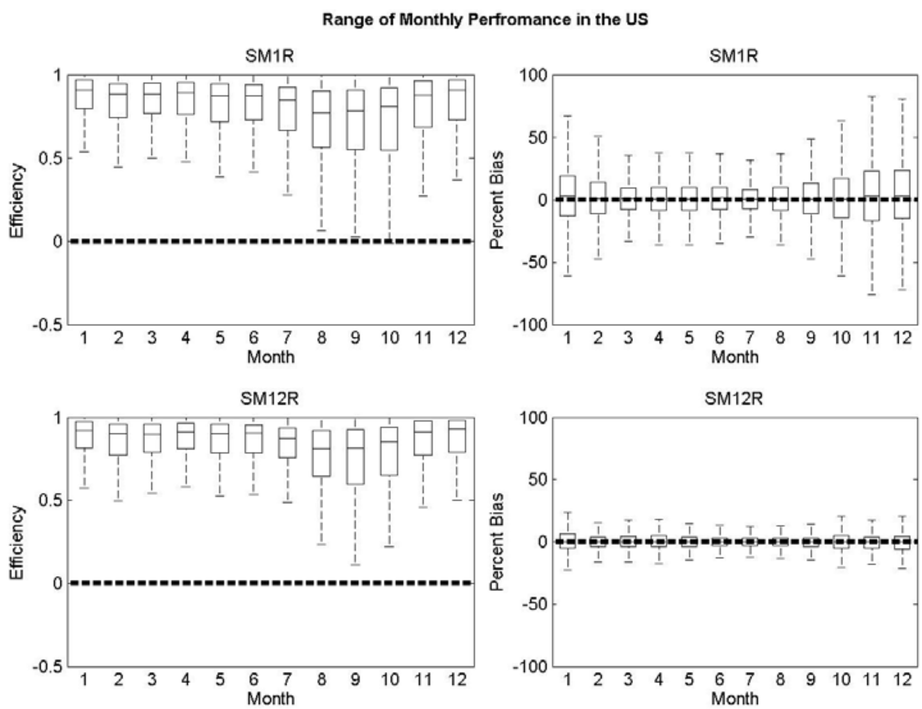

Figure 3. 10. National, monthly performance of SM1R and SM12R.

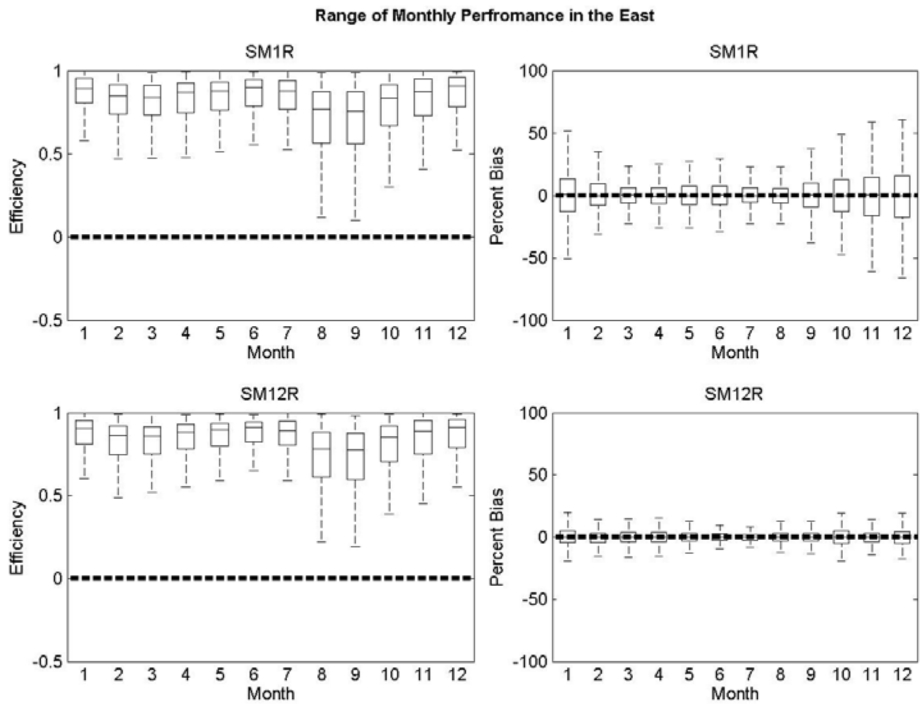

Figure 3. 11. Monthly performance of SM1R and SM12R in the East.

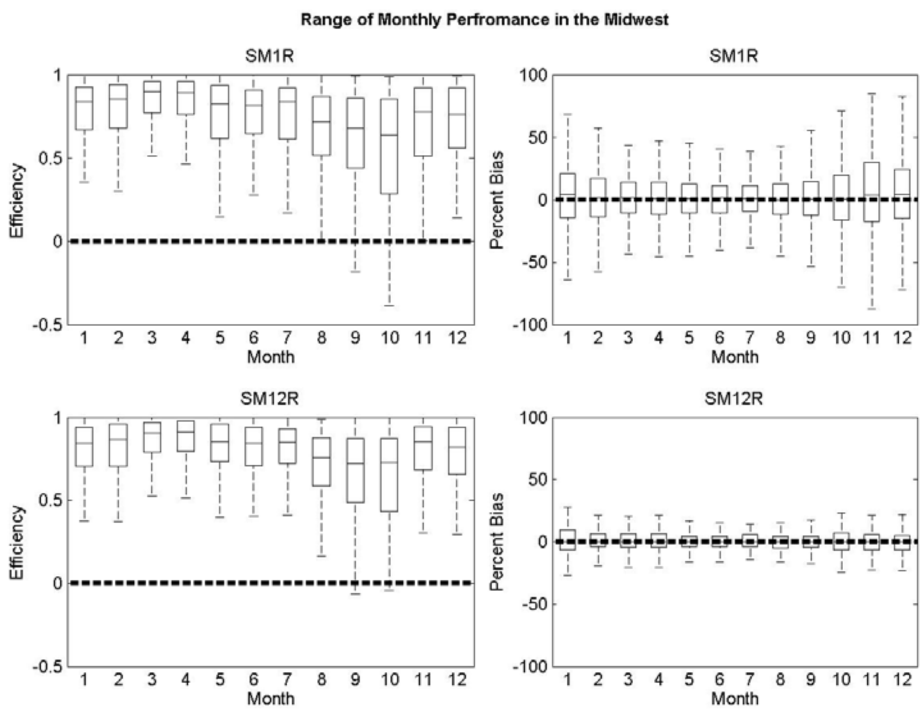

Figure 3. 12. Monthly performance of SM1R and SM12R in the Midwest.

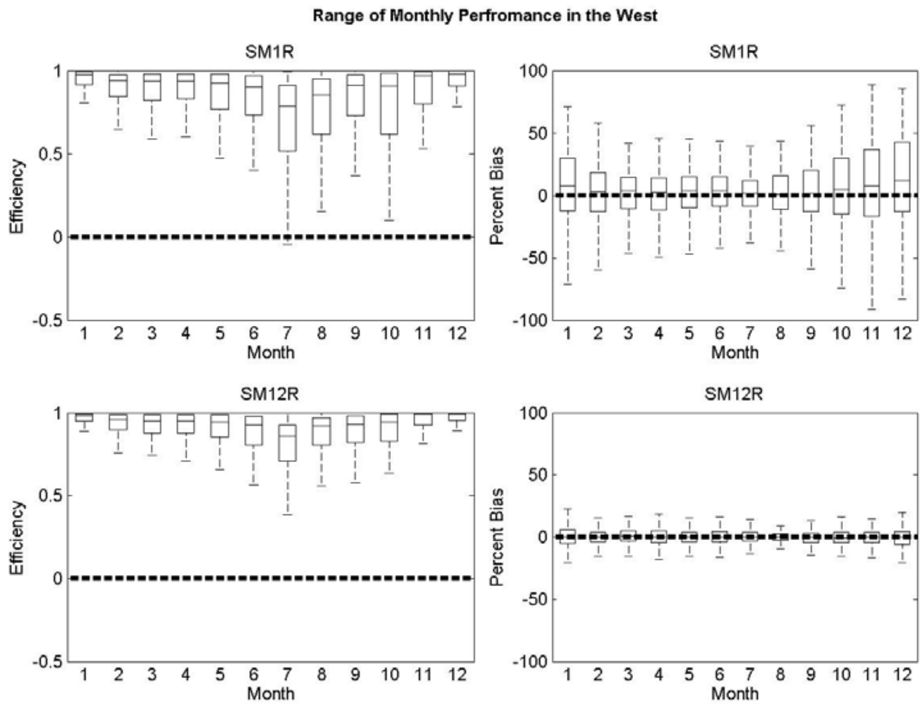

Figure 3. 13. Monthly performance of SM1R and SM12R in the West.

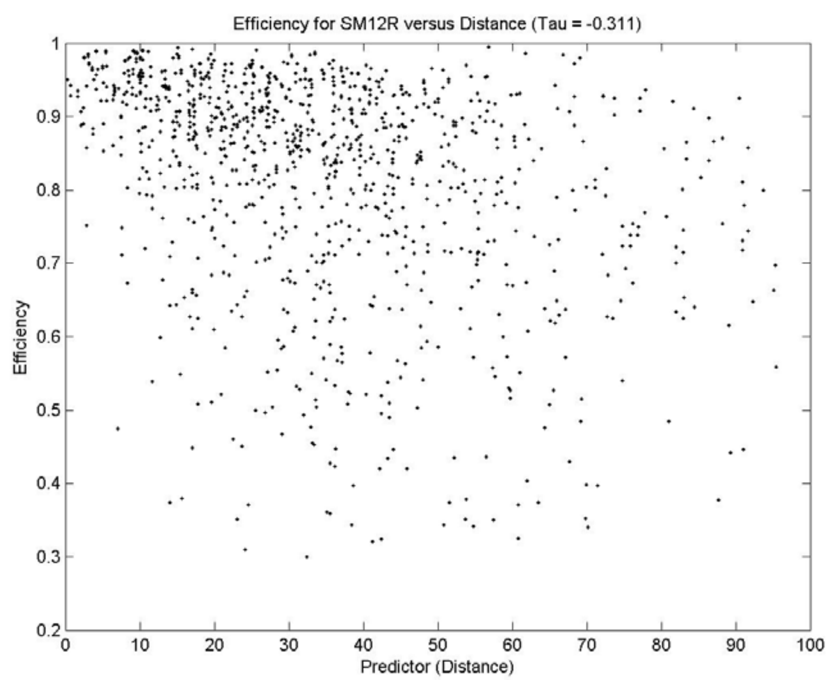

Figure 3. 14. Correlation of distance to index gauge with Nash-Sutcliffe Efficiency.

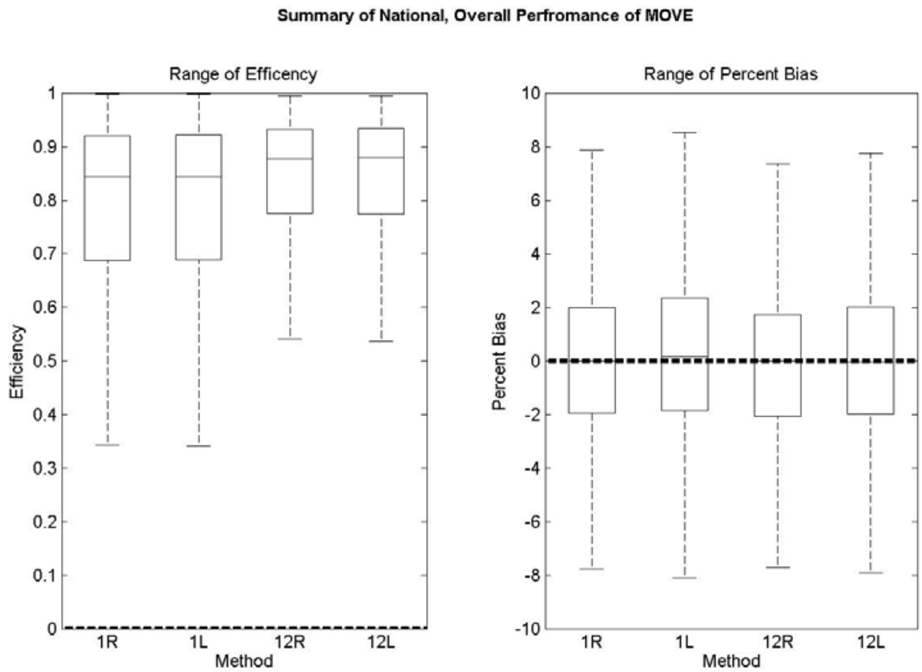

Figure 3. 15. National, overall performance of MOVE variants.

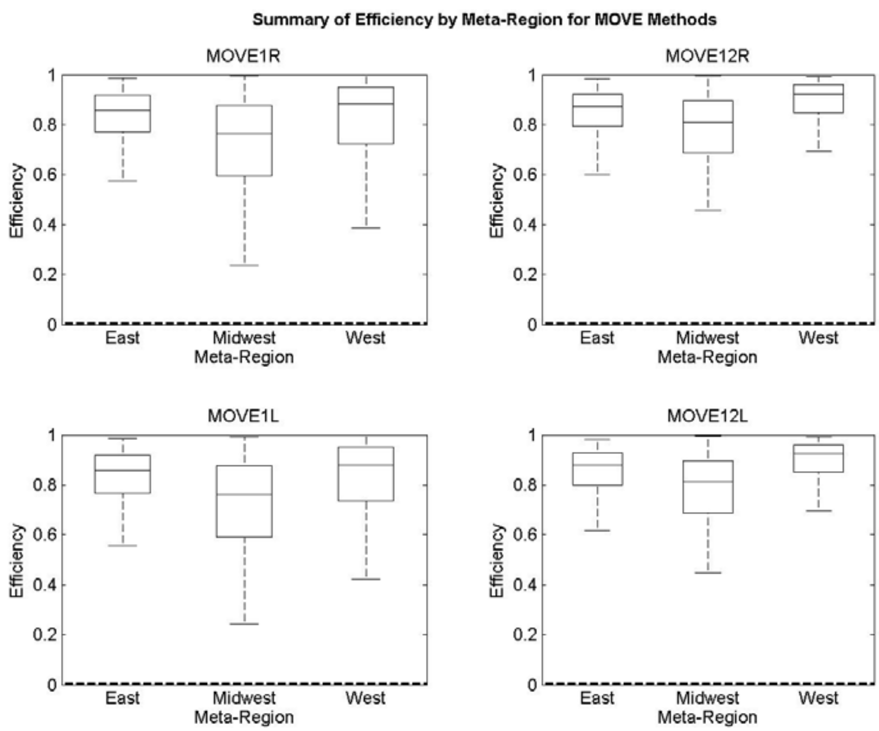

Figure 3. 16. Meta-regional range of NSE for four MOVE variants.

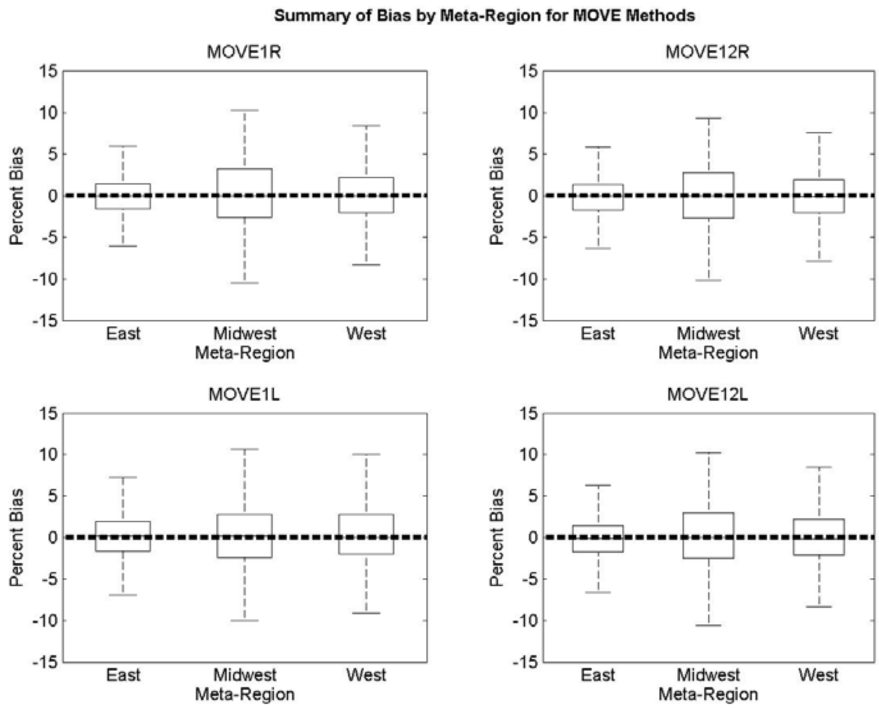

Figure 3. 17. Meta-regional range of bias for four MOVE variants.

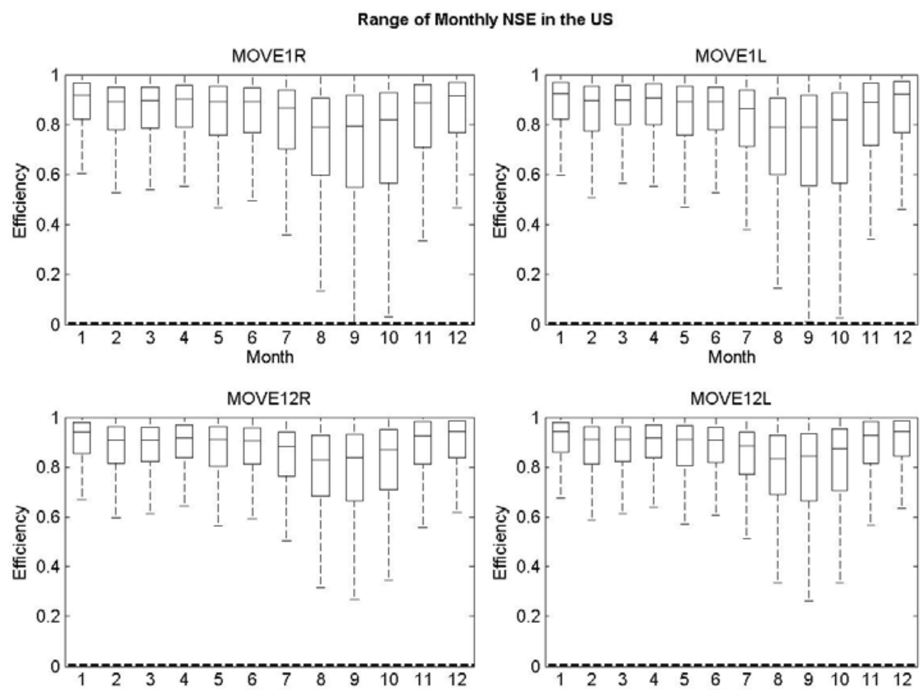

Figure 3. 18. National monthly range of NSE for four MOVE variants.

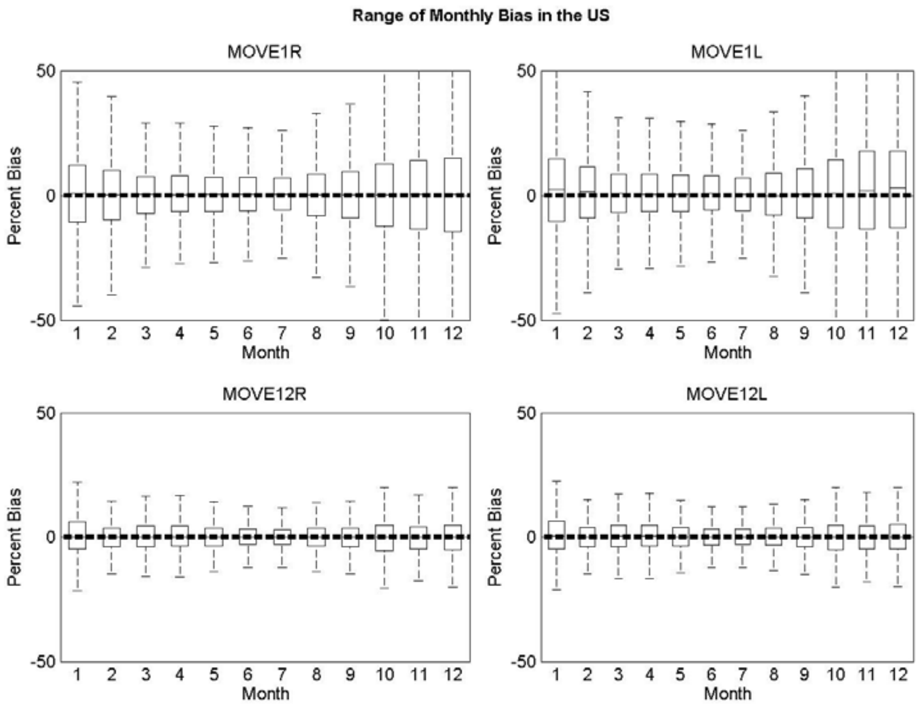

Figure 3. 19. National monthly range of bias for four MOVE variants.

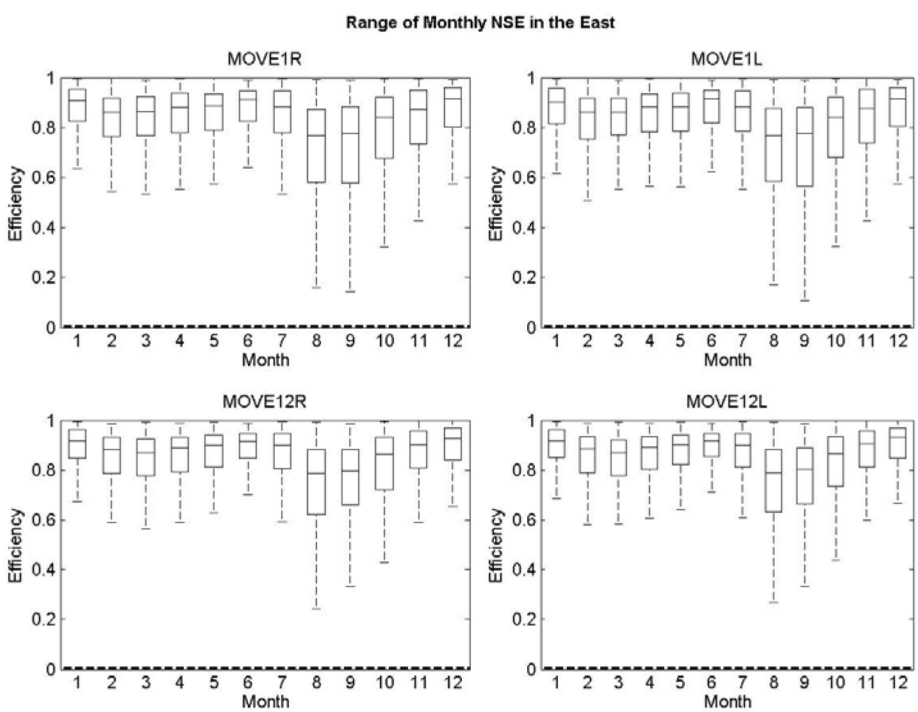

Figure 3. 20. Monthly range of NSE for four MOVE variants in the East.

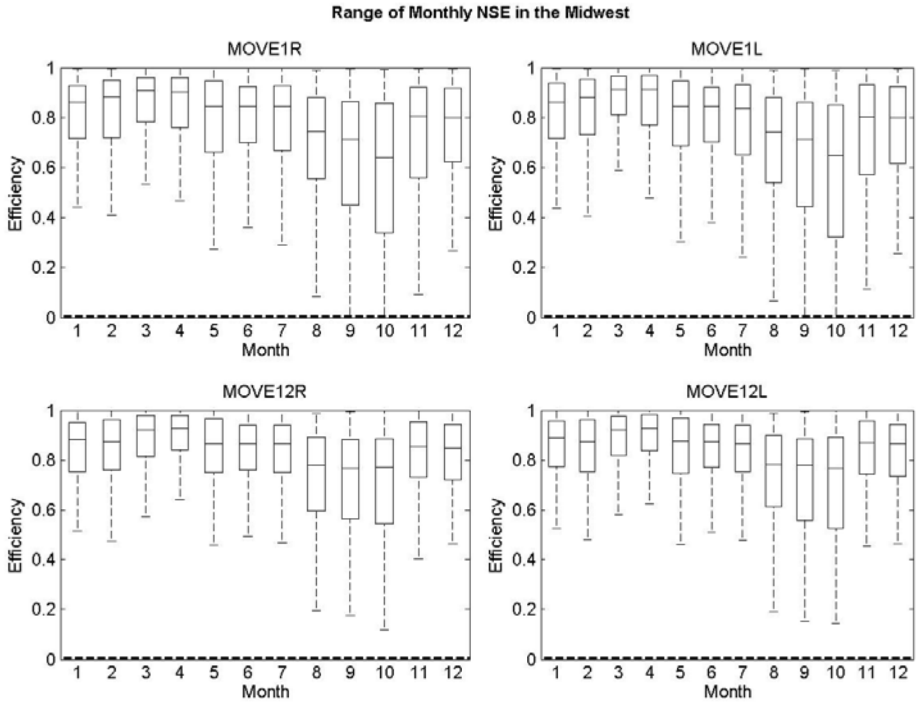

Figure 3. 21. Monthly range of NSE for four MOVE variants in the Midwest.

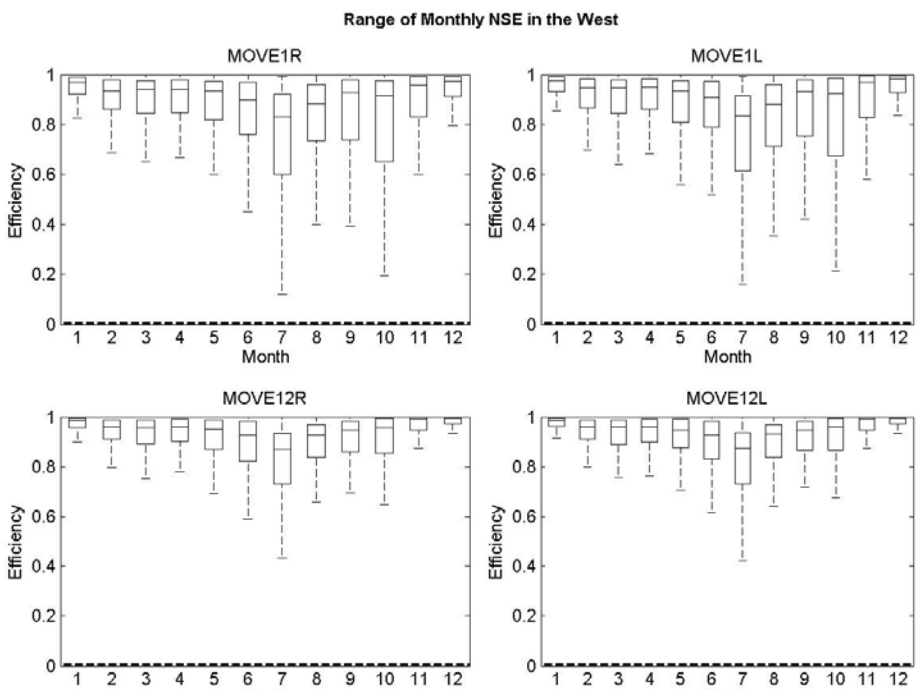

Figure 3. 22. Monthly range of NSE for four variants of MOVE in the West.

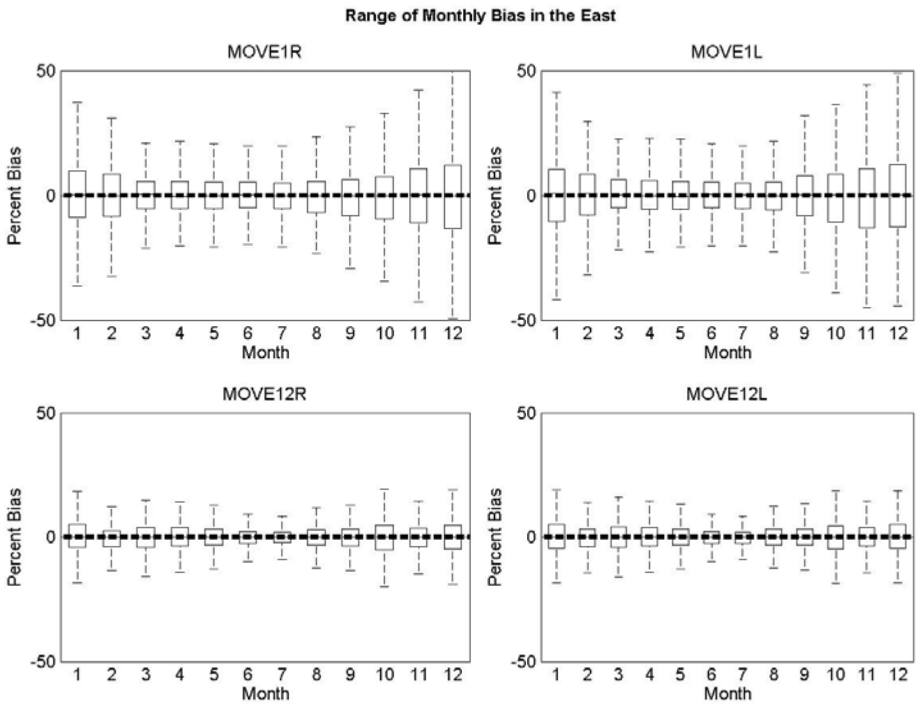

Figure 3. 23. Monthly range of bias for four variants of MOVE in the East.

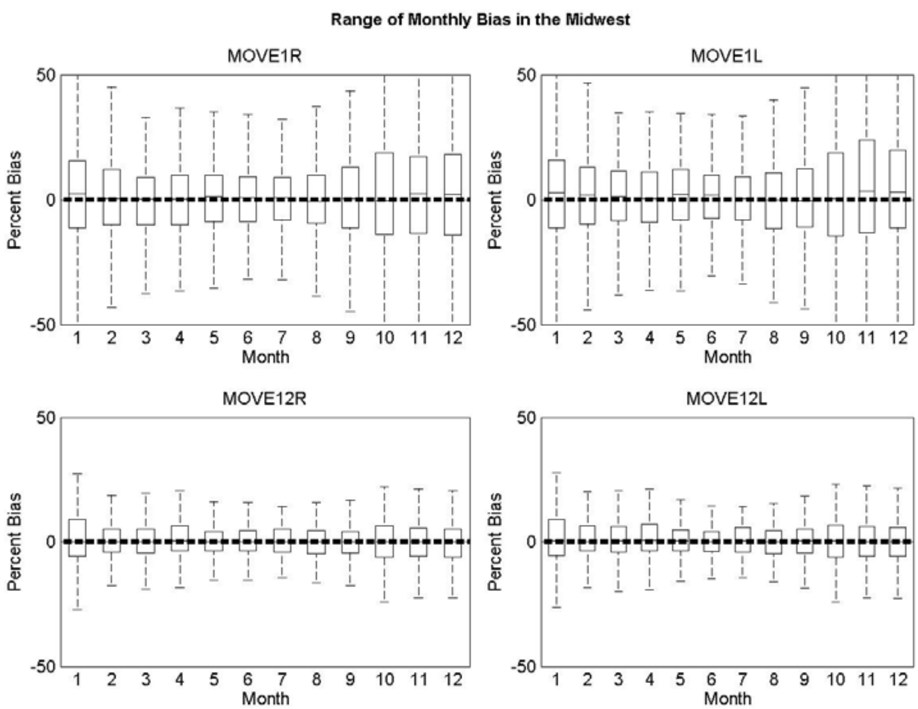

Figure 3. 24. Monthly range of bias for four MOVE variants in the Midwest.

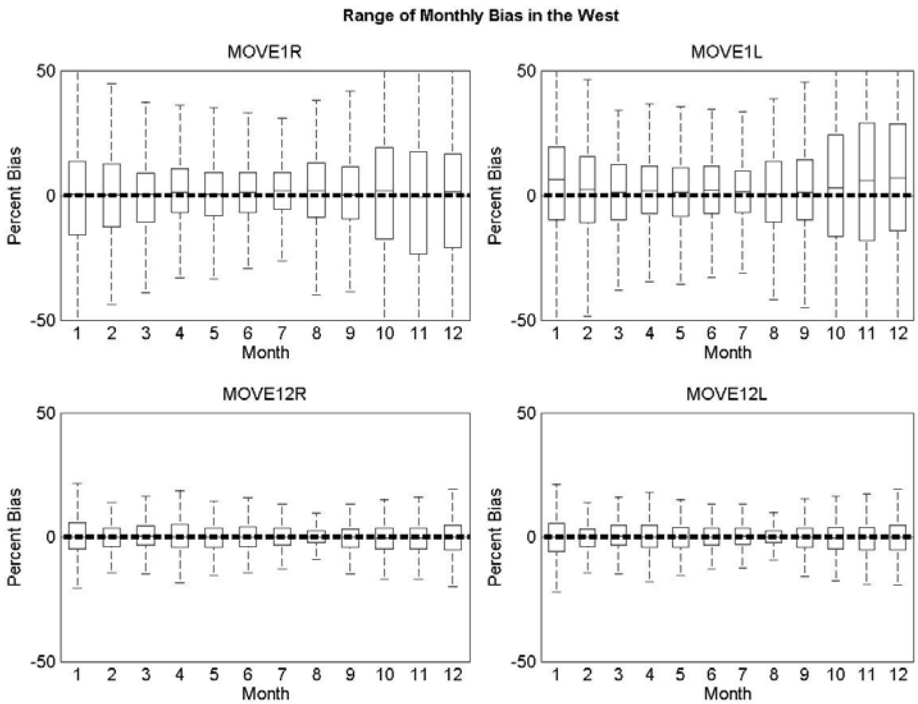

Figure 3. 25. Monthly range of bias for four MOVE variants in the West.

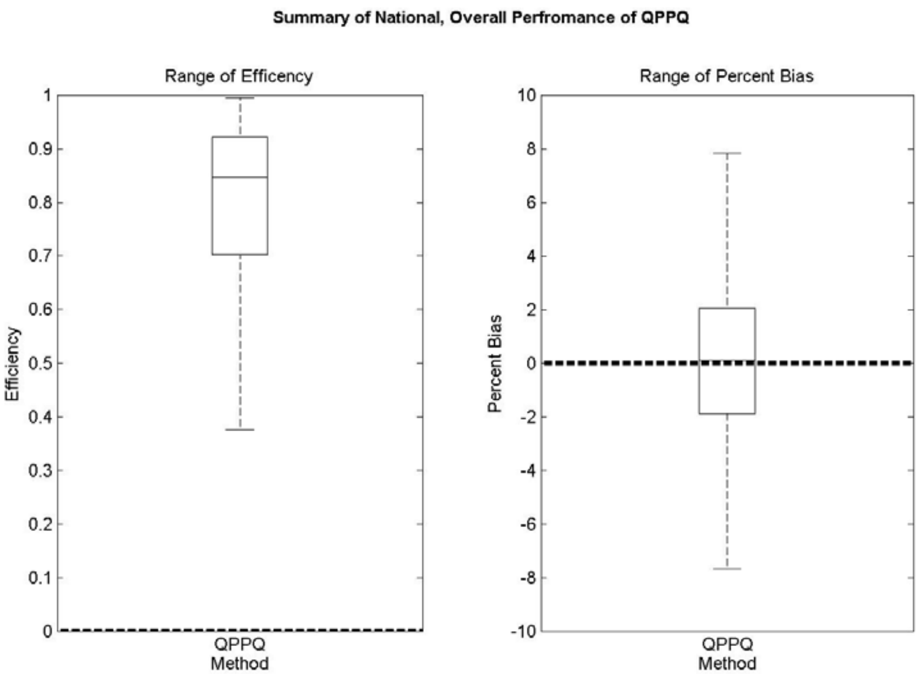

Figure 3. 26. National, overall performance of QPPQ.

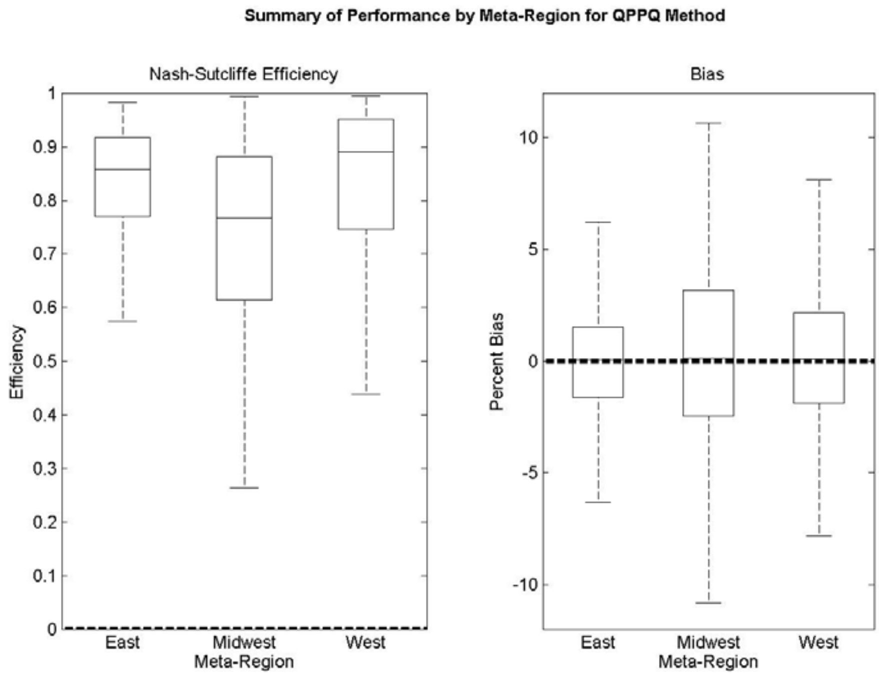

Figure 3. 27. Meta-regional, overall performance of QPPQ.

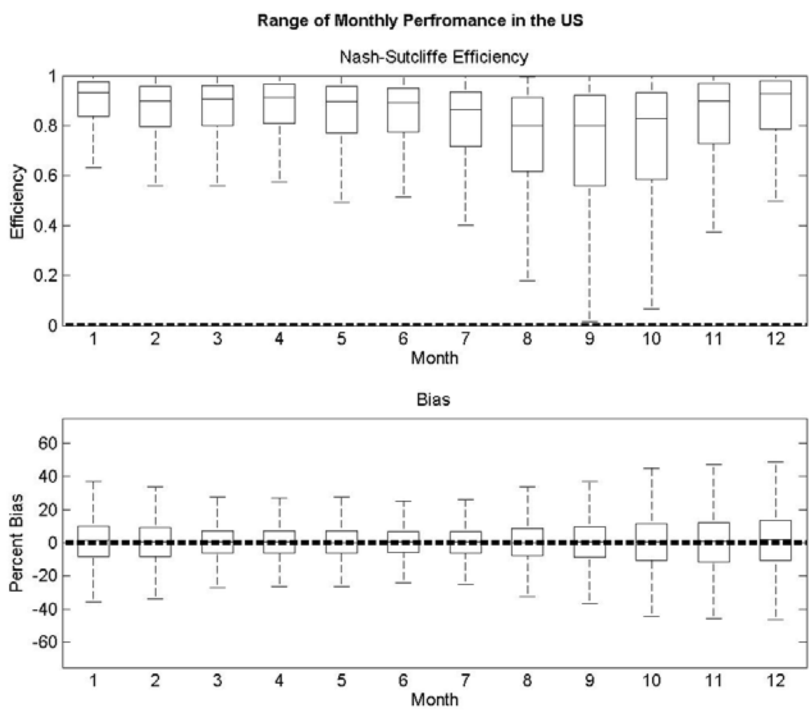

Figure 3. 28. National, monthly performance of QPPQ.

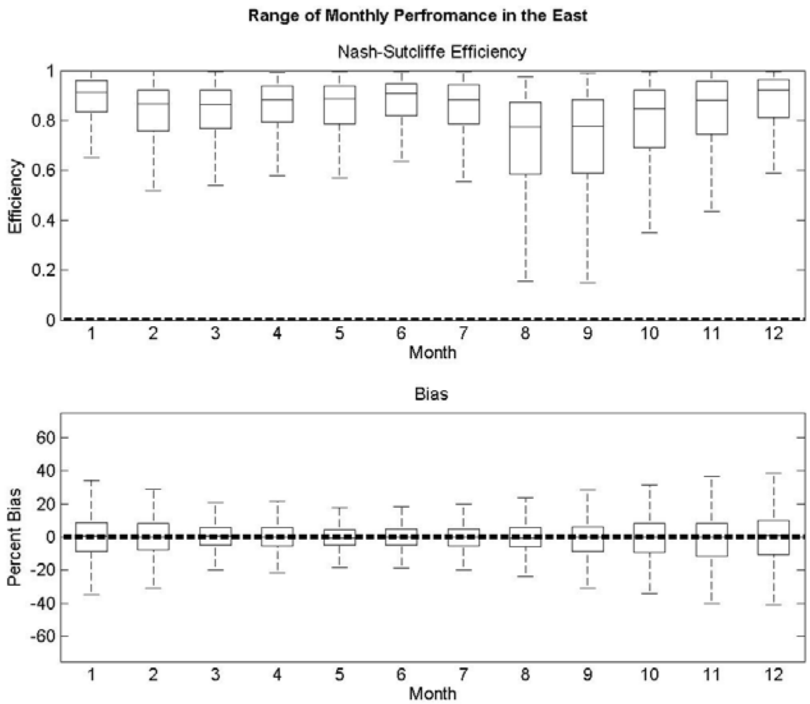

Figure 3. 29. Monthly performance of QPPQ in the East.

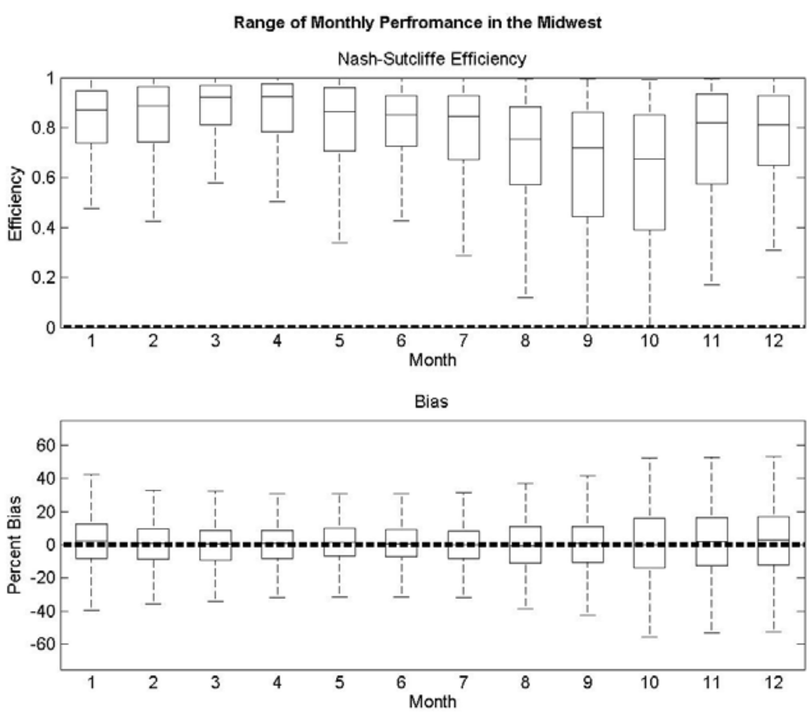

Figure 3. 30. Monthly performance of QPPQ in the Midwest.

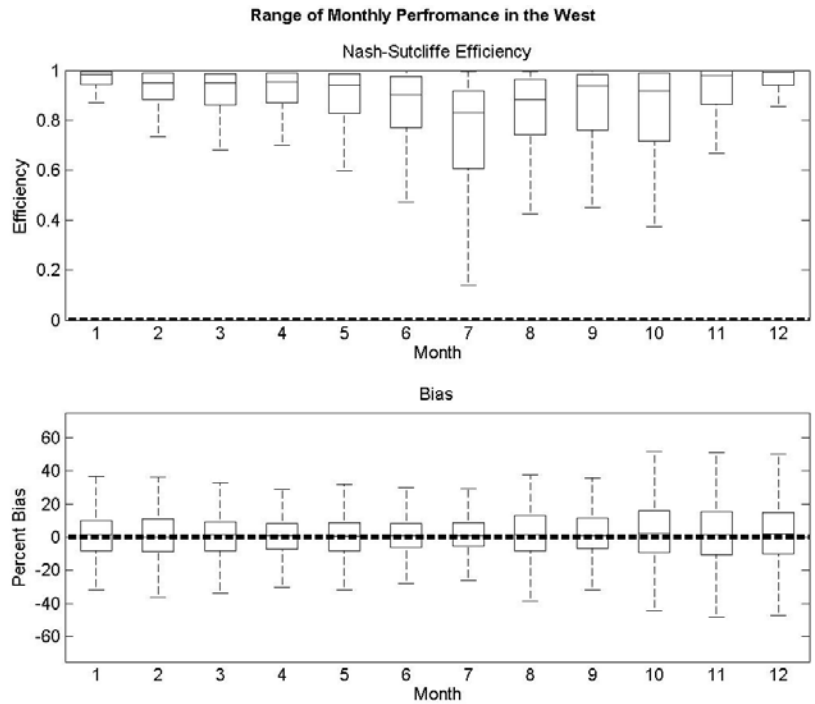

Figure 3. 31. Monthly performance of QPPQ in the West.

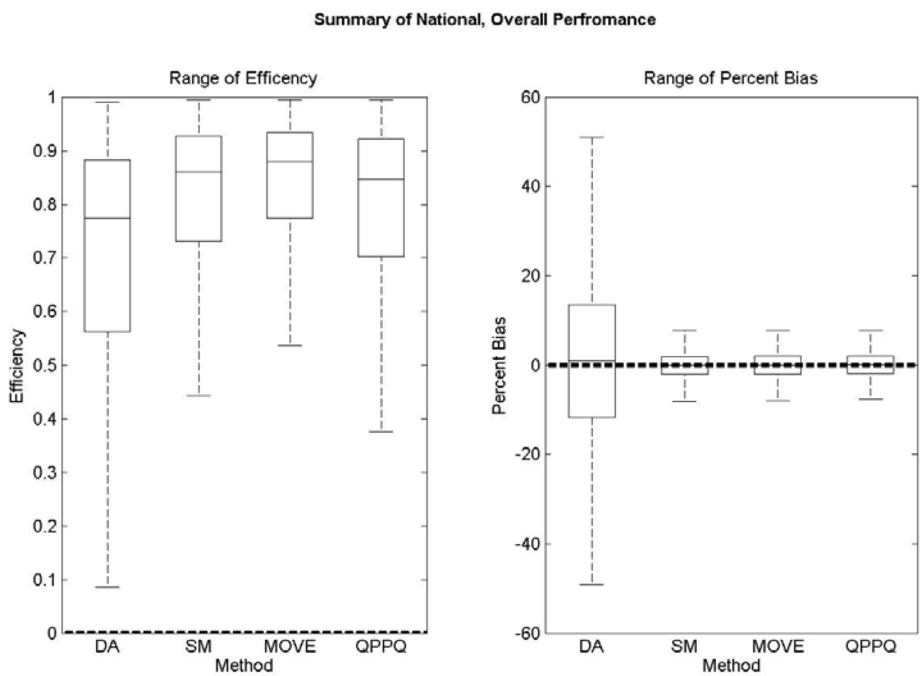

Figure 3. 32. National, overall performance of four flow-transfer techniques.

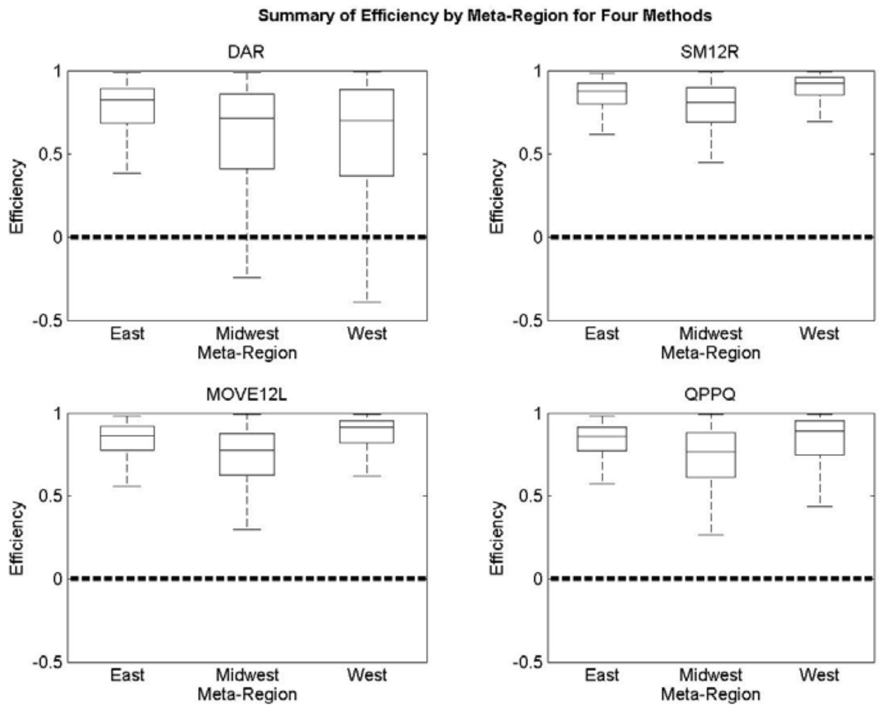

Figure 3. 33. Meta-regional range of NSE for four flow-transfer techniques.

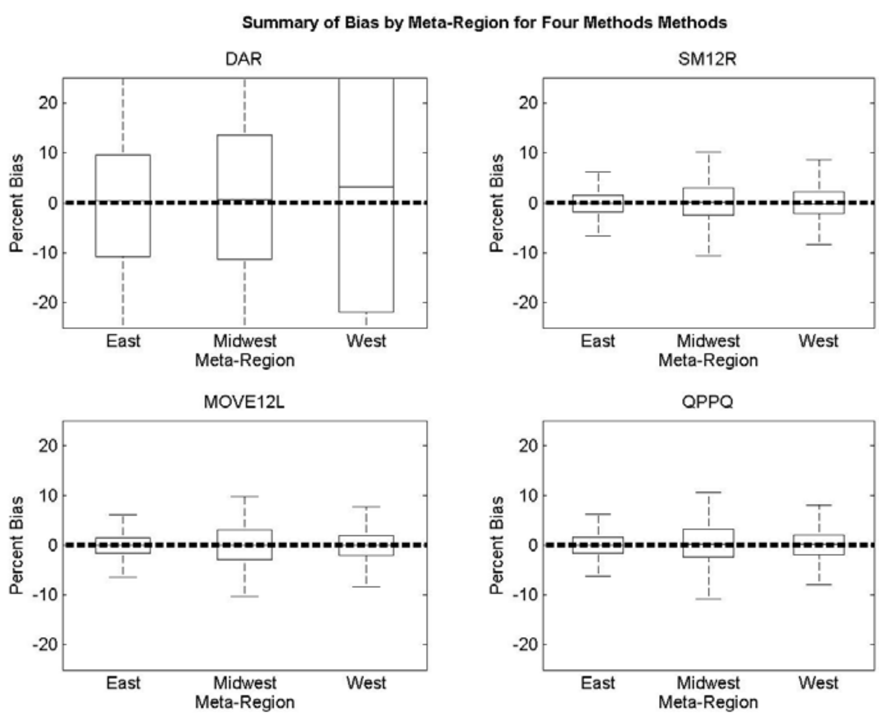

Figure 3. 34. Meta-regional range of bias for four flow-transfer techniques.

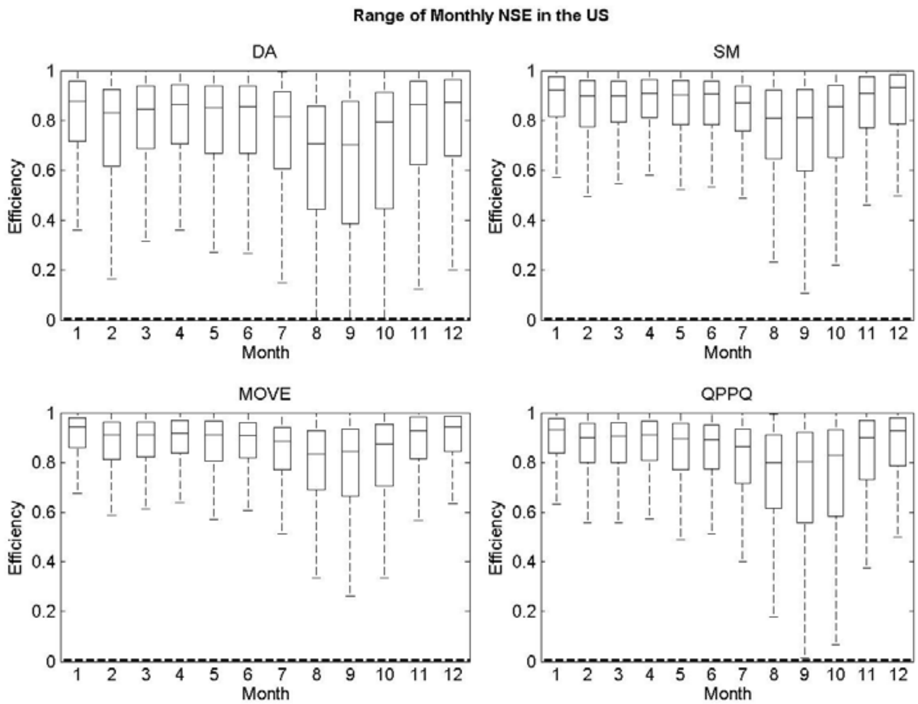

Figure 3. 35. National, monthly range of NSE for four flow-transfer techniques.

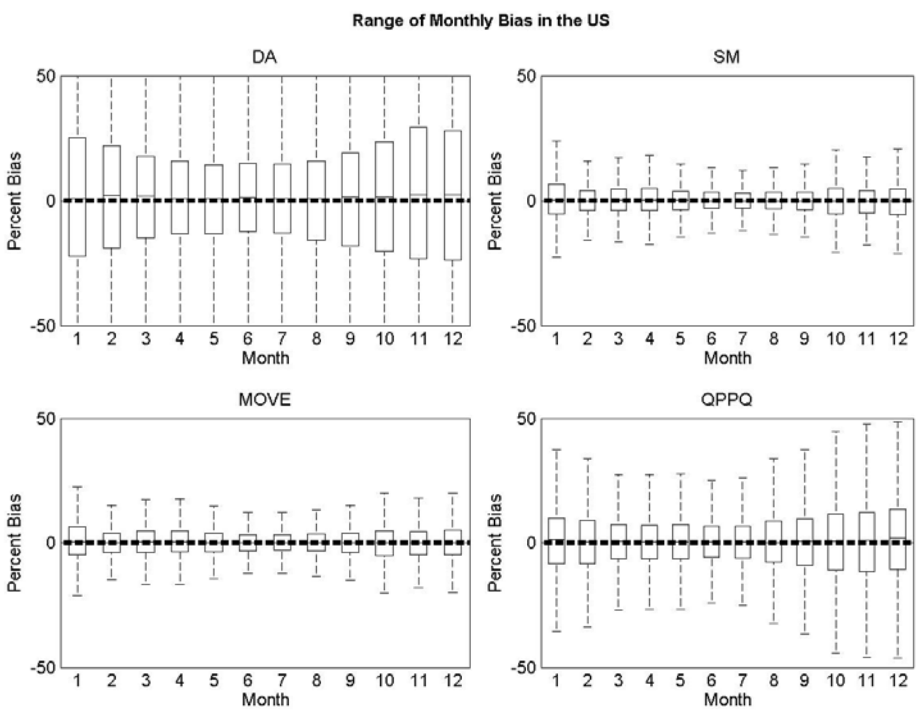

Figure 3. 36. National, monthly range of bias for four flow-transfer techniques.

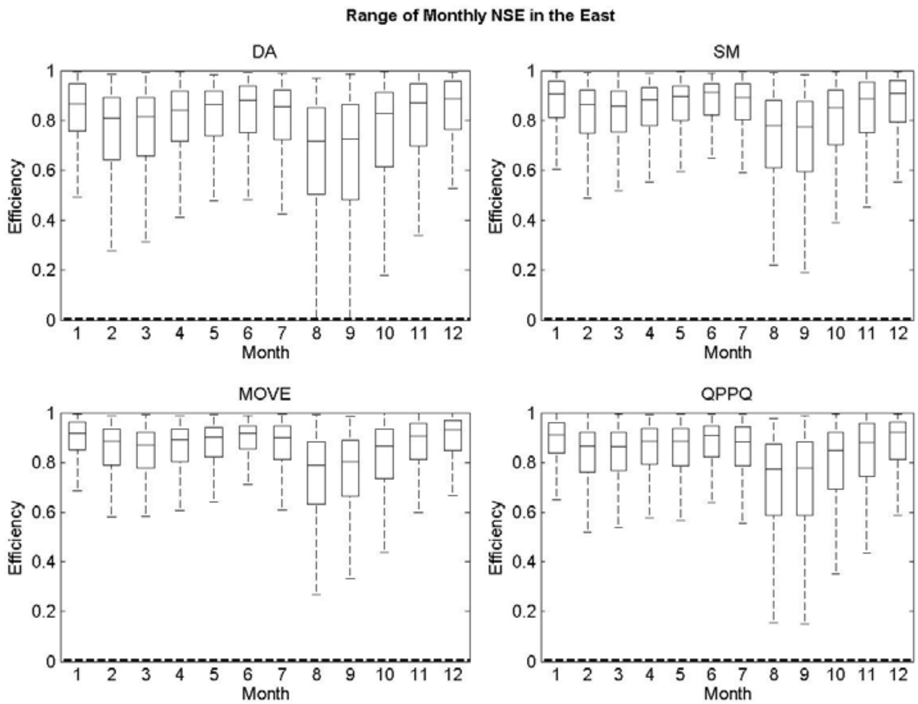

Figure 3. 37. Monthly range of NSE for four flow-transfer techniques in the East.

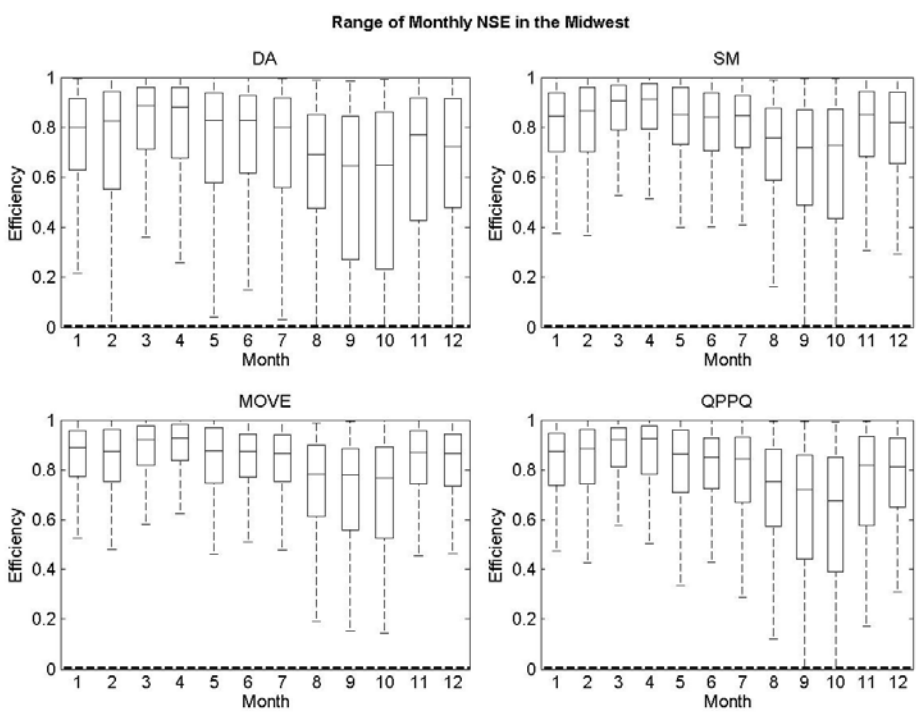

Figure 3. 38. Monthly range of NSE for four flow-transfer techniques in the Midwest.

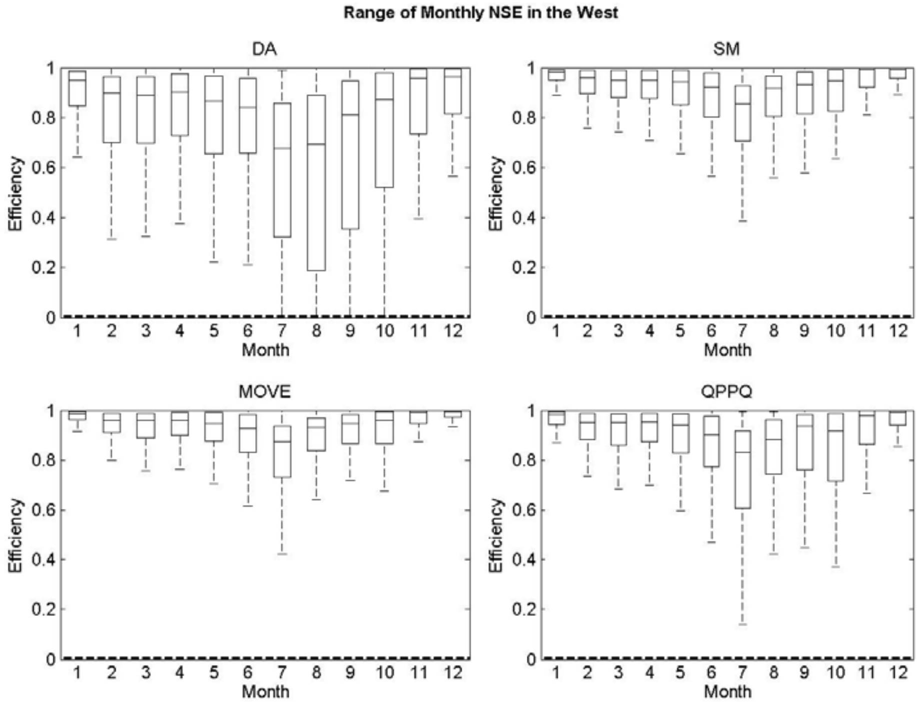

Figure 3. 39. Monthly range of NSE for four flow-transfer techniques in the West.

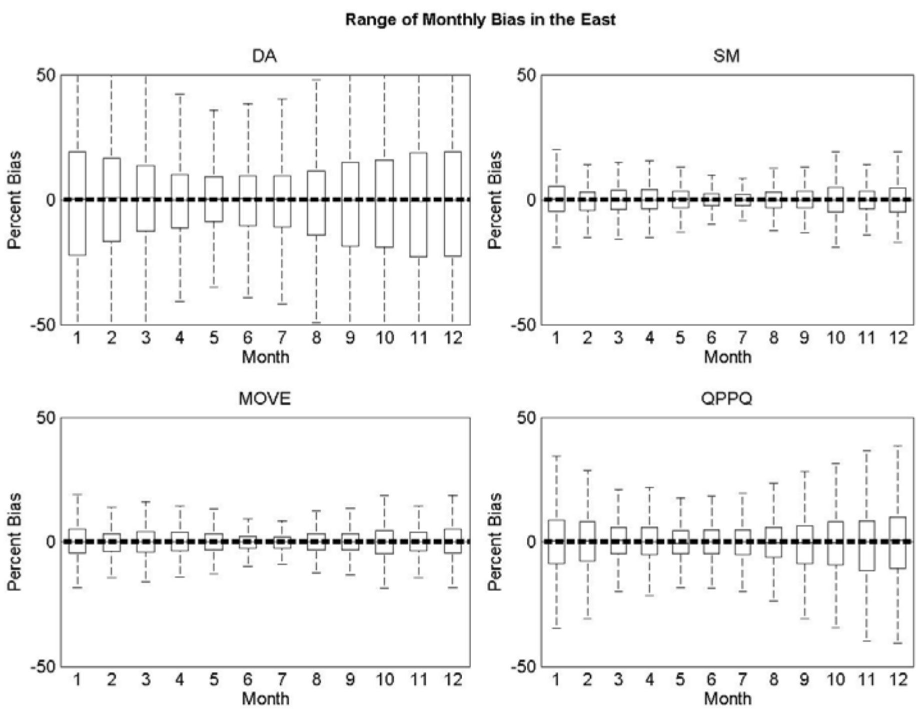

Figure 3. 40. Monthly range of bias for four flow-transfer techniques in the East.

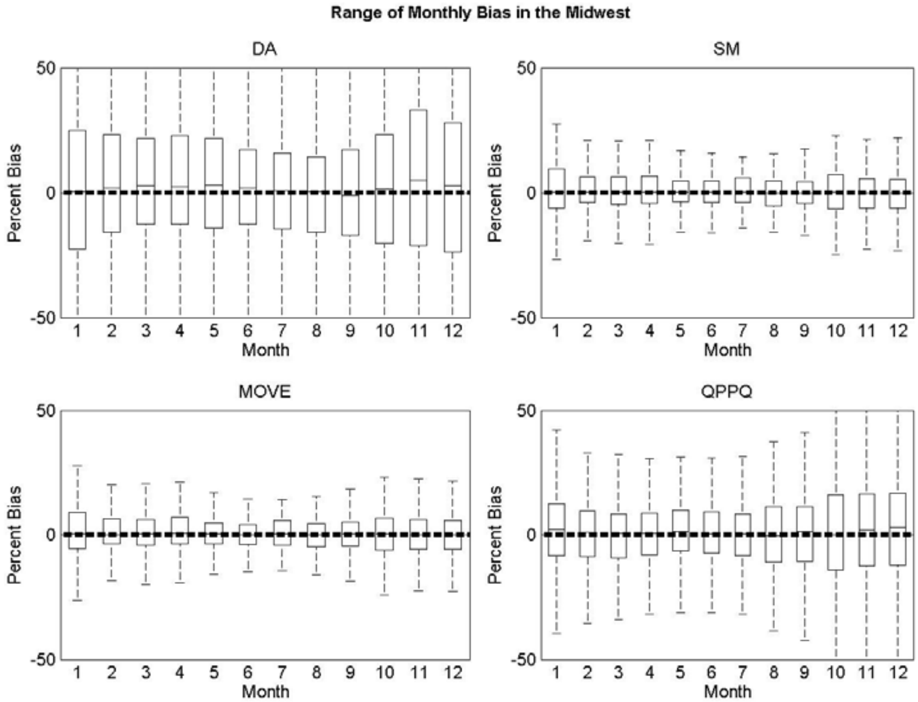

Figure 3. 41. Monthly range of bias for four flow-transfer techniques in the Midwest.

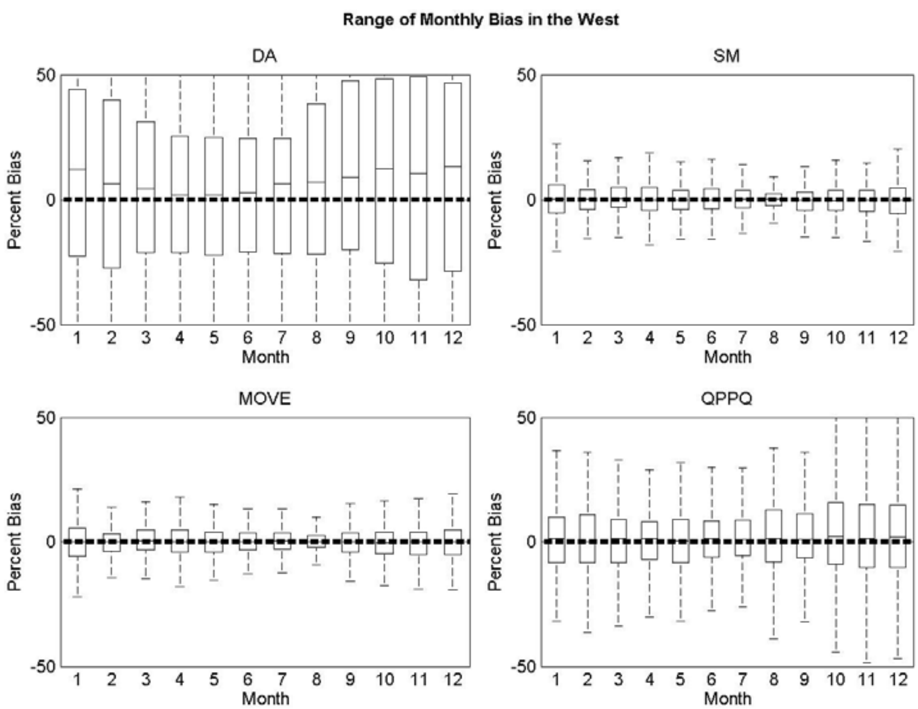

Figure 3. 42. Monthly range of bias for four flow-transfer techniques in the West.

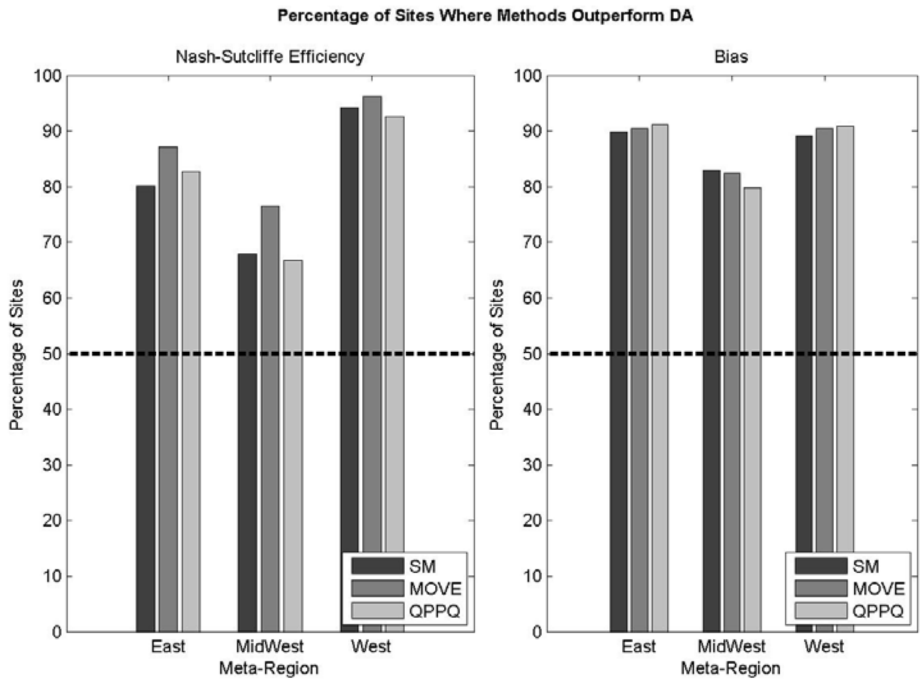

Figure 3. 43. Overall percentage of sites where each method outperforms DAR.

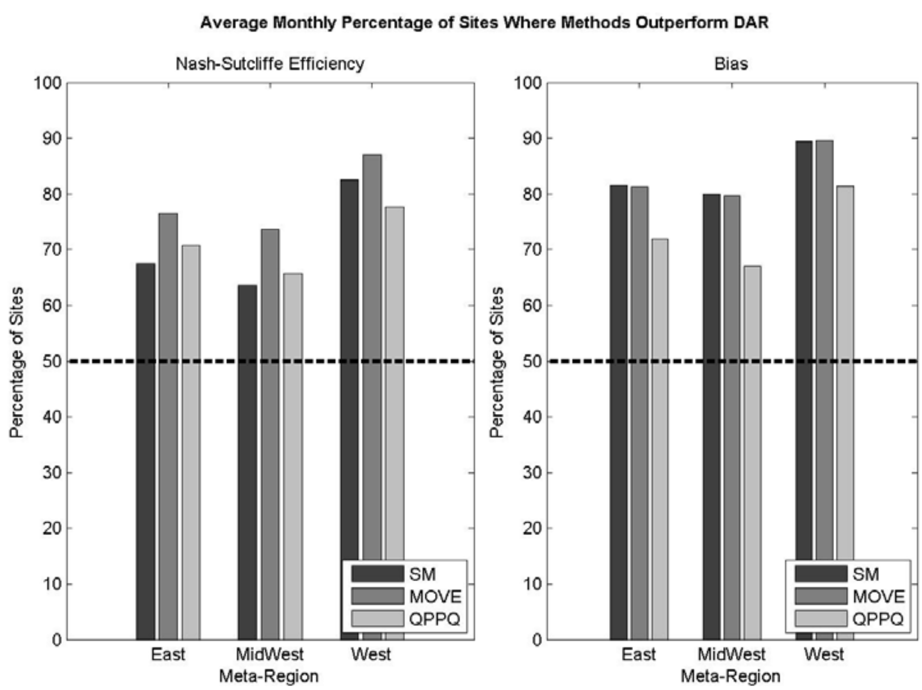

Figure 3. 44. Monthly average percentage of sites where each method outperforms DAR.

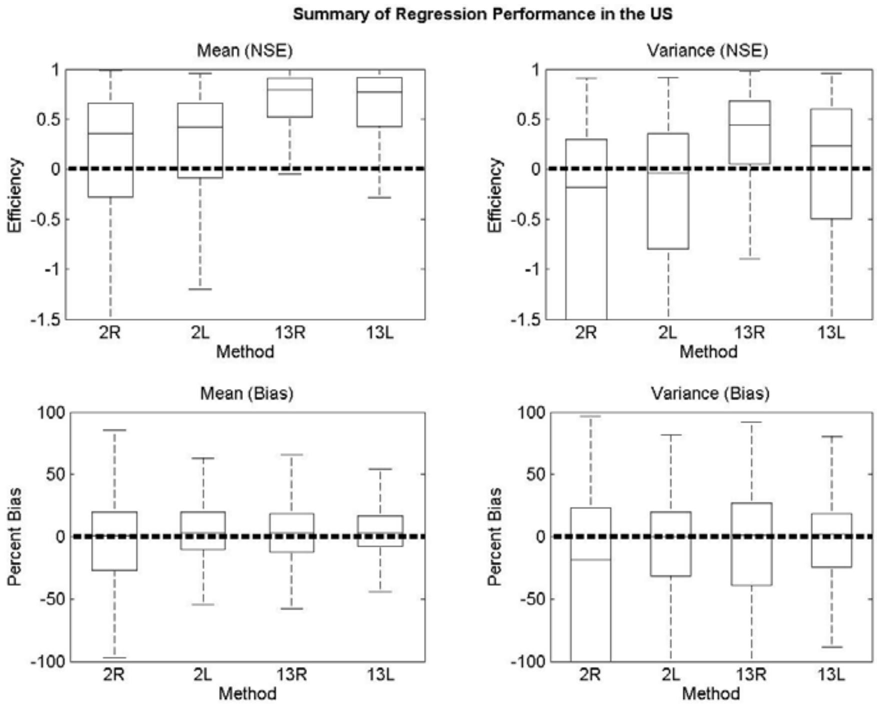

Figure 4. 1: Range of National Performance of Four Regression Methods.

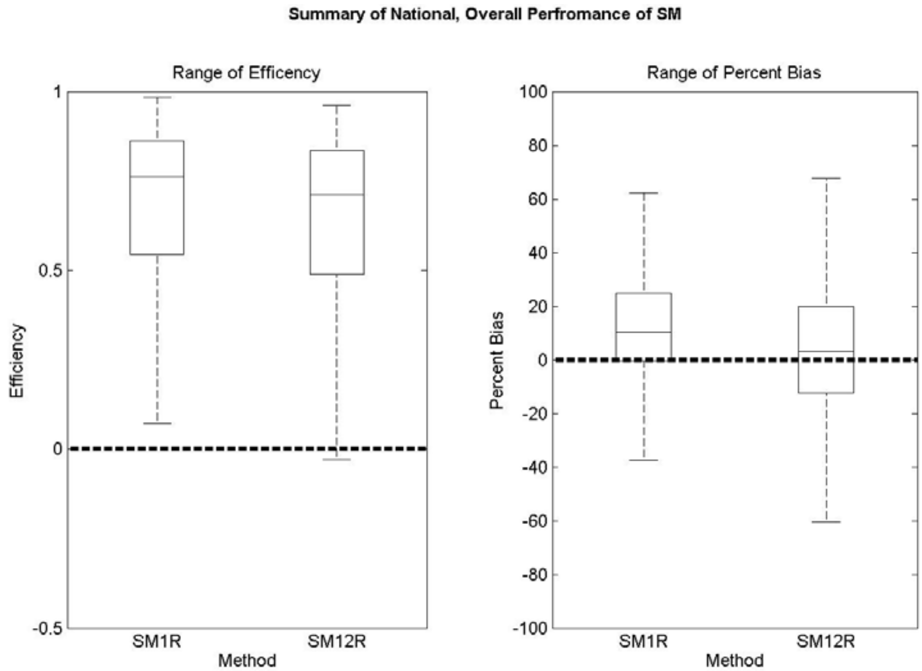

Figure 5. 1. National, overall performance of SM with regional regression.

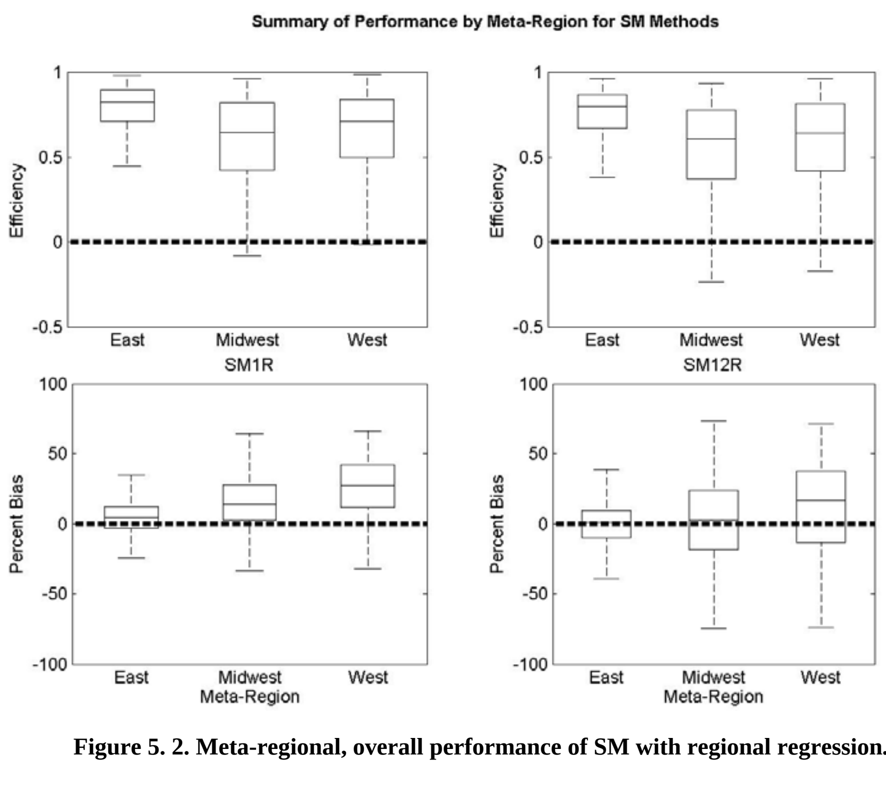

Figure 5. 2. Meta-regional, overall performance of SM with regional regression.

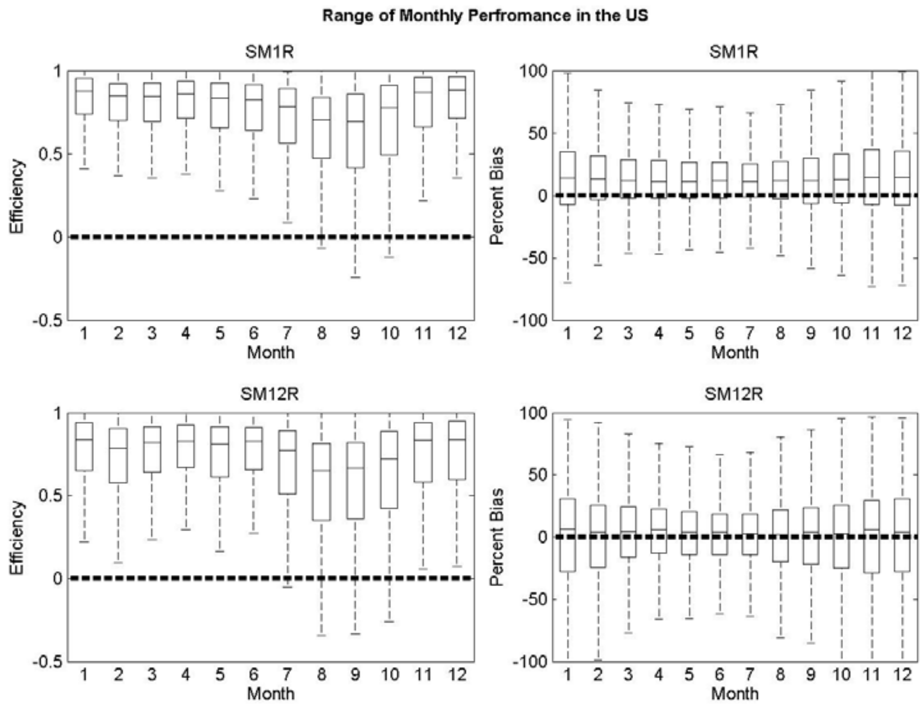

Figure 5. 3. National, monthly performance of SM methods with regional regression by meta-region.

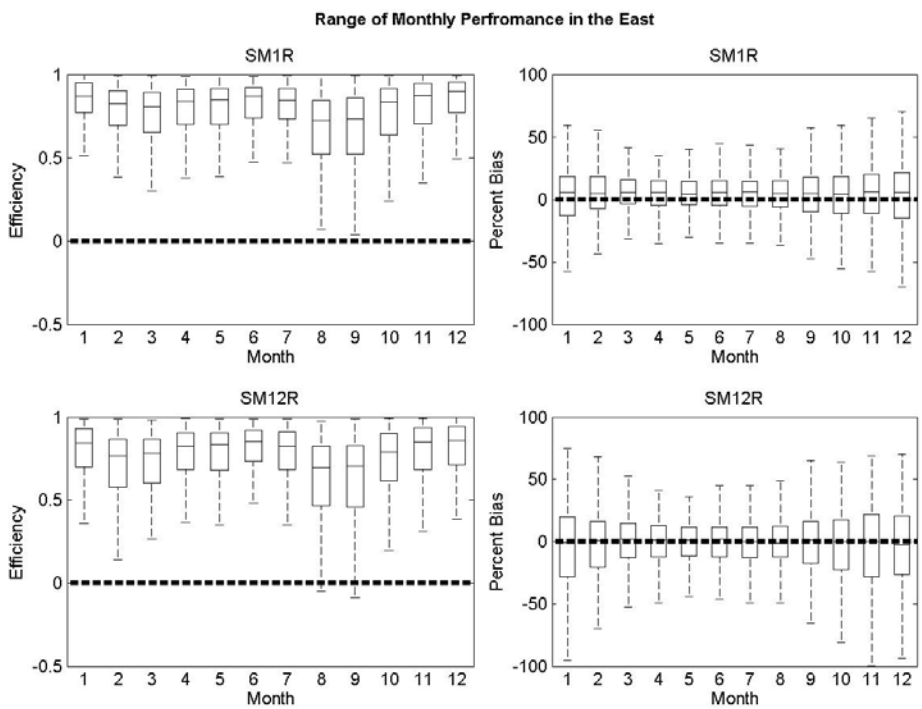

Figure 5. 4. Monthly performance of SM methods with regional regression in the East.

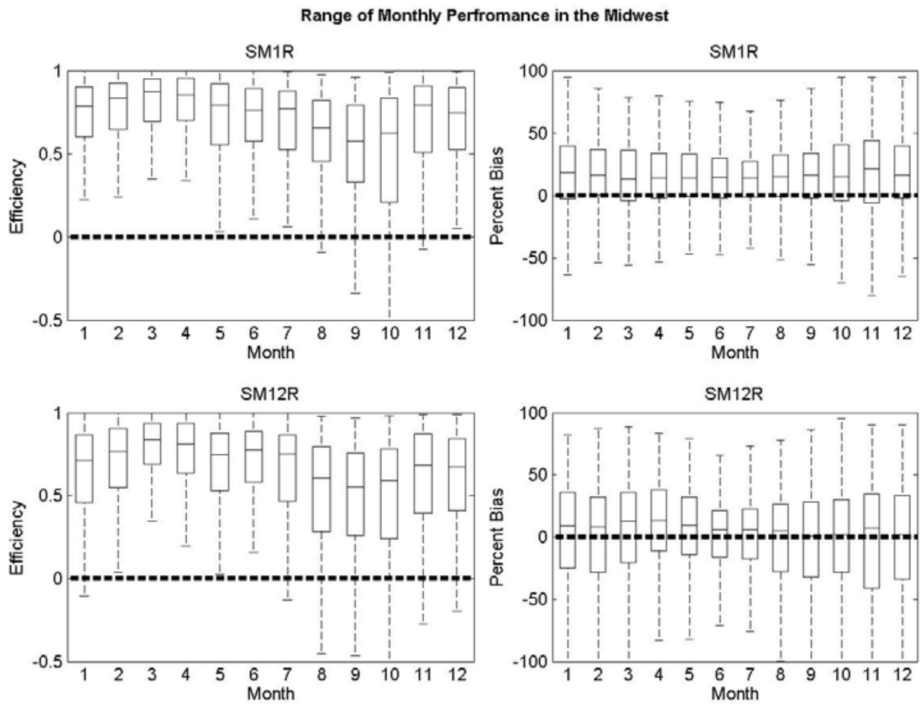

Figure 5. 5. Monthly performance of SM methods with regional regression in the Midwest.

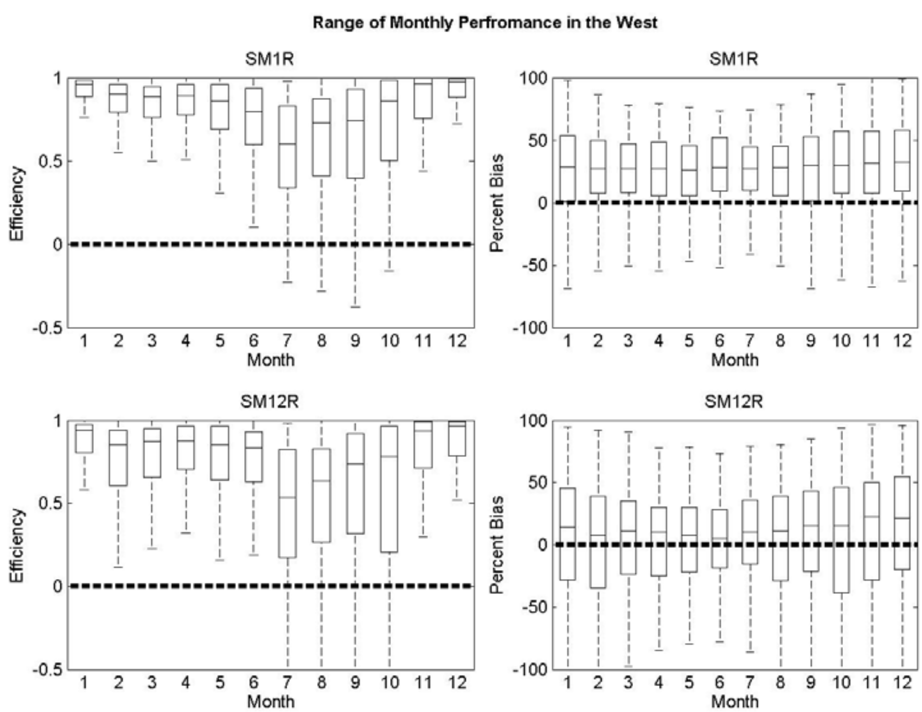

Figure 5. 6. Monthly performance of SM methods with regional regression in the West.

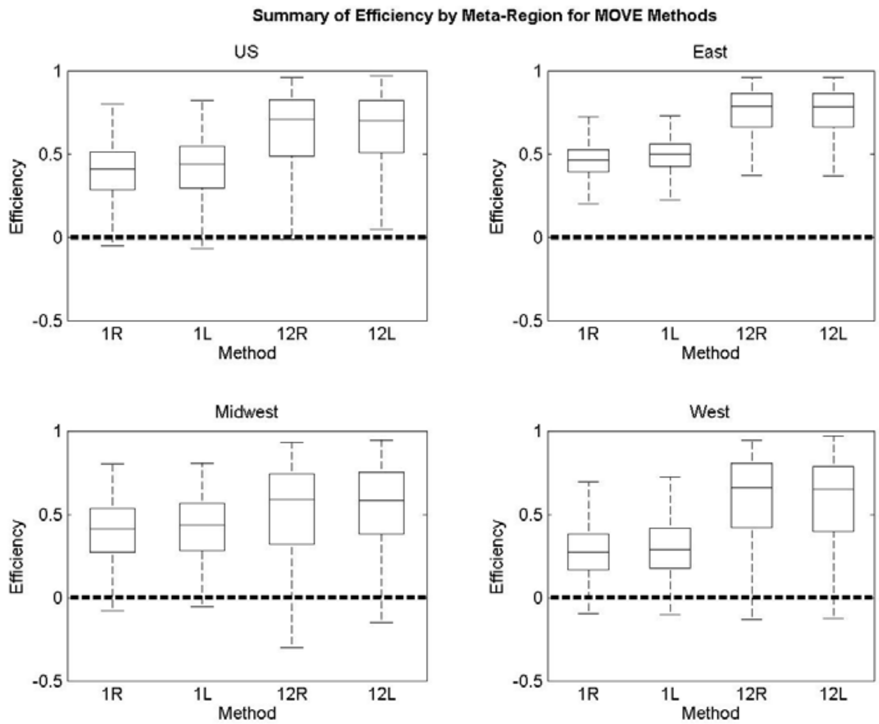

Figure 5. 7. Overall range of NSE for four variants of MOVE with regional regression.

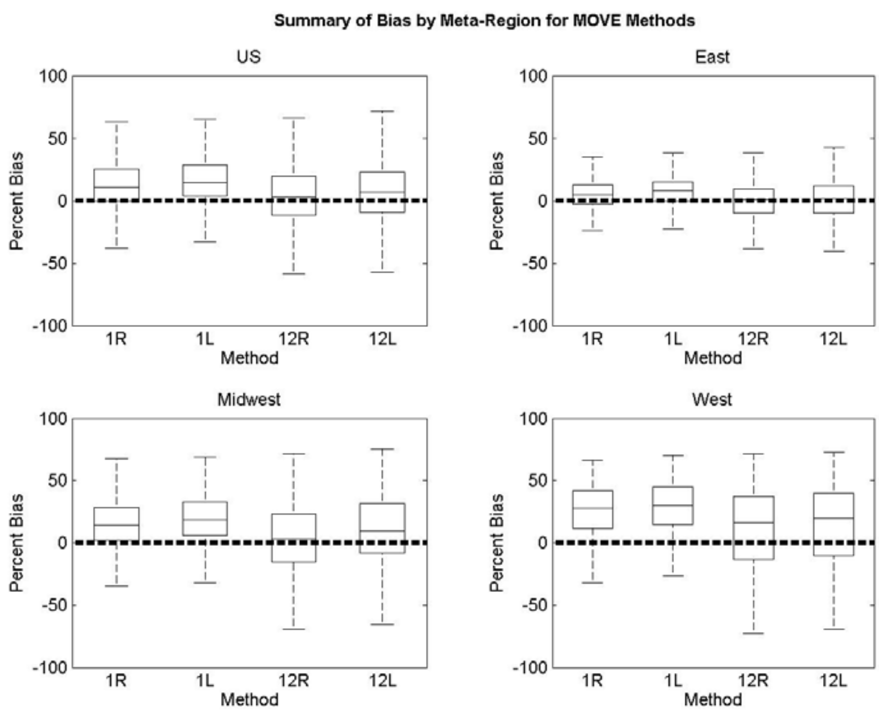

Figure 5. 8. Overall range of bias for four variants of MVOE with regional regression.

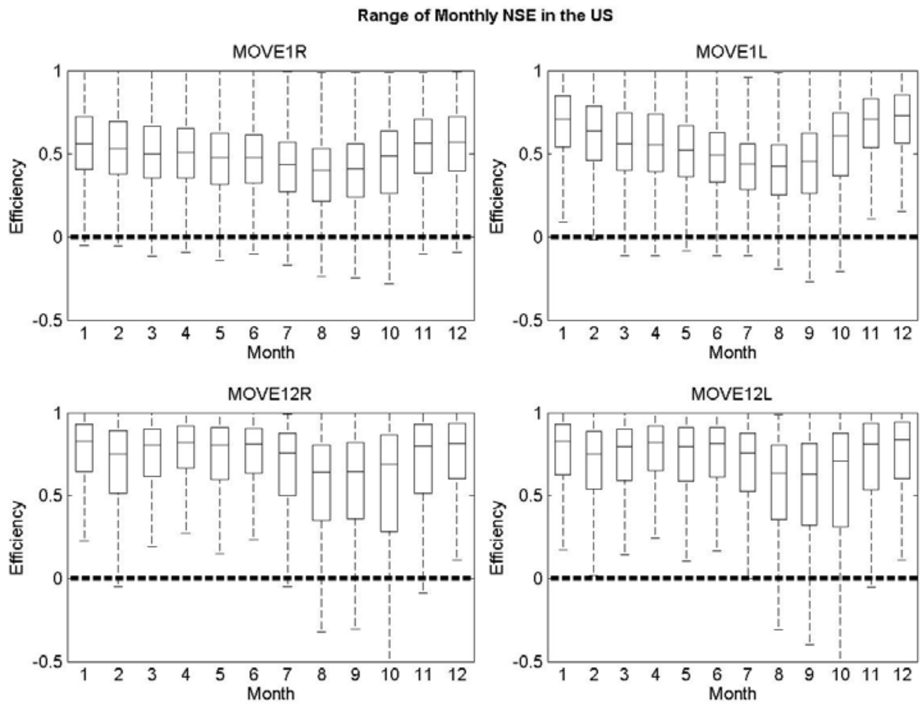

Figure 5. 9. Range of monthly NSE for four variants of MOVE with regional regression.

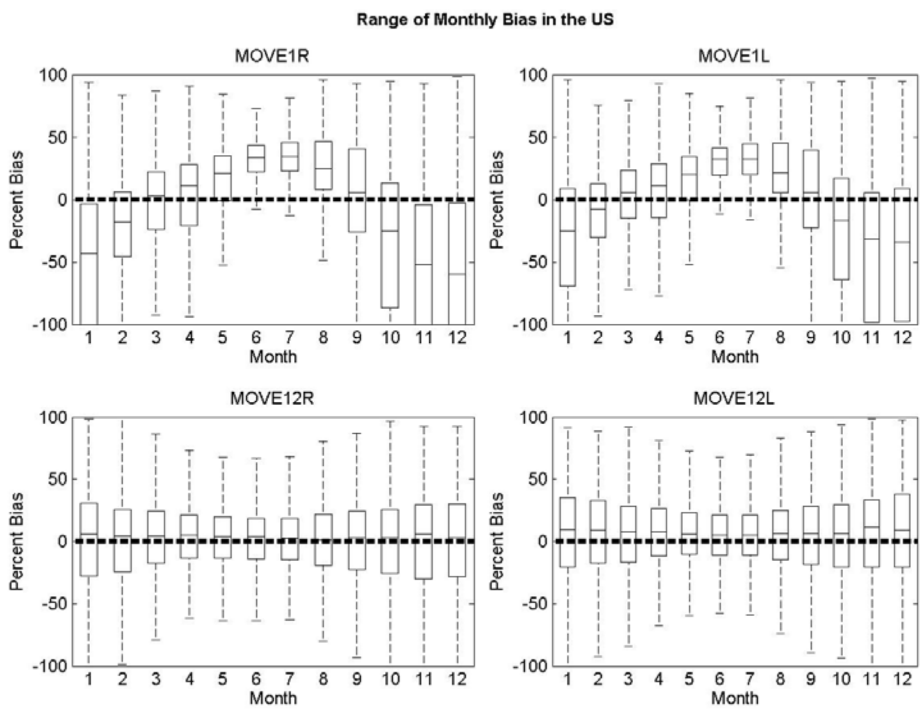

Figure 5. 10. Range of monthly bias for four variants of MOVE with regional regression.

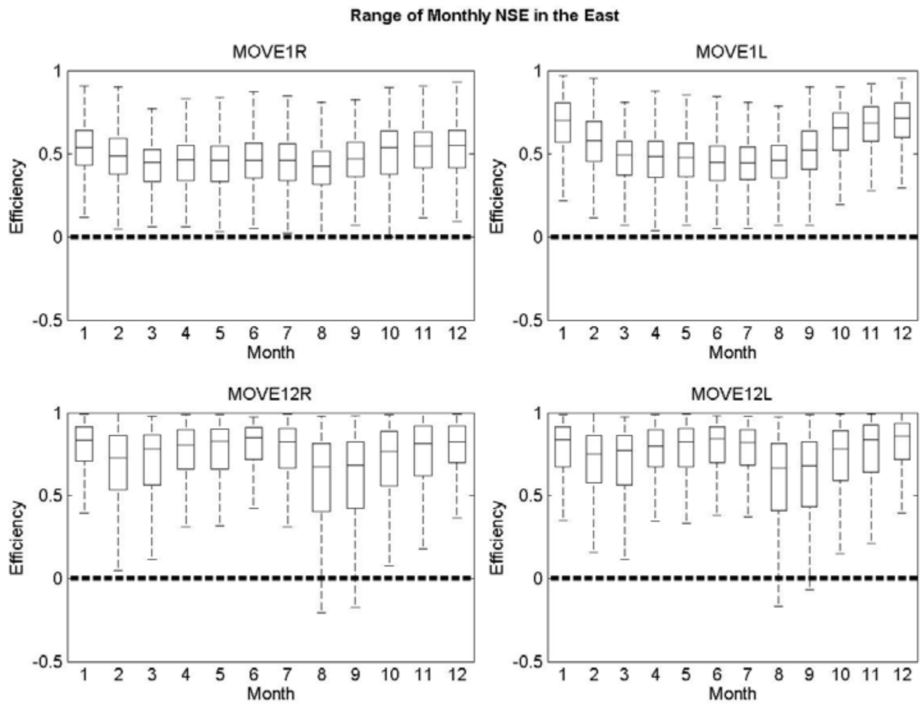

Figure 5. 11. Range of monthly NSE for four variants of MOVE with regional regression in the East.

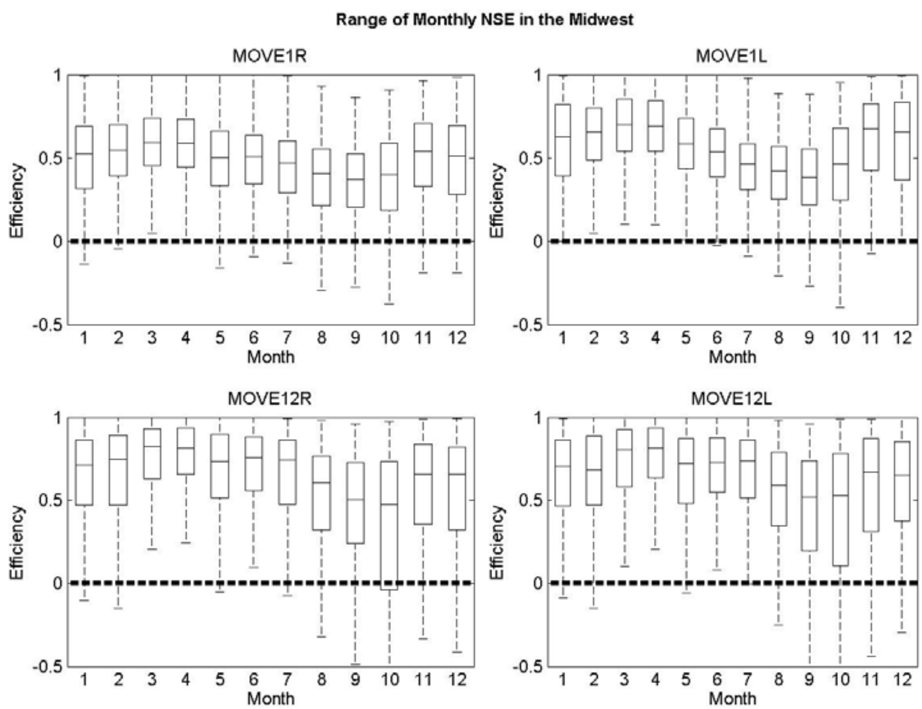

Figure 5. 12. Range of monthly NSE for four variants of MOVE with regional regression in the Midwest.

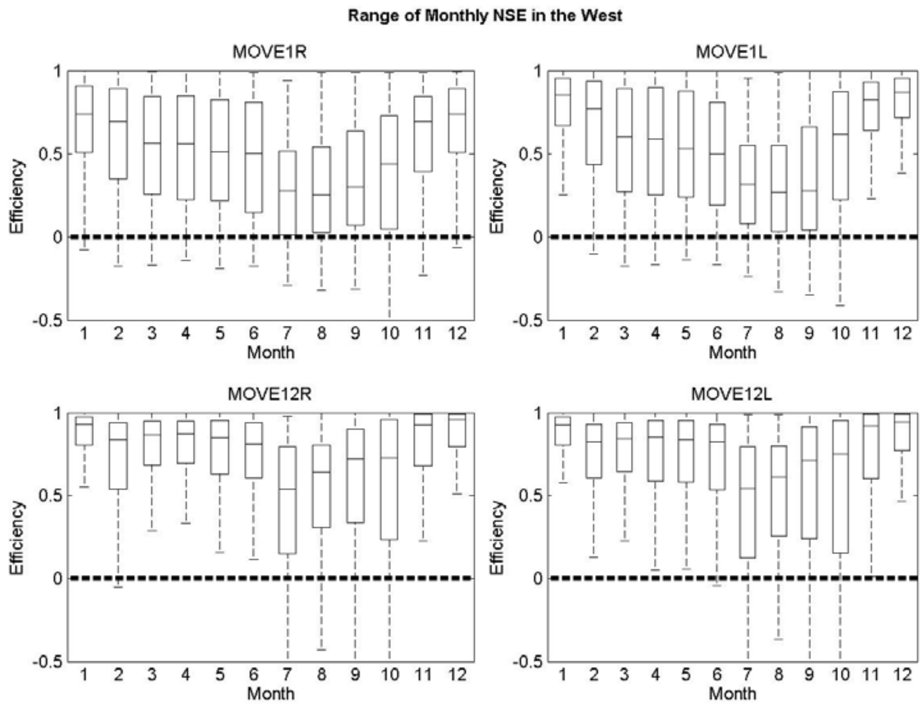

Figure 5. 13. Range of monthly NSE for four variants of MOVE with regional regression in the West.

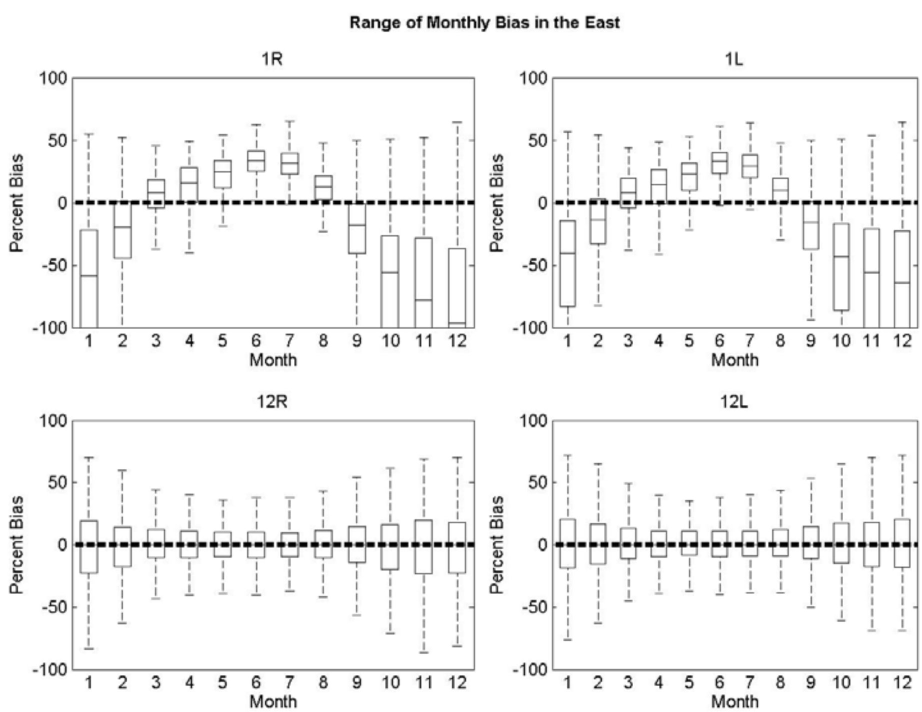

Figure 5. 14. Range of monthly bias for four variants of MOVE with regional regression in the East.

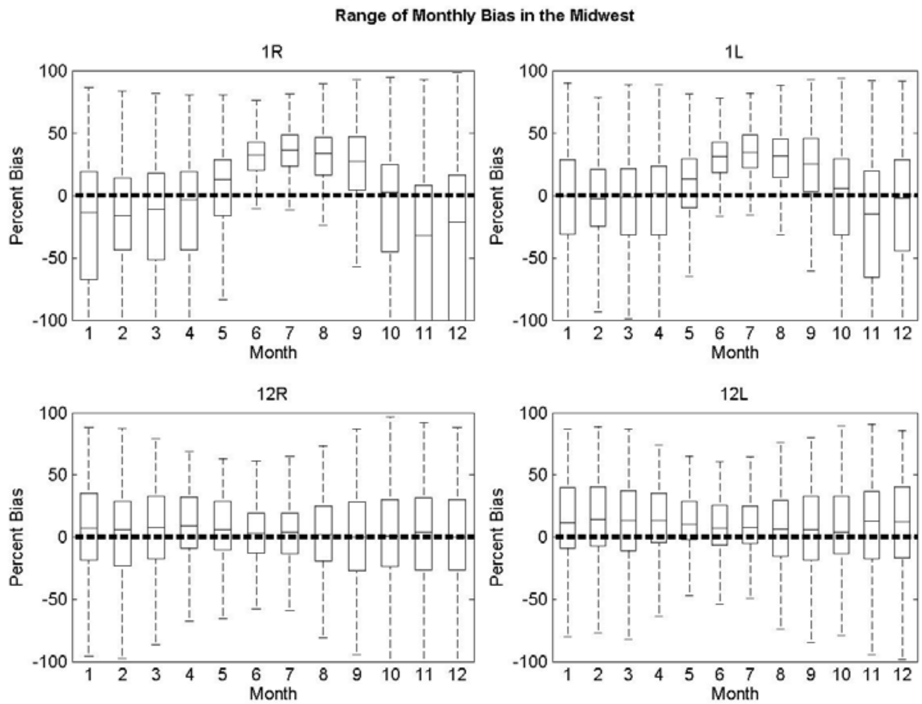

Figure 5. 15. Range of monthly bias for four variants of MOVE with regional regression in the Midwest.

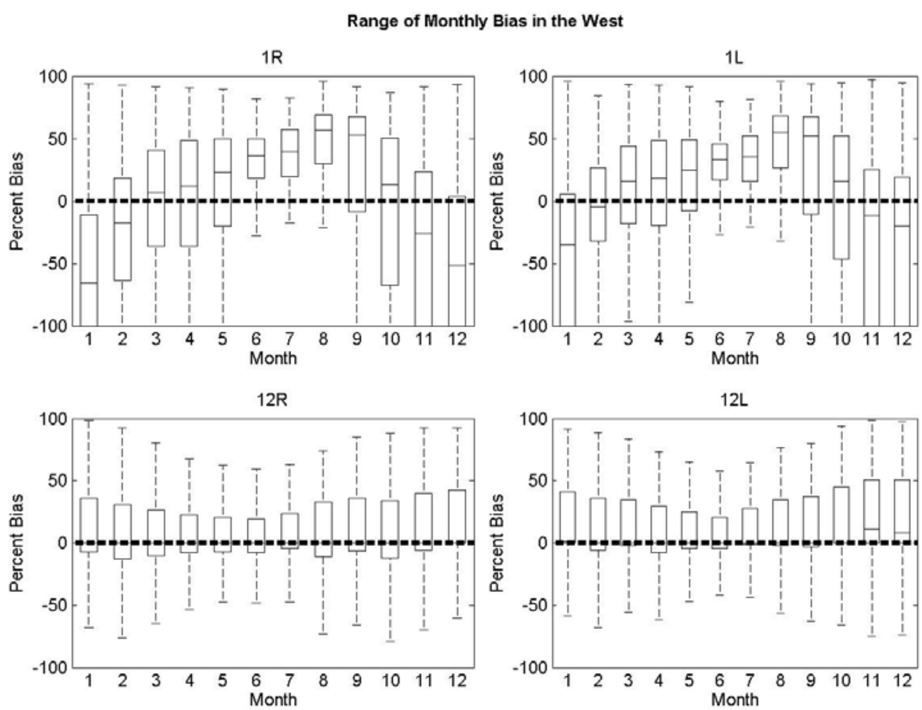

Figure 5. 16. Range of monthly bias for four variants of MOVE with regional regression in the West.

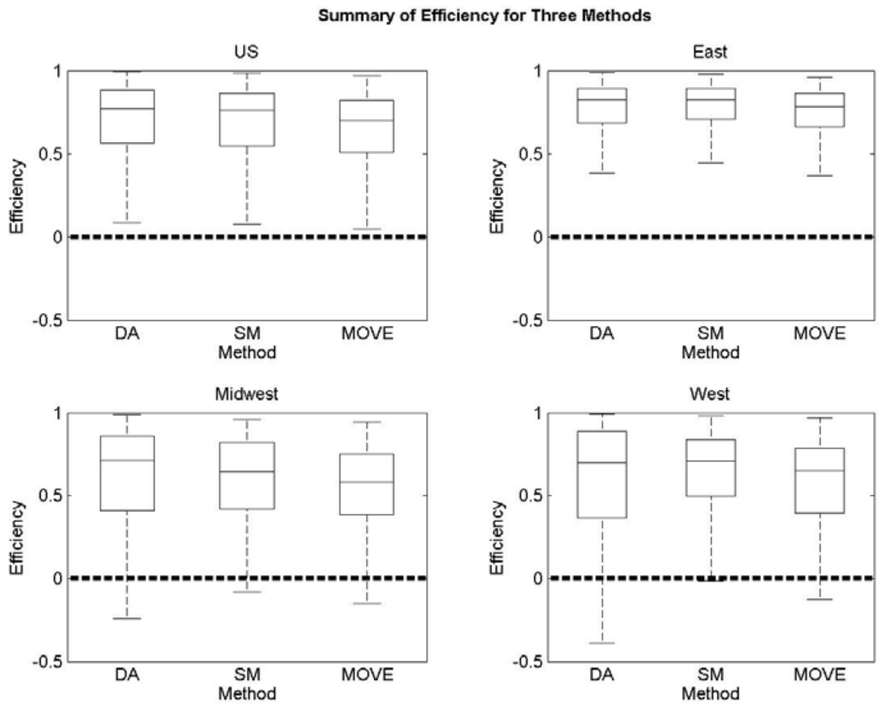

Figure 5. 17. Range of overall NSE for three flow-estimation techniques.

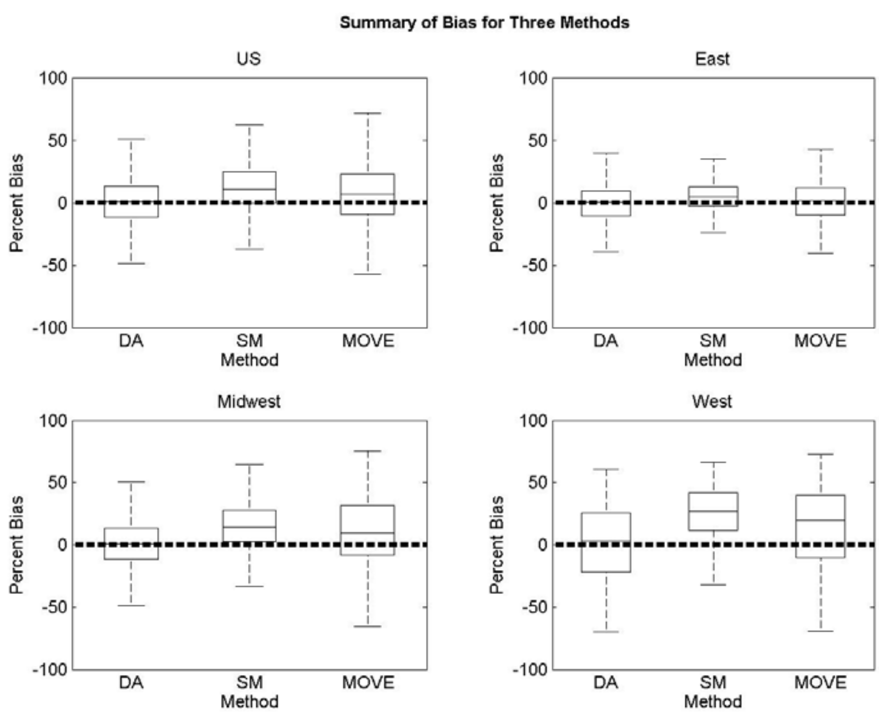

Figure 5. 18. Range of overall bias for three flow-estimation techniques.

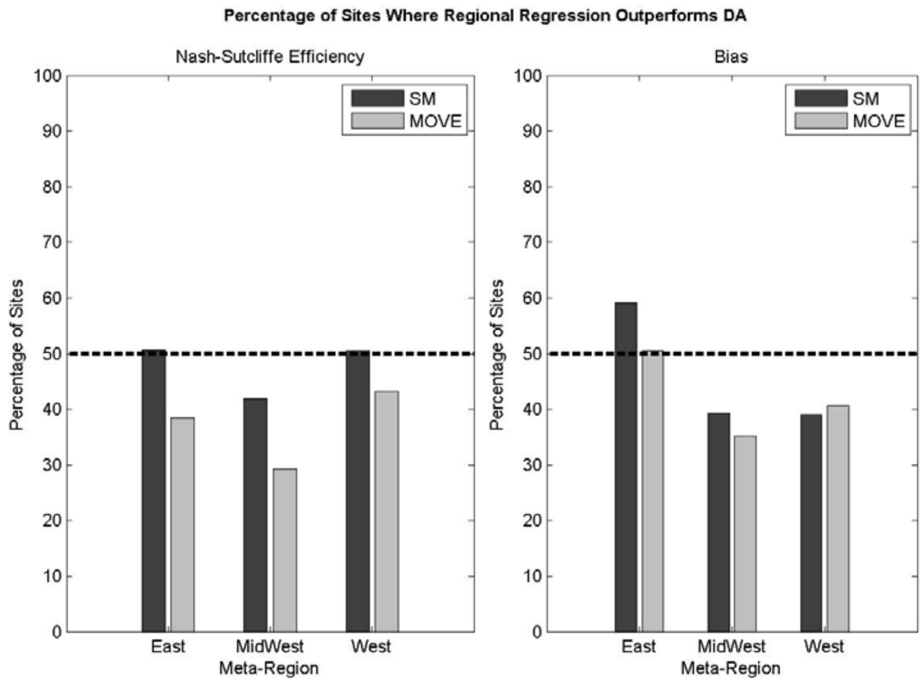

Figure 5. 19. Percentage of sites where each method outperforms DAR.

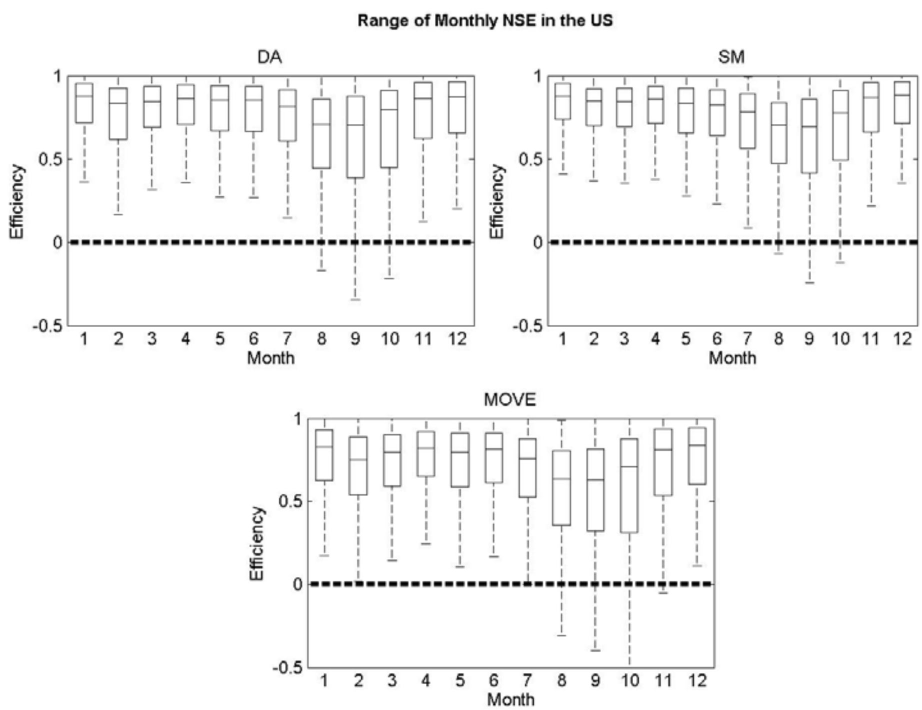

Figure 5. 20. Monthly range of NSE for three flow-estimation techniques.

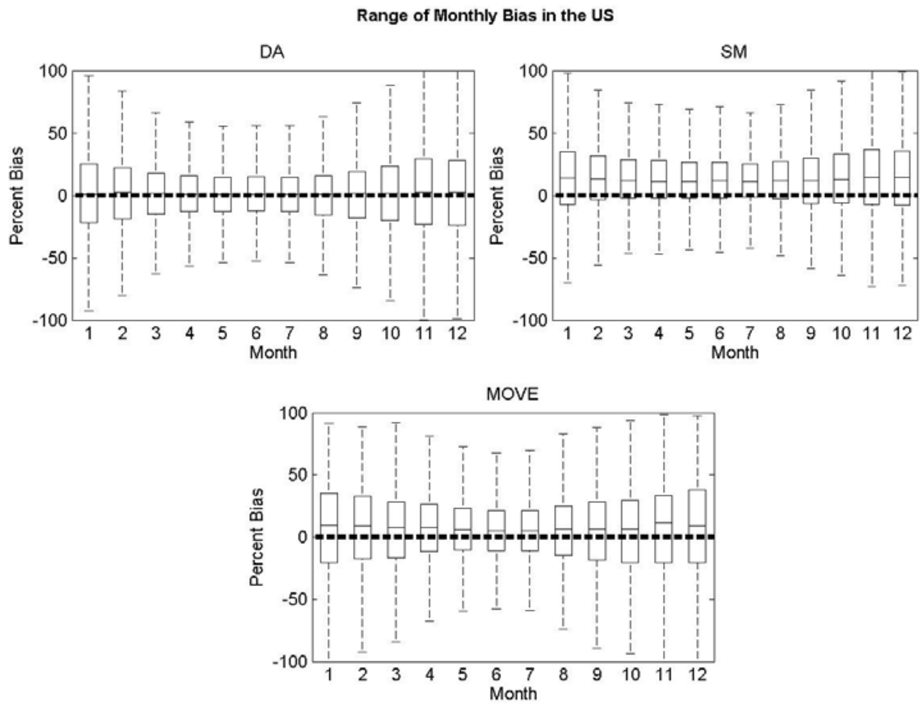

Figure 5. 21. Range of monthly bias of three flow-estimation techniques.

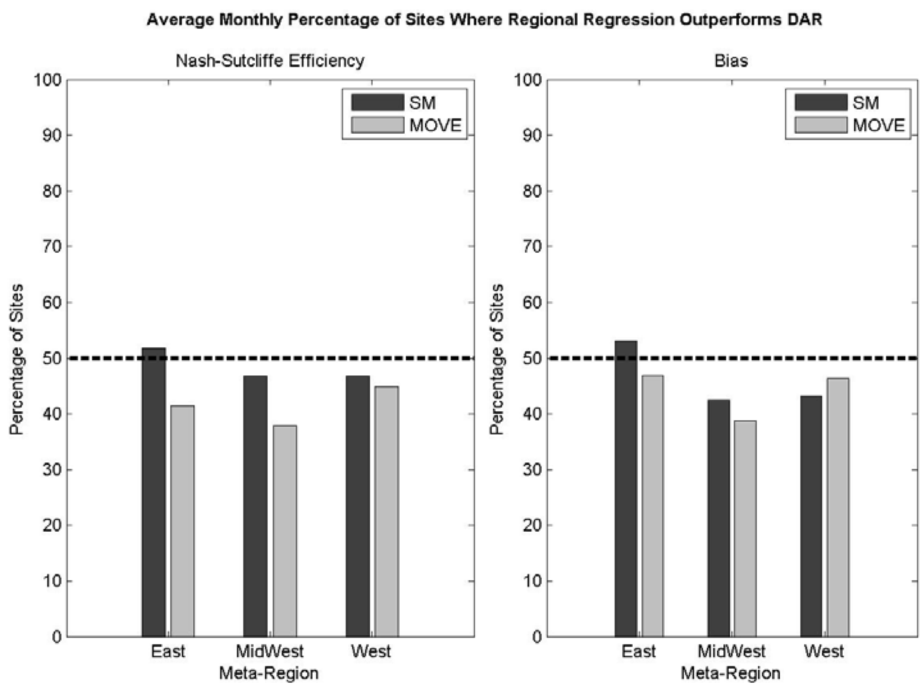

Figure 5. 22. Average percentage of sites where each method outperforms DAR.

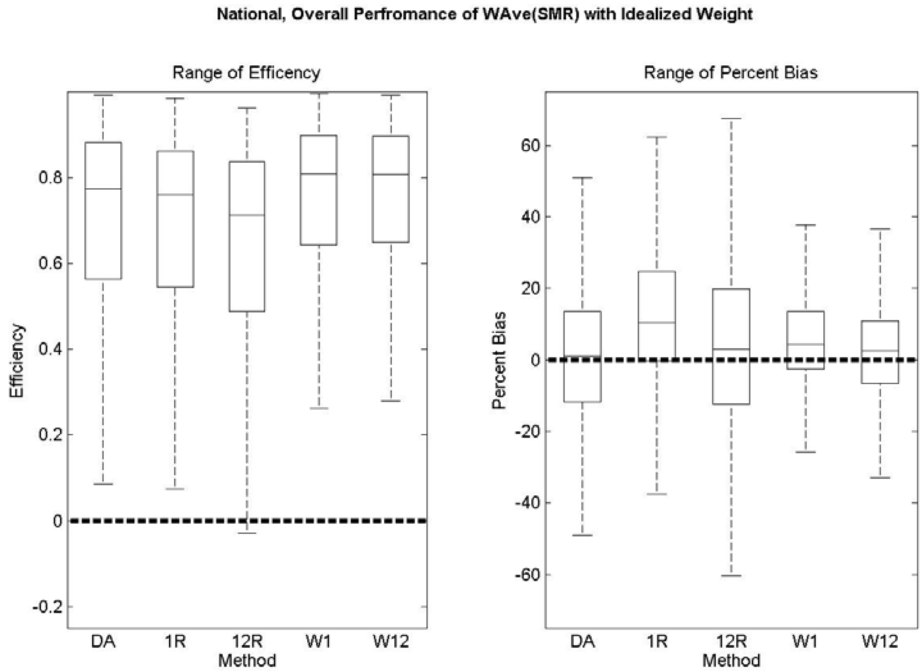

Figure 6. 1. National, overall performance of WAve(SM) methods with idealized weights.

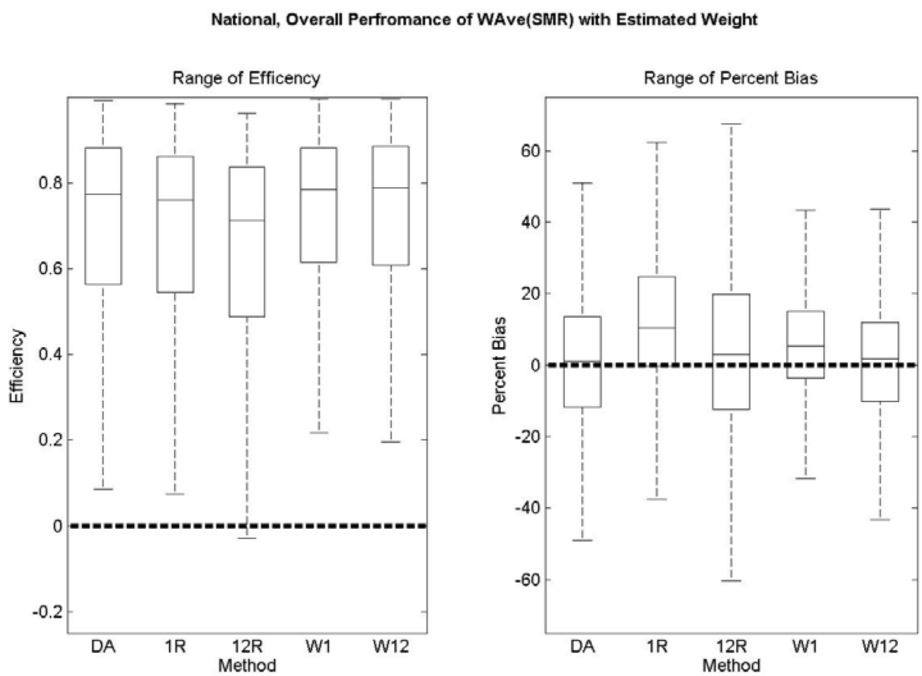

Figure 6. 2. National, overall performance of WAve(SM) methods with estimated weights.

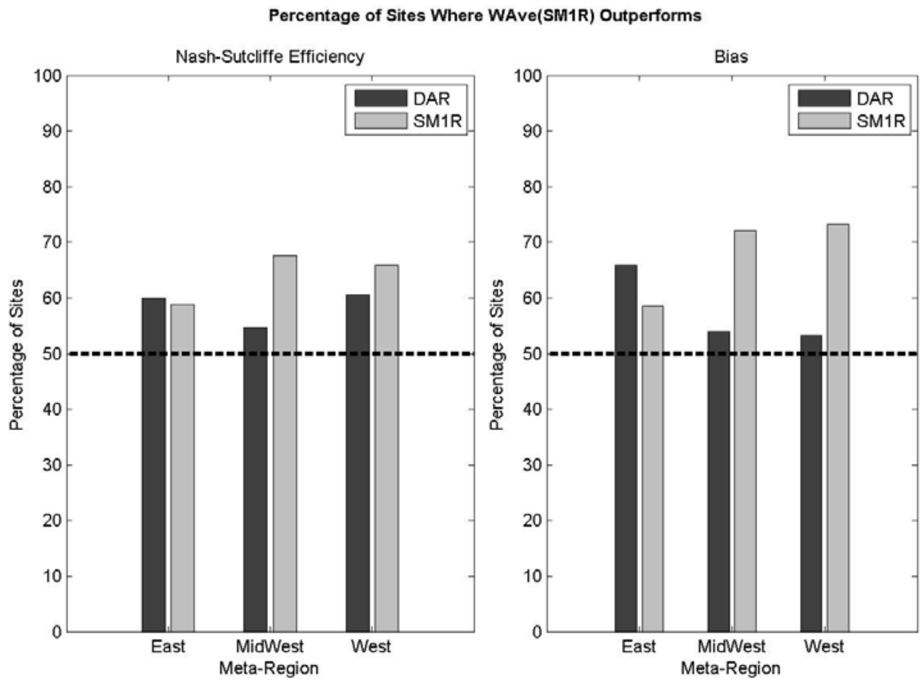

Figure 6. 3. Percentage of sites where WAve(SM1R) outperforms its component parts.

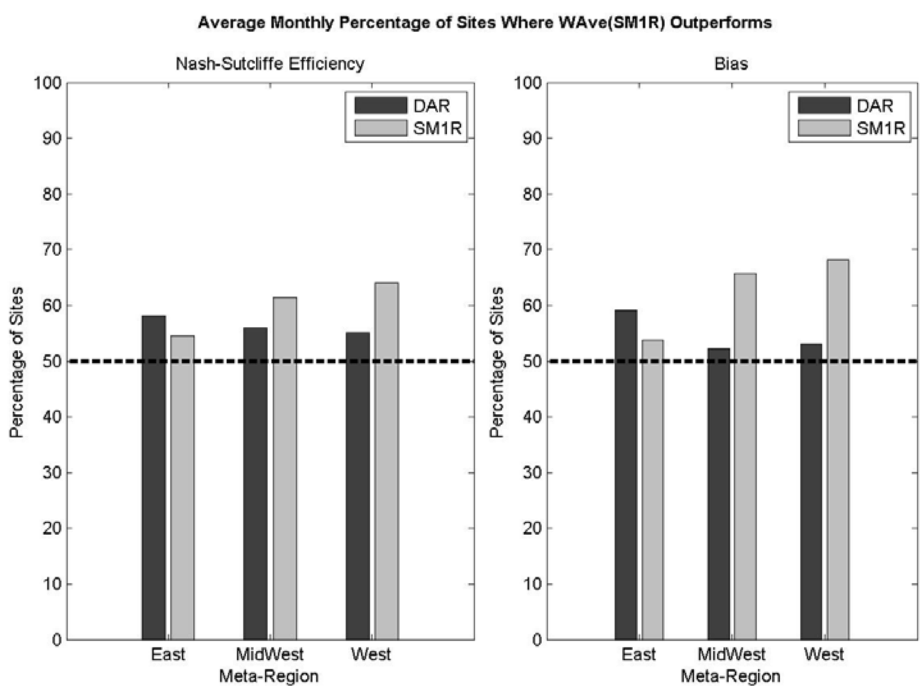

Figure 6. 4. Average monthly percentage of sites where WAve(SM1R) outperforms its component parts.

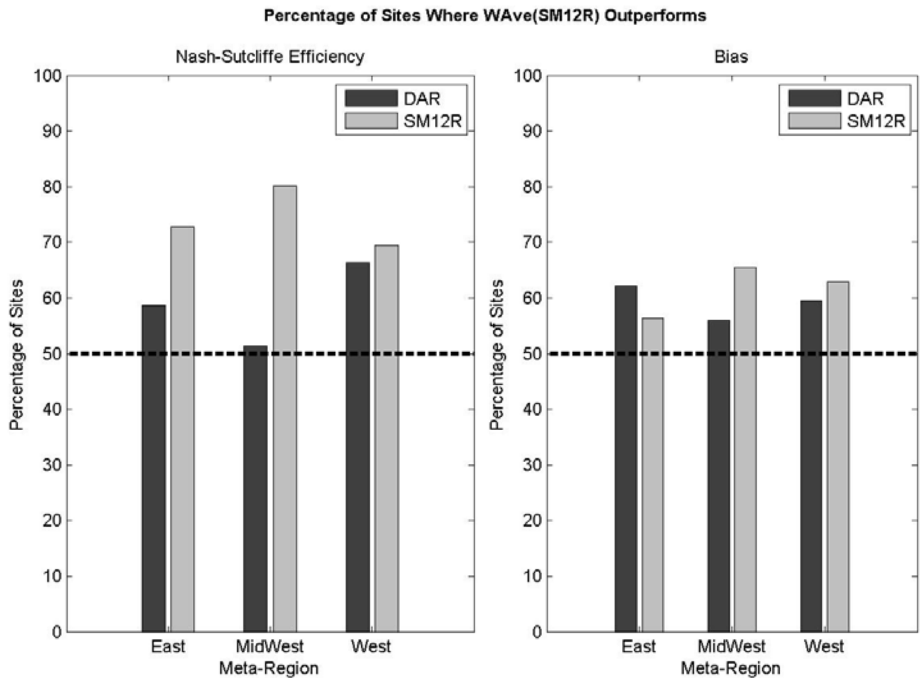

Figure 6. 5. Overall percentage of sites where WAve(SM12R) outperforms its component parts.

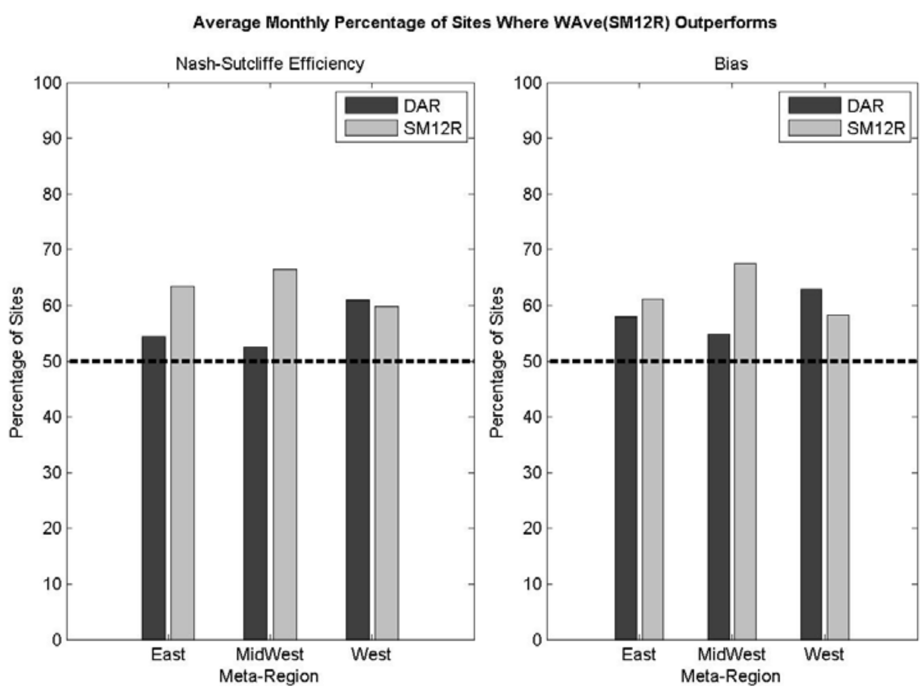

Figure 6. 6. Average monthly percentage of sites where WAve(SM12R) outperforms its component parts.

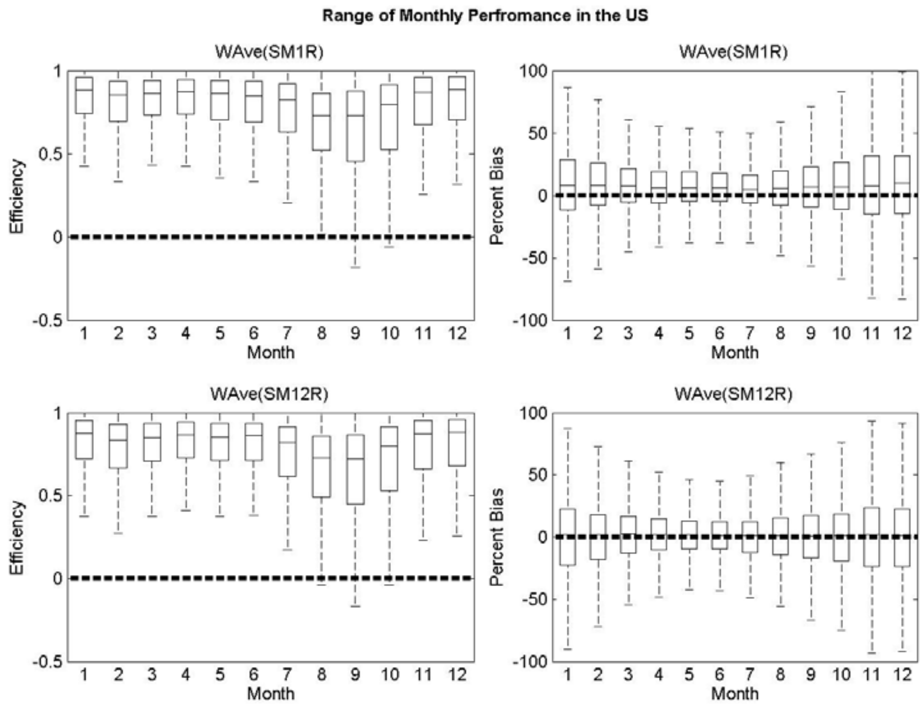

Figure 6. 7. Range of monthly performance for WAve(SM) techniques with estimated weights.

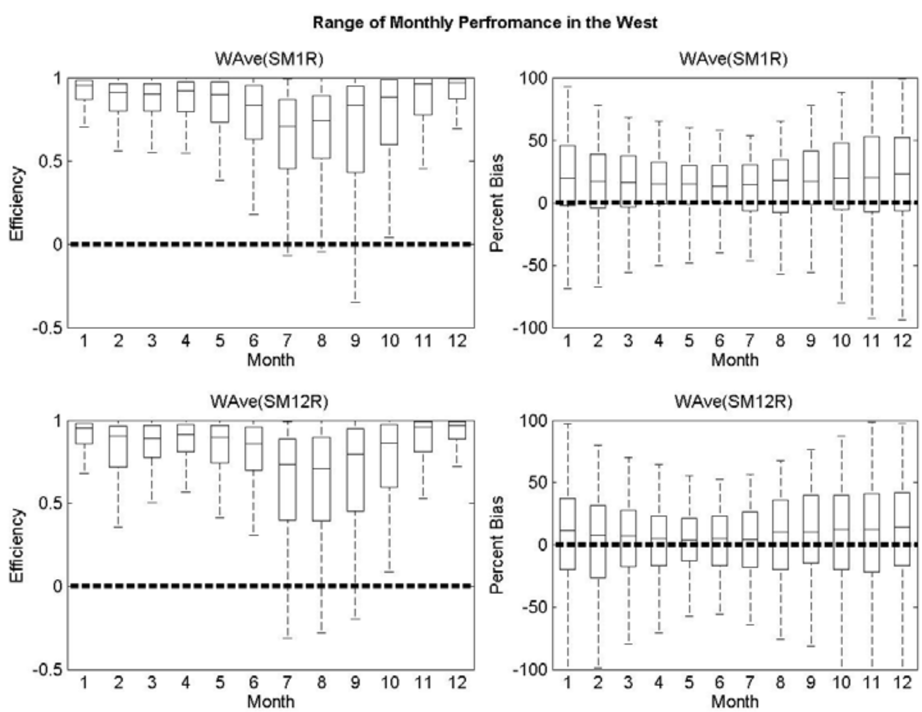

Figure 6. 8. Range of monthly performance for WAve(SM) techniques in the West.

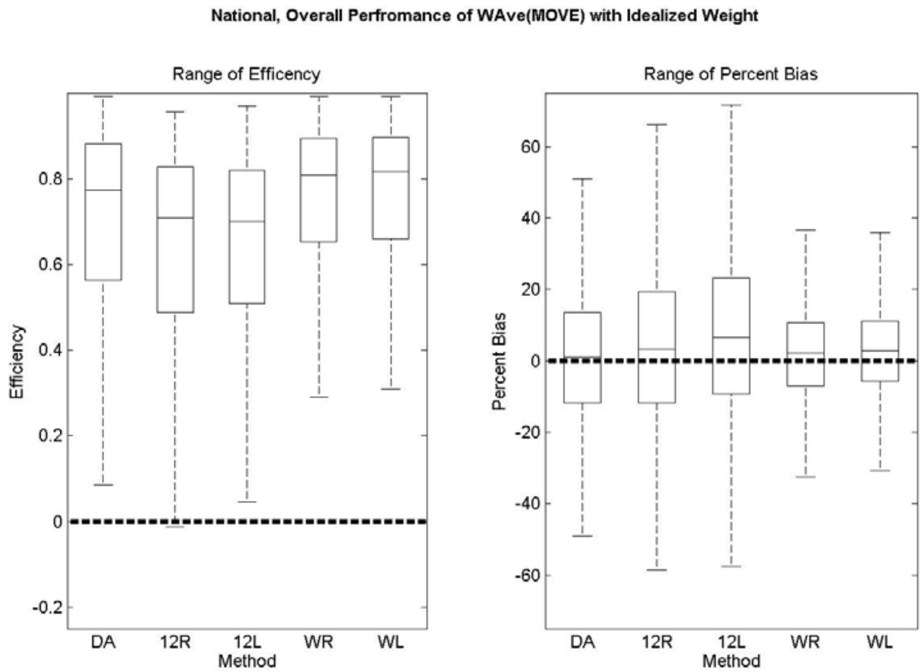

Figure 6. 9. National, overall performance of WAve(MOVE) methods with idealized weights.

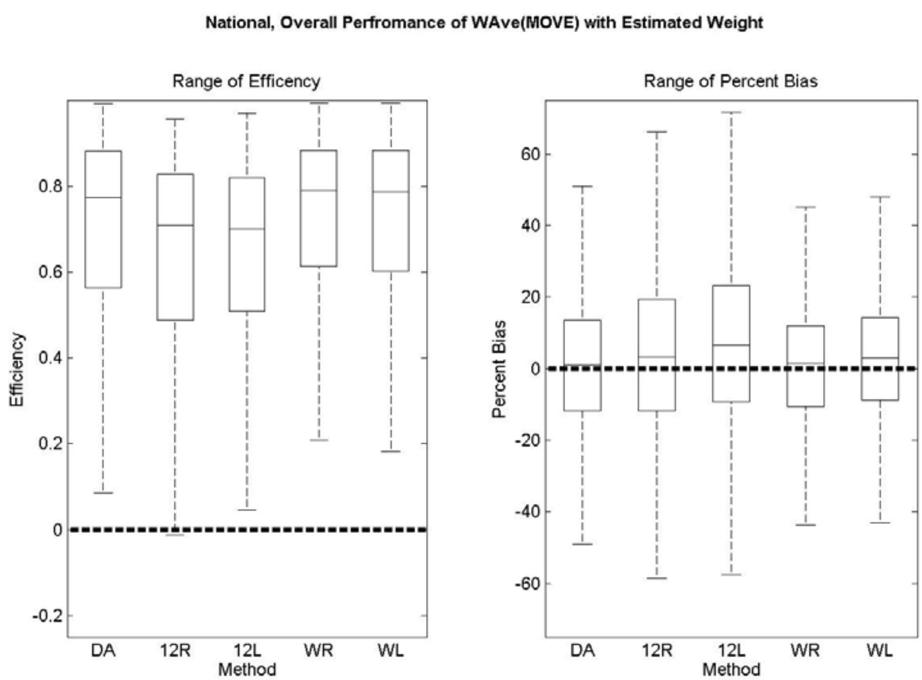

Figure 6. 10. National, overall performance of WAve(MOVE) with estimated weights.

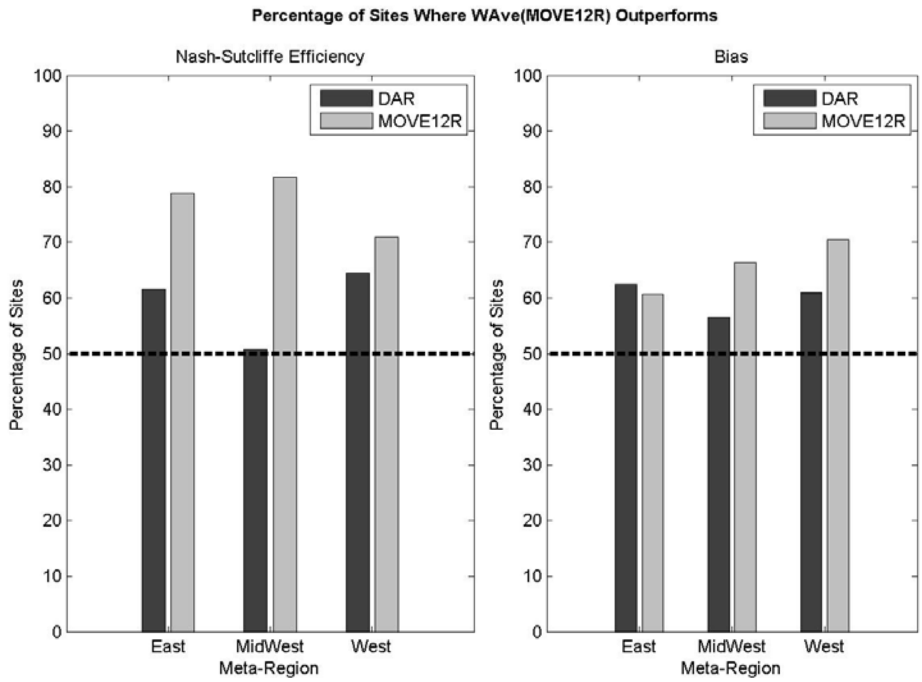

Figure 6. 11. Overall percentage of sites where WAve(MOVE12R) outperforms its component parts.

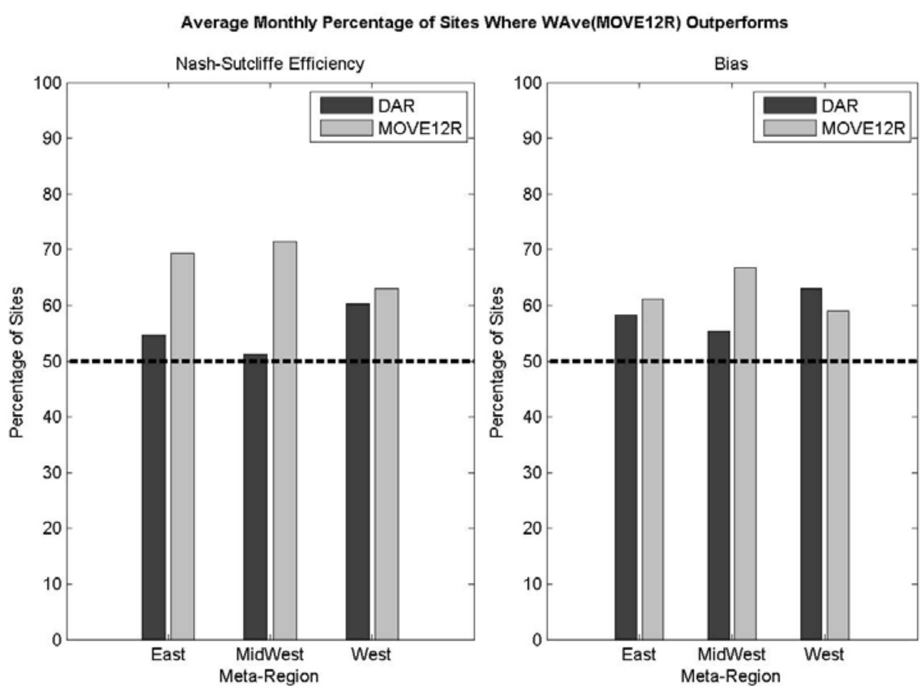

Figure 6. 12. Average monthly percentage of sites where WAve(MOVE12R) outperforms its component parts.

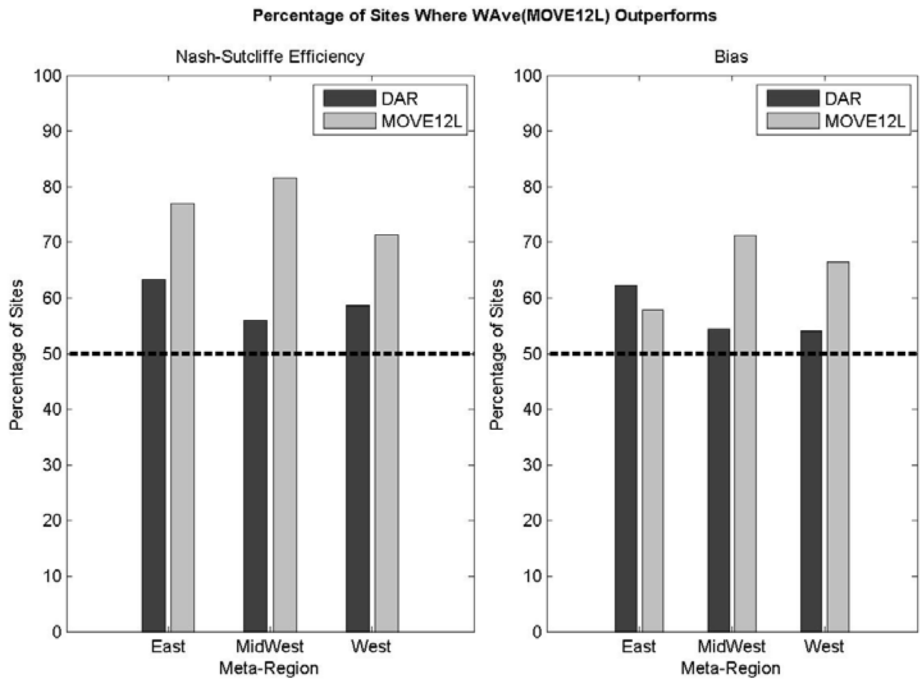

Figure 6. 13. Overall percentage of sites where WAve(MOVE12L) outperforms its component parts.

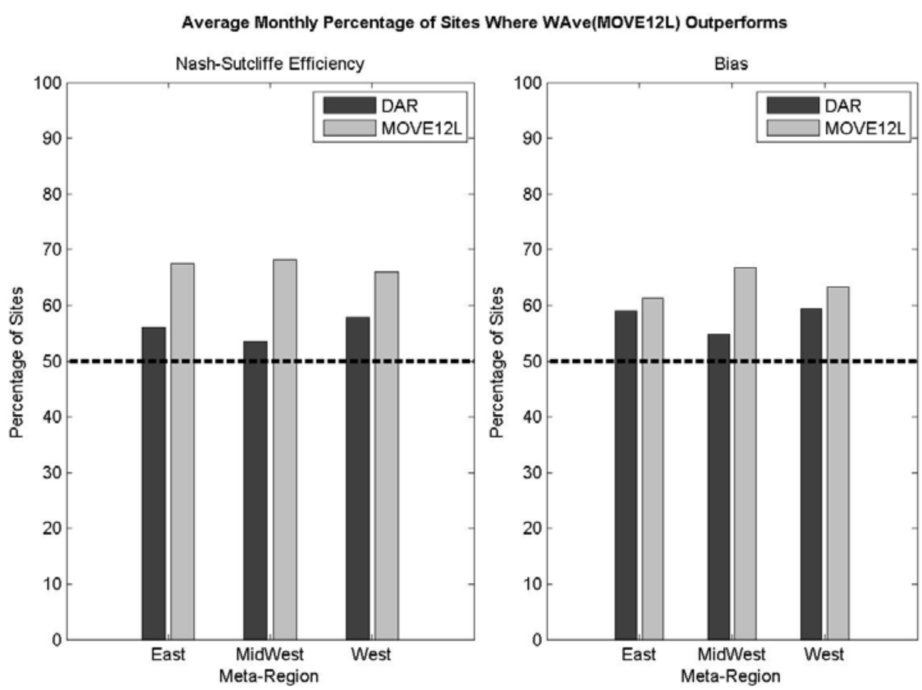

Figure 6. 14. Average monthly percentage of sites where WAve(MOVE12L) outperforms its component parts.

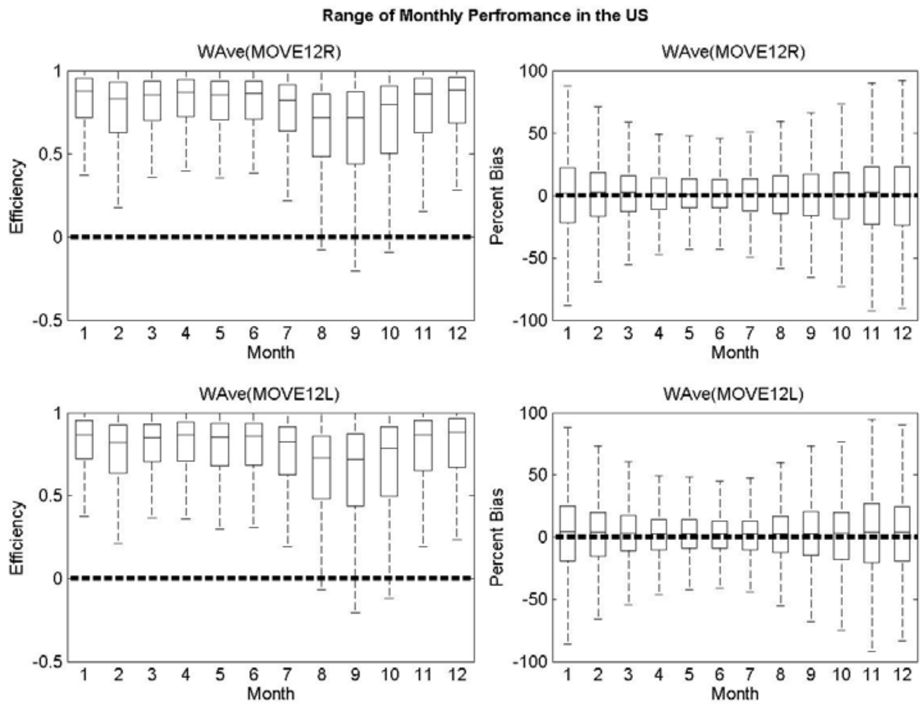

Figure 6. 15. Range of monthly performance for WAve(MOVE) methods with estimated weights.

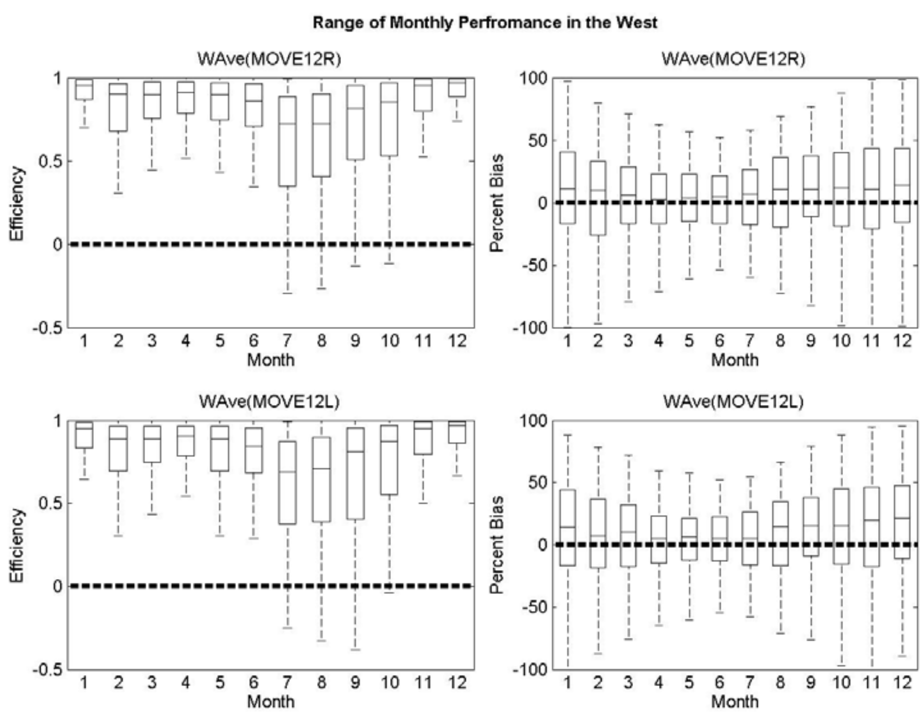

Figure 6. 16. Range of monthly performance for WAve(MOVE) methods in the West.

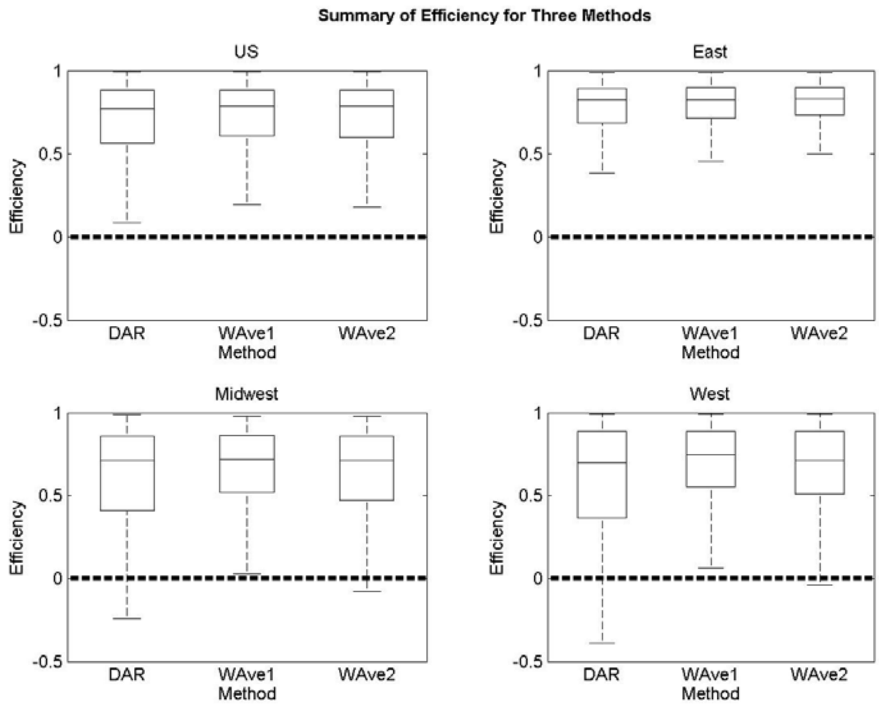

Figure 6. 17. Overall range of NSE for DAR and WAve techniques.

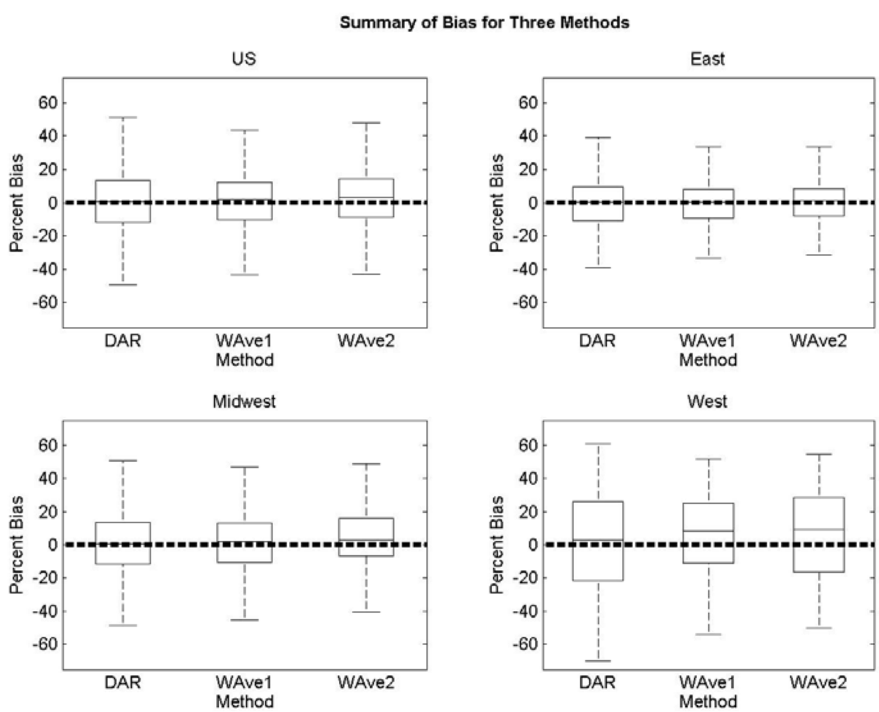

Figure 6. 18. Overall range of bias for DAR and WAve techniques.

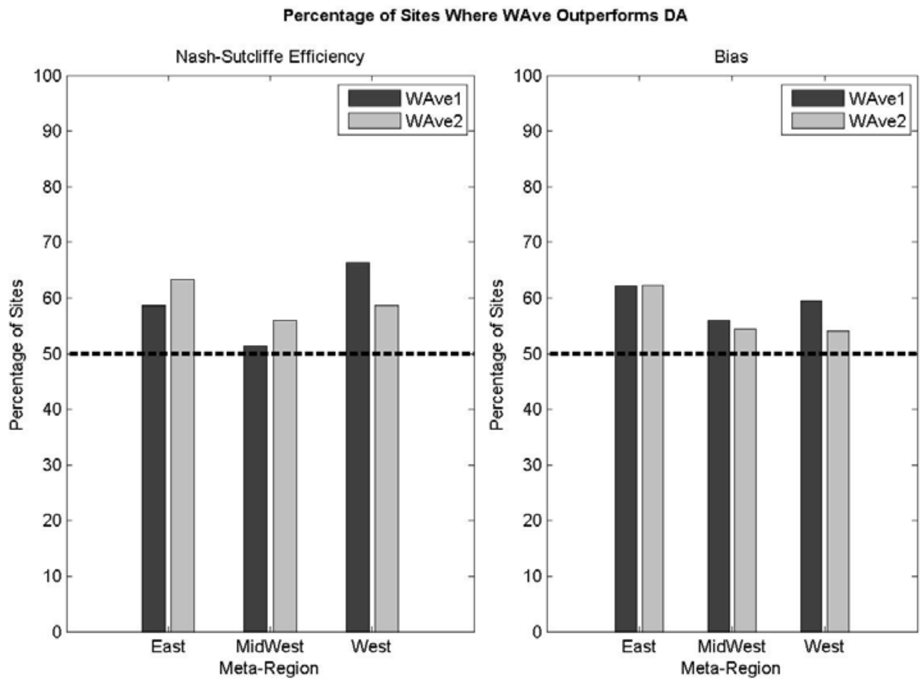

Figure 6. 19. Overall percentage of sites where WAve methods outperform DAR.

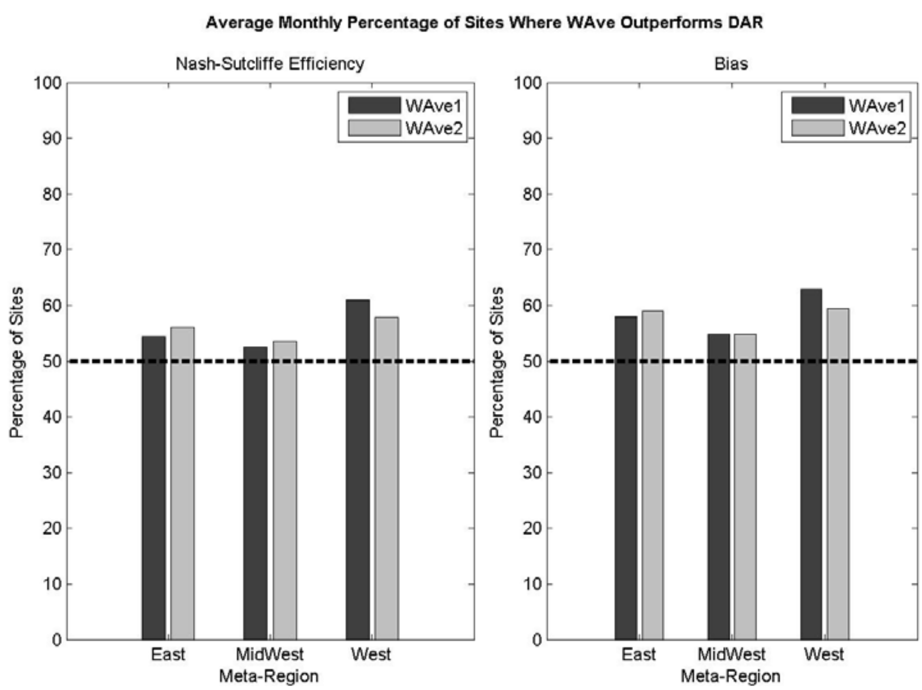

Figure 6. 20. Average monthly percentage of sites where WAve methods outperform DAR.

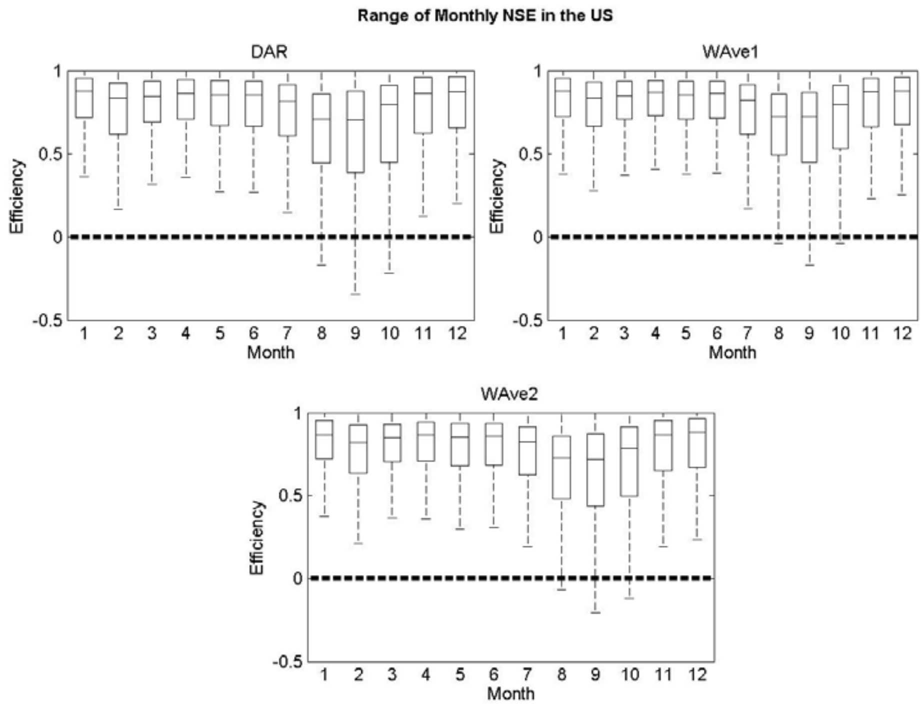

Figure 6. 21. Monthly range of NSE for DAR and WAve methods.

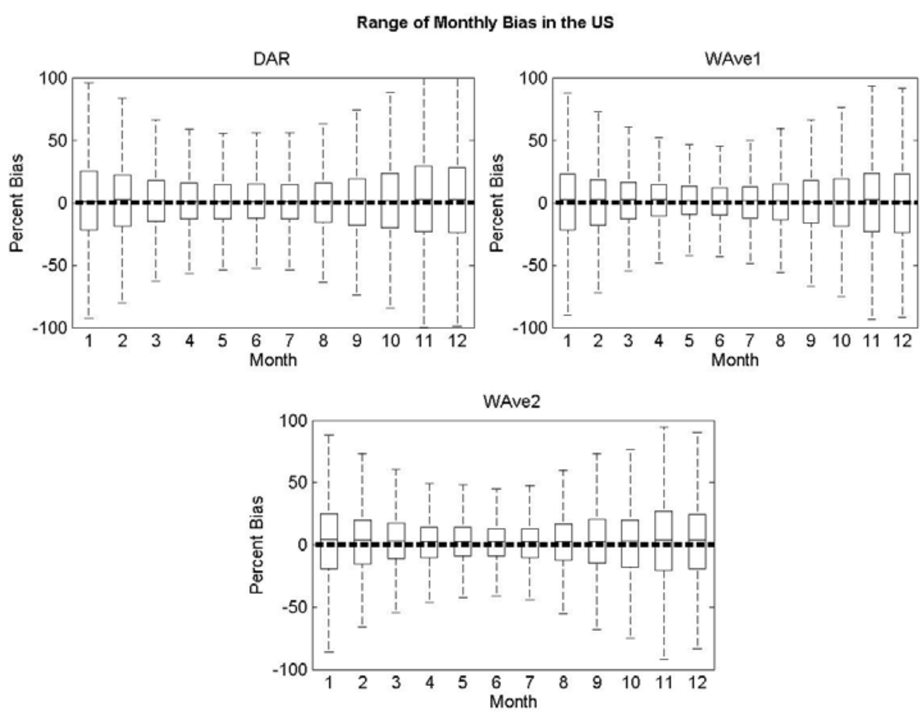

Figure 6. 22. Monthly range of bias for DAR and WAve methods.

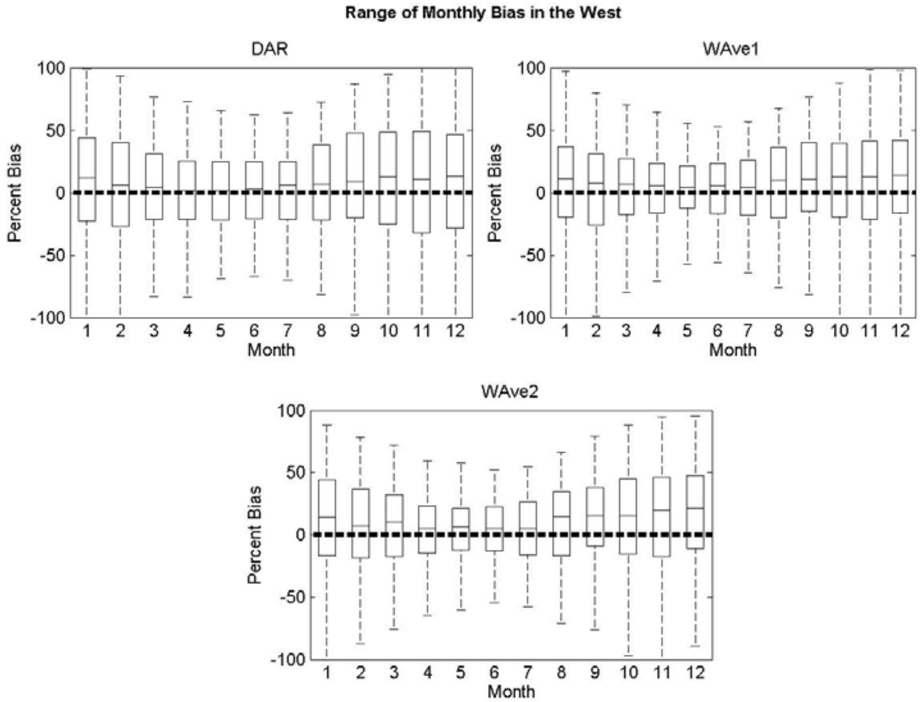

Figure 6. 23. Monthly range of bias for DAR and WAve methods in the West.

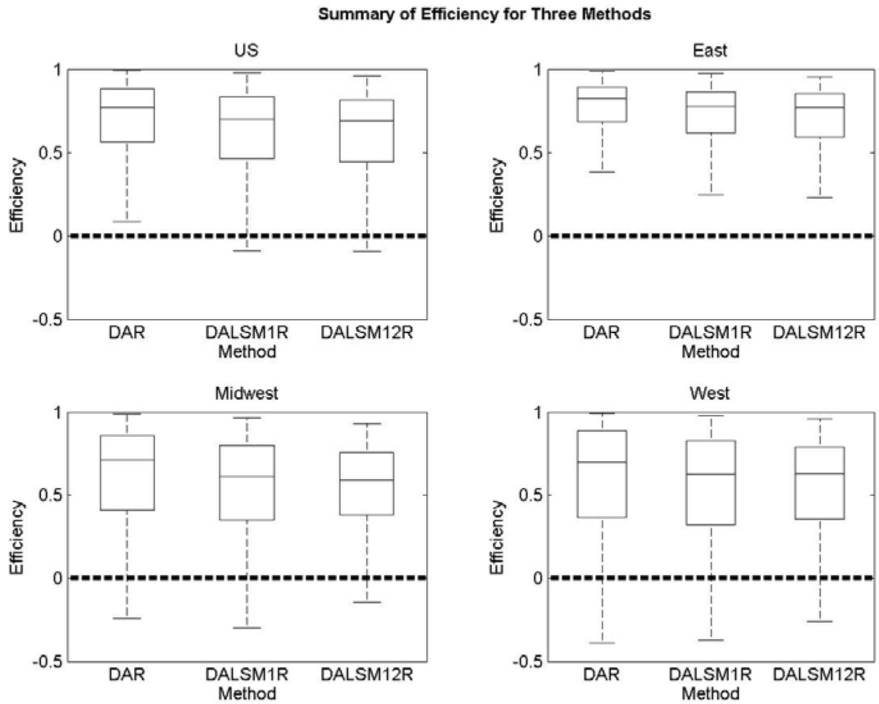

Figure 7. 1. Overall range of NSE for DAR and two combinations of DA and SM.

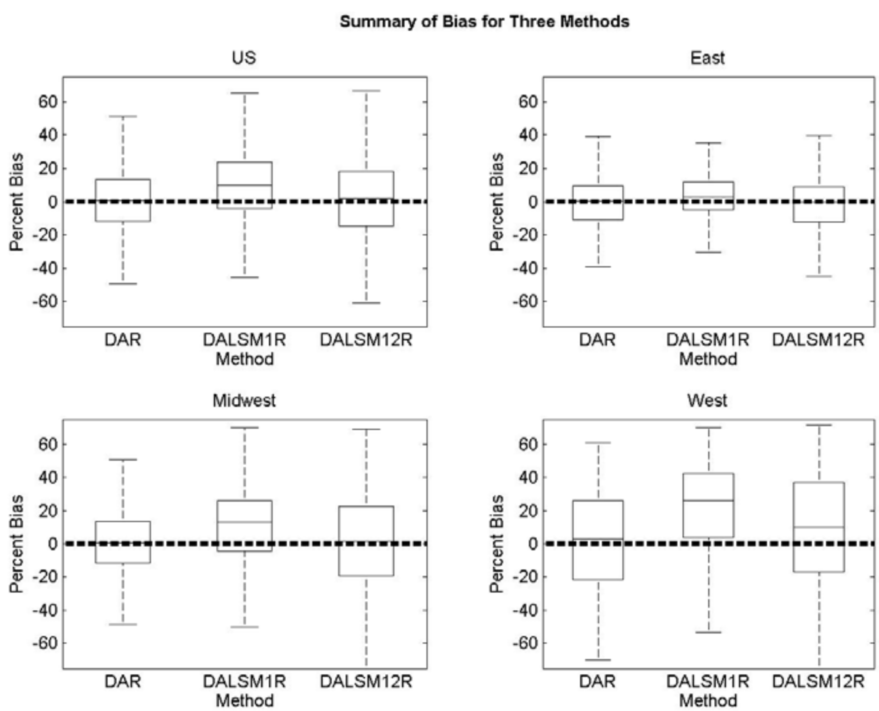

Figure 7. 2. Overall range of bias for DAR and two combinations of DA and SM.

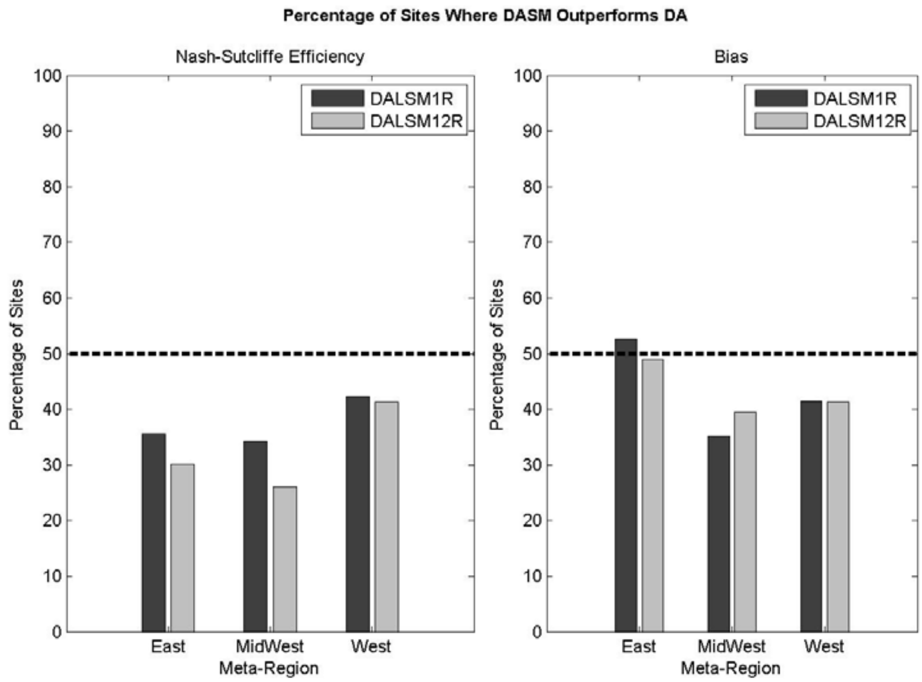

Figure 7. 3. Overall percentage of sites where DA-SM combinations outperform DAR.

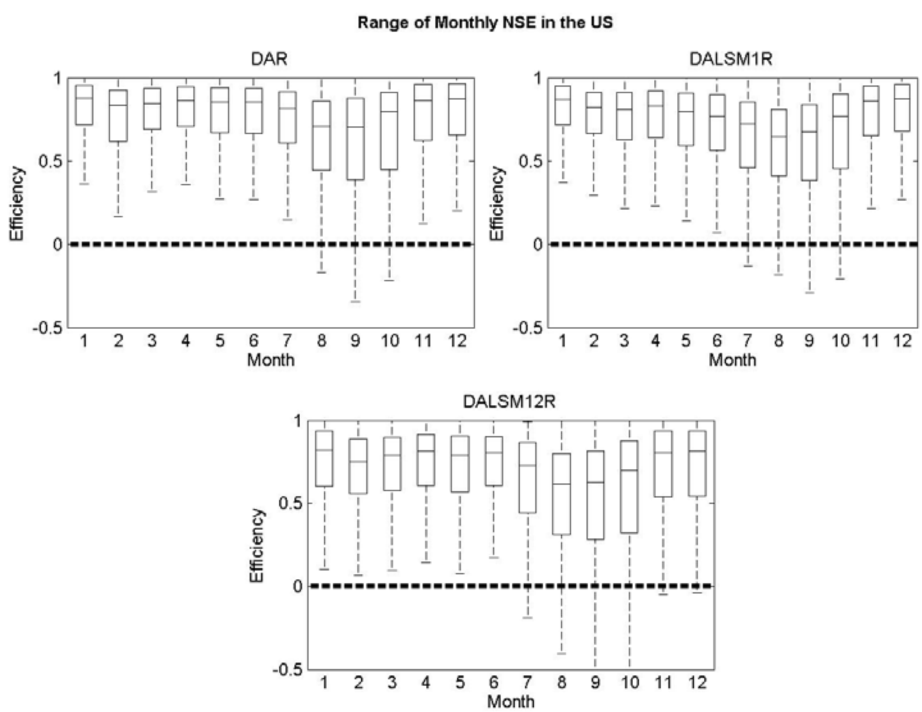

Figure 7. 4. Monthly range of NSE for DAR and two DA-SM combinations.

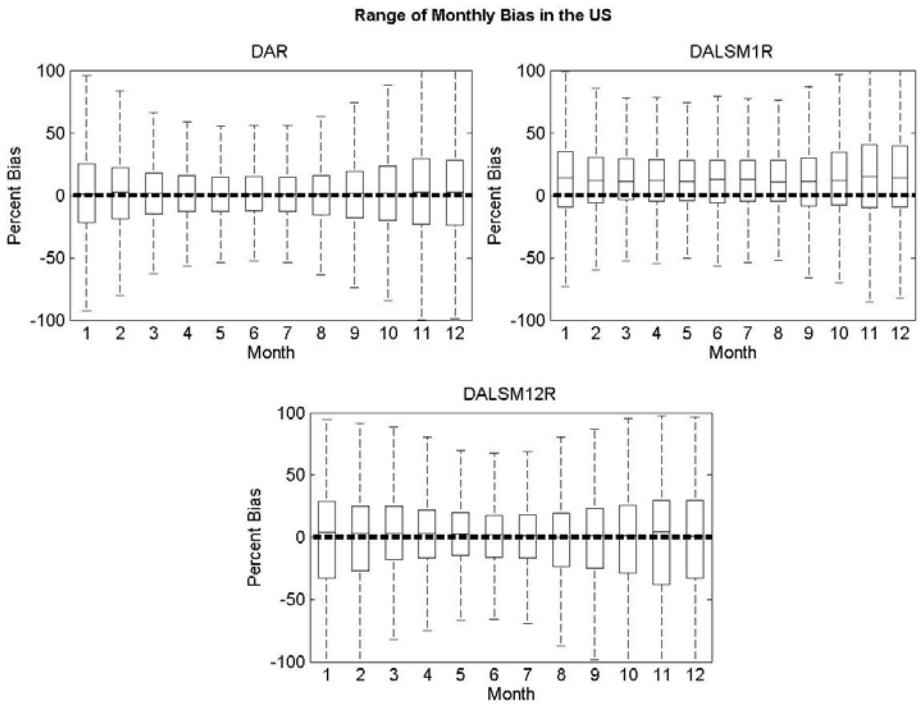

Figure 7. 5. Monthly range of bias for DAR and two DA-SM combinations.

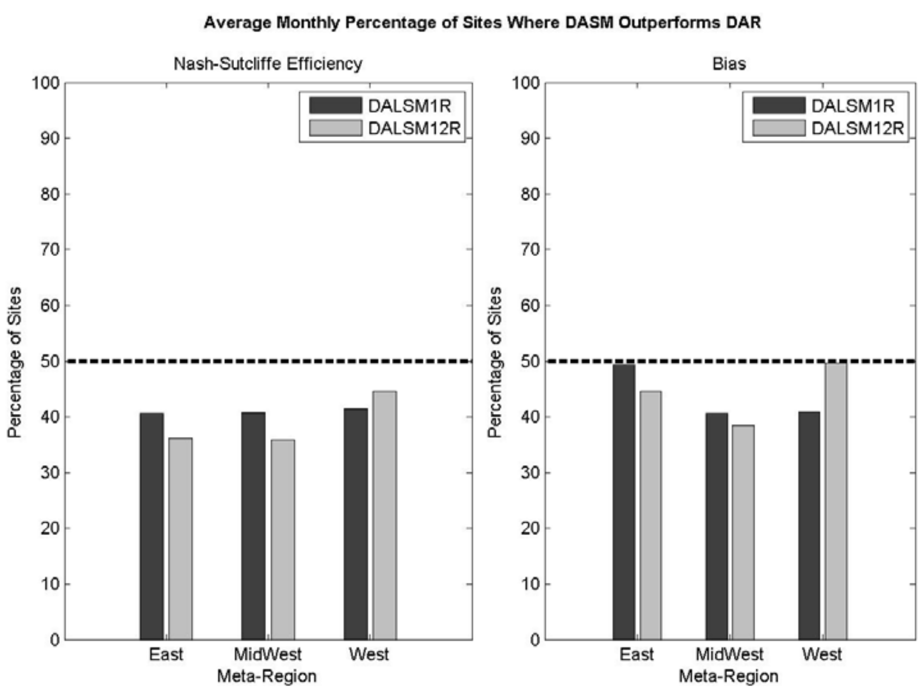

Figure 7. 6. Average monthly percentage of sites where DA-SM combinations outperform DAR.

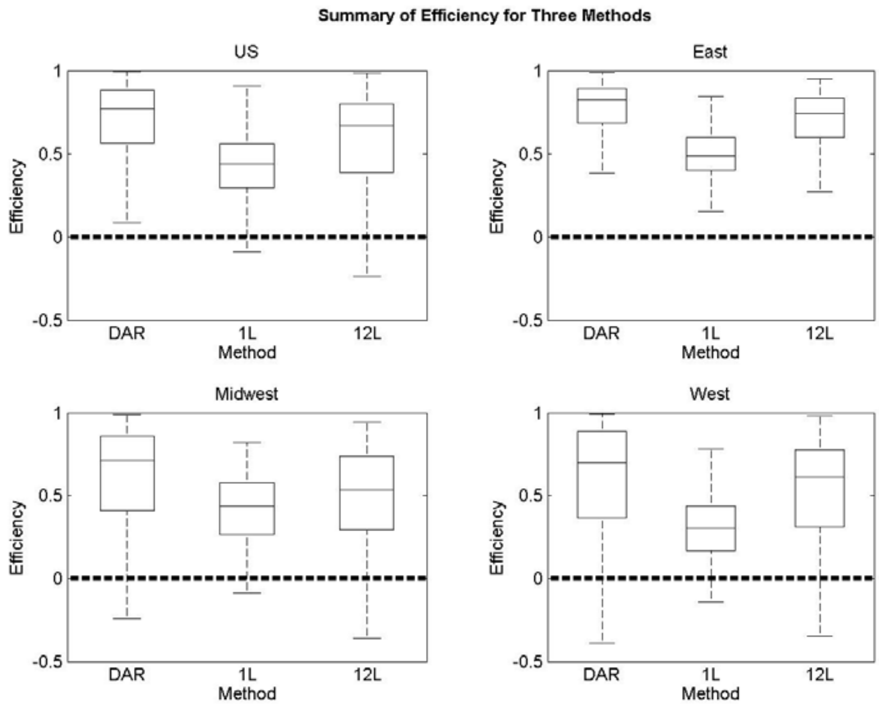

Figure 7. 7. Overall range of NSE for DAR and two DA-MOVE combinations.

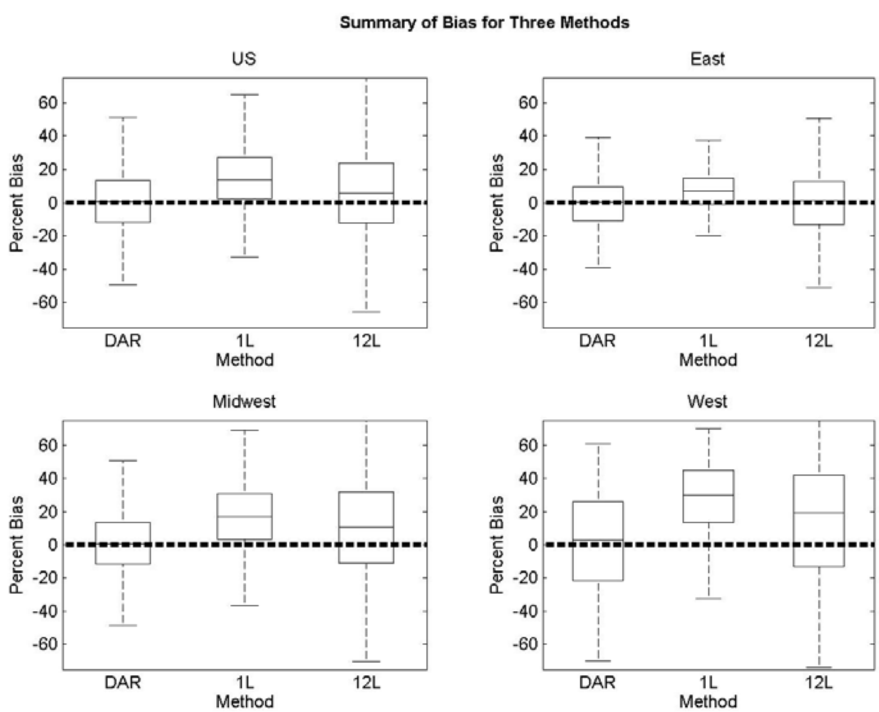

Figure 7. 8. Overall range of bias for DAR and two DA-MOVE combinations.

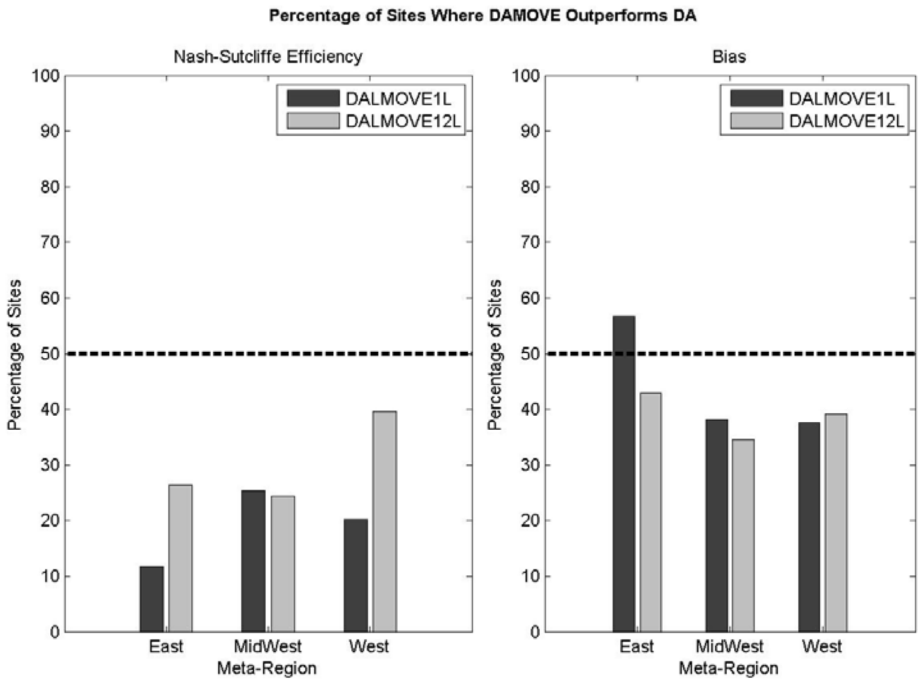

Figure 7. 9. Overall percentage of sites where DA-MOVE combinations outperform DAR.

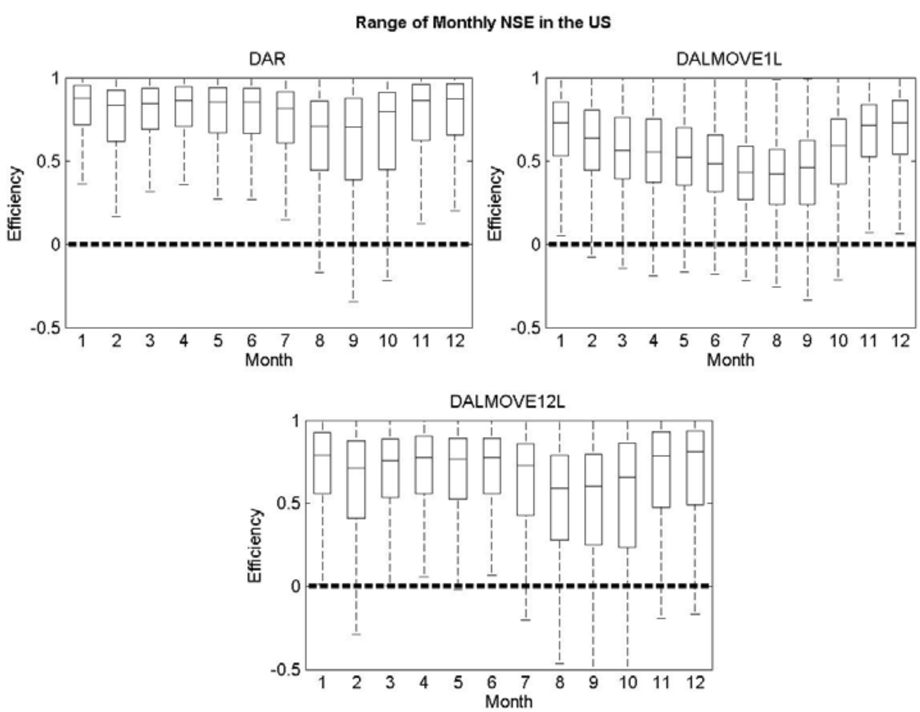

Figure 7. 10. Range of monthly NSE for DAR and two DA-MOVE combinations.

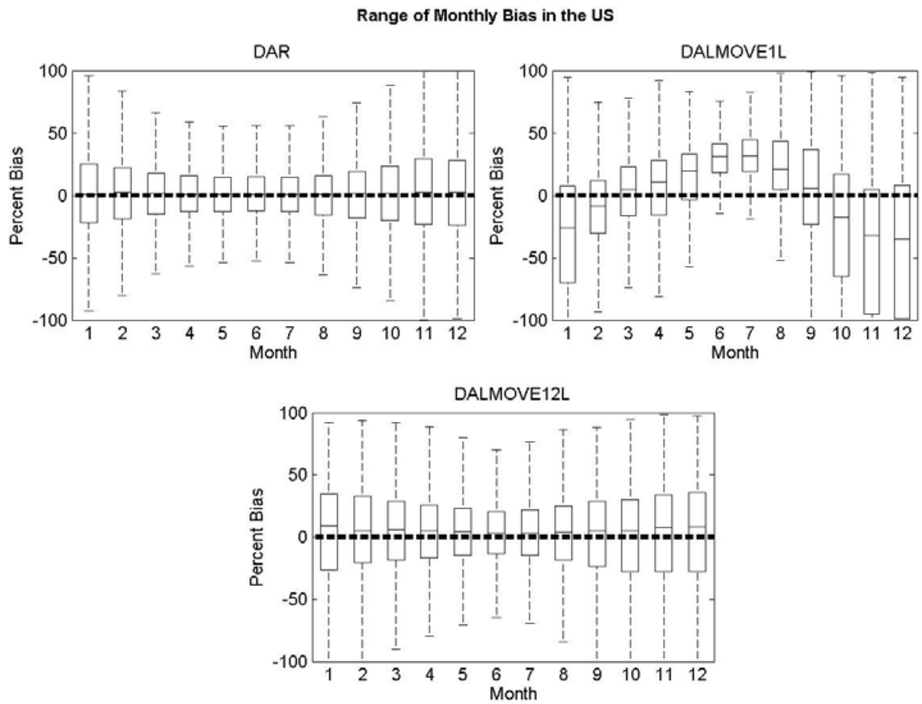

Figure 7. 11. Range of monthly bias for DAR and two DA-MOVE combinations.

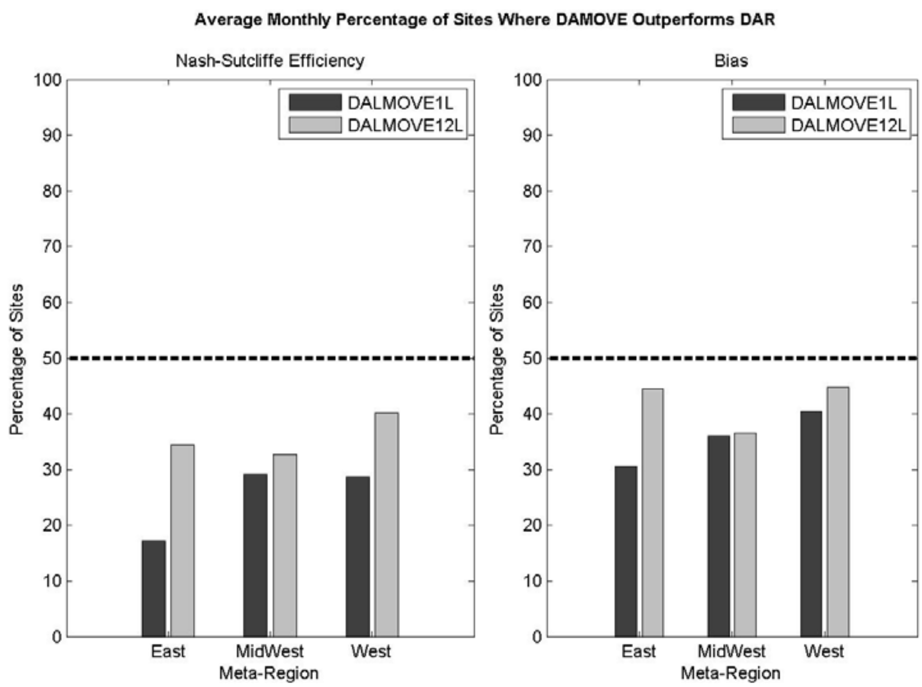

Figure 7. 12. Average monthly percentage of sites where DA-MOVE combinations outperform DAR.

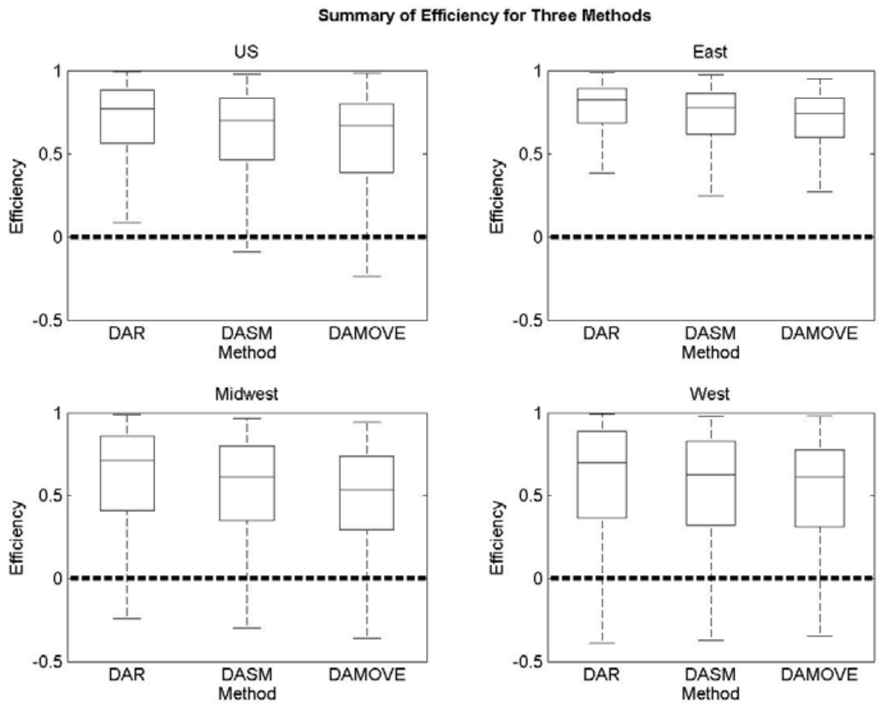

Figure 7. 13. Overall range of NSE for DAR and two combination methods.

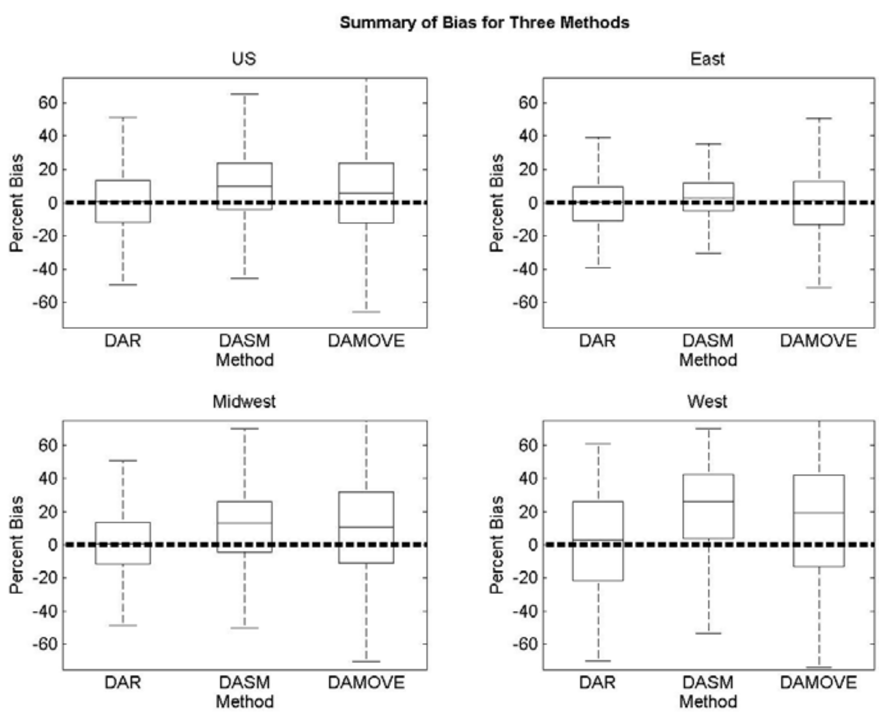

Figure 7. 14. Overall range of bias for DAR and two combination methods.

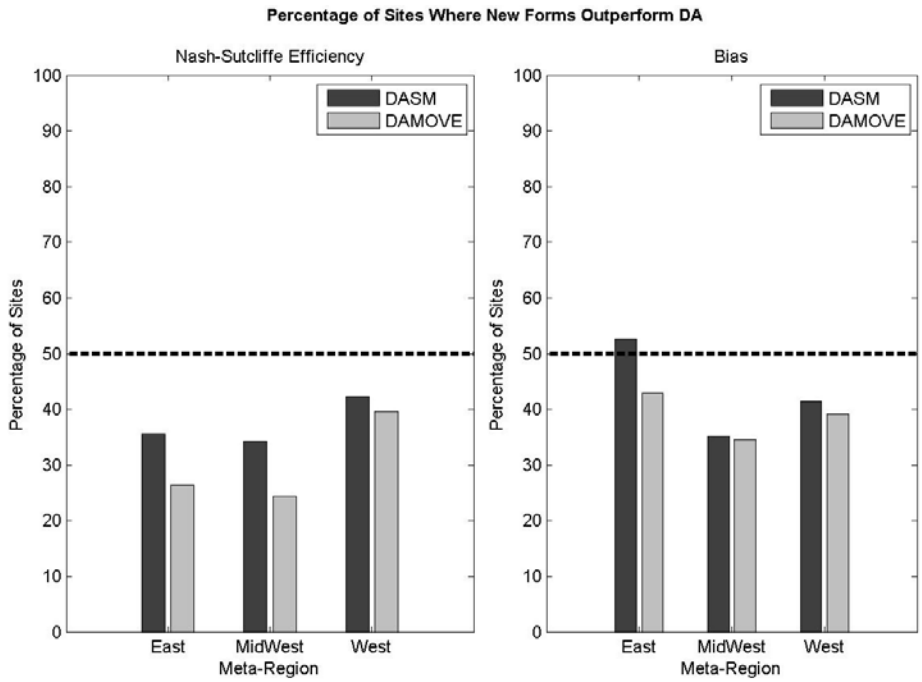

Figure 7. 15. Overall percentage of sites where combination methods outperform DAR.

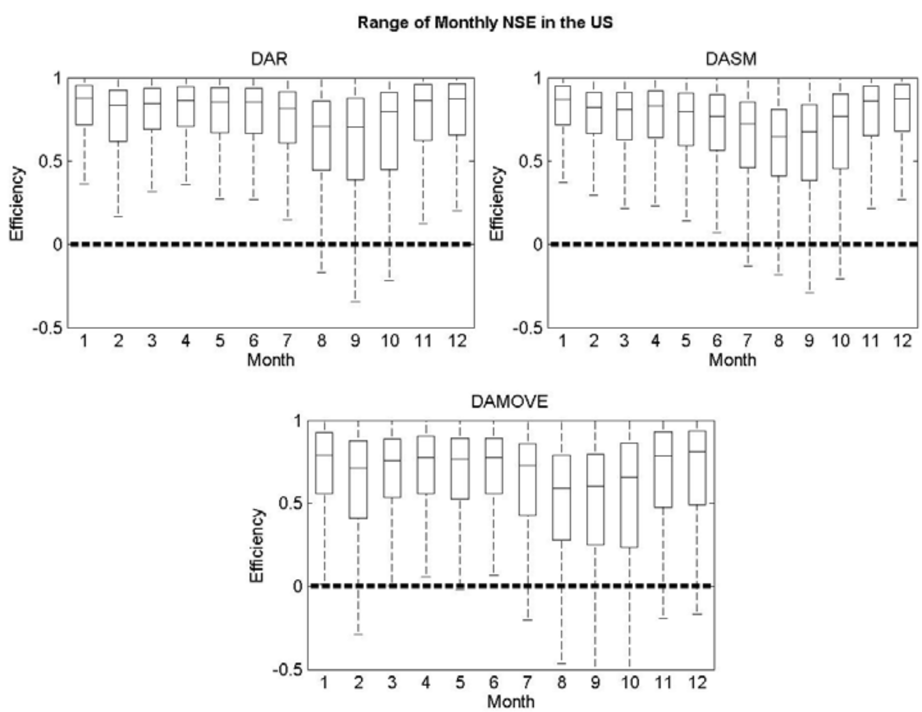

Figure 7. 16. Monthly range of NSE for DAR and two combination methods.

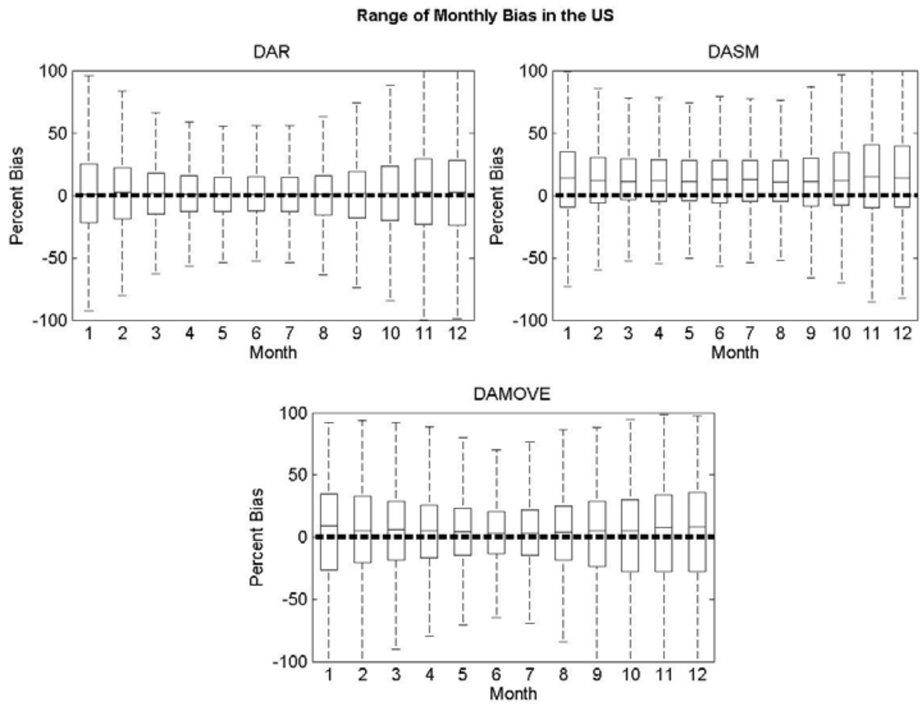

Figure 7. 17. Monthly range of bias for DAR and two combination methods.

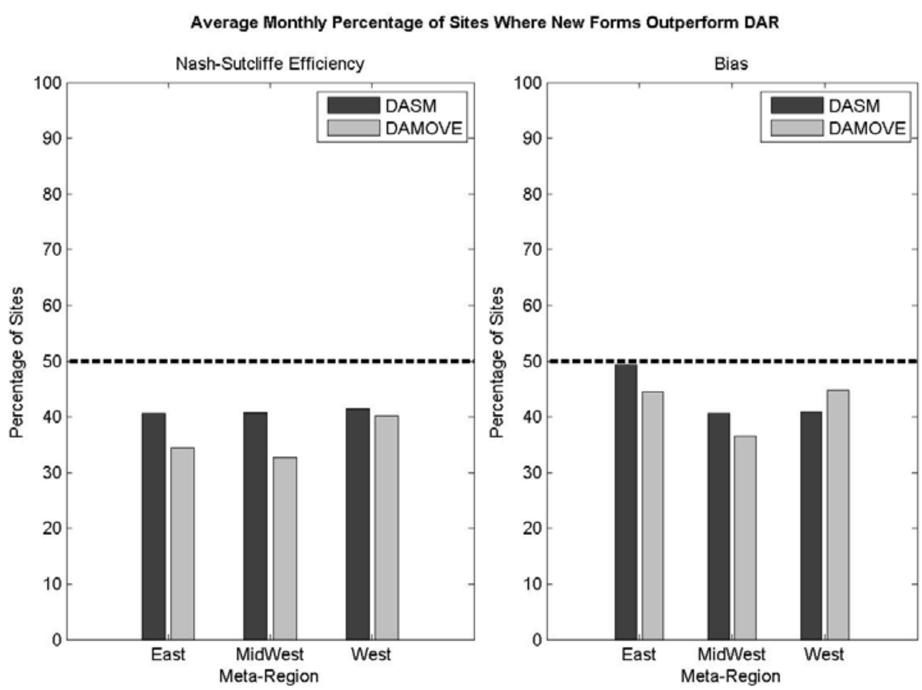

Figure 7. 18. Average monthly percentage of sites where combination methods outperform DAR.