Appendix C

Contents

Figures

C24. The distribution of at-site percent errors in the estimated 10-percent-exceedance streamflow for each method of prediction in ungaged basins is considered here

C25. The distribution of at-site percent errors in the estimated mean daily streamflow for each method of prediction in ungaged basins is considered here

C26. The distribution of at-site percent errors in the estimated coefficient of variation of daily streamflow for each method of prediction in ungaged basins is considered here

C27. The distribution of at-site percent errors in the estimated skewness of daily streamflows for each method of prediction in ungaged basins is considered here

C28. The distribution of at-site percent errors in the estimated kurtosis of daily streamflows for each method of prediction in ungaged basins is considered here

C29. The distribution of at-site percent errors in the estimated lag-1 autocorrelation of daily streamflows for each method of prediction in ungaged basins is considered here

C30. The distribution of at-site percent errors in the estimated amplitude of the sinusoidal seasonal trend of daily streamflows for each method of prediction in ungaged basins is considered here

C31. The distribution of at-site percent errors in the estimated phase of the sinusoidal seasonal trend of daily streamflows for each method of prediction in ungaged basins is considered here

C32. The distribution of the at-site root-mean-square-normalized errors of estimated Fundamental Daily Streamflow Statistics for each method of prediction in ungaged basins is considered here

Appendix C figures are located in a separate PDF file, available at *http://pubs.usgs.gov/sir/2014/5321*.

Introduction

This appendix supplements the material presented in the body of this report. It is included only as documentation of the complete analysis and is meant to encourage further research. Herein, little attempt is made to explain any of the particularities of data. Full figures of all metrics considered and discussed in the report body are included.

Overall Goodness-of-Fit of Estimated Streamflow Records

Nash-Sutcliffe Efficiency of Estimated Daily Streamflow

The Nash-Sutcliffe efficiency of daily streamflow (NSE) quantifies how accurately the day-by-day observations of streamflow are reproduced in the estimated record. The distribution of the at-site NSEs for each prediction of ungaged basins (PUB) method is shown in figure C1. The horizontal axis indicates the PUB methods while the vertical axis shows the calculated efficiency. Most of the methods do poorly, with those based on annual moments performing worst. In general, it can be seen that the transfer-based methods using the nearest-neighbor algorithm to select an index gage perform better than those that rely on the map-correlation algorithm. The best methods are the nearest-neighbor implementations of the drainage area ratio, QPPQ, monthly standardizations (SMS12 and SM12), and the Precipitation Runoff Modeling System (PRMS) and NN-AFINCH.

Nash-Sutcliffe Efficiency of the Logarithms of Estimated Daily Streamflow

The Nash-Sutcliffe efficiency of the logarithms of non-zero daily streamflow (NSEL) captures the day-by-day performance of the model but is somewhat less sensitive to outliers in the analysis. Taking the logarithms of the non-zero flows removes some of the extreme skew in the record. As a result, the statistic is more robust. While NSE is largely dominated by large errors in large flows, the NSEL may be more sensitive to errors in the smaller flows. The distribution of NSEL for each PUB method is shown in figure C2. Even though the values of NSEL are slightly greater than the NSE, suggesting better performance, the conclusions remain similar: Methods relying on annual moments do not reproduce the record well. Transfer-based approaches using nearest-neighbor are superior to those relying on map correlation. The best methods remain the nearest-neighbor implementations of the drainage area ratio, QPPQ, monthly standardizations (SMS12 and SM12), and the PRMS and NN-AFINCH.

Root-Mean-Square Error of Estimated Daily Streamflow

The root-mean-square error (RMSE) of daily streamflow is a function of the Nash-Sutcliffe efficiency and the variability of the observed record. Because the RMSE is a mean value of highly skewed data, the statistic is highly sensitive to extreme values. The result is the large values of RMSE seen in figure C3. The conclusions reached by considering the NSE are nearly identical to those seen derived from the RMSE. The errors are greatest for methods relying on annual moments; nearest-neighbor outperforms map correlation. The nearest-neighbor implementations of the drainage area ratio, QPPQ, monthly standardizations (SMS12 and SM12), and the PRMS and NN-AFINCH all perform well.

Root-Mean-Square-Normalized Error of Estimated Daily Streamflow

Because the RMSE is in streamflow units, it is difficult to compare the RMSE at one site to the RMSE at another site, especially if the streamflows between the sites are starkly different. This can be corrected by normalizing each error by the observation, much like a percent difference. Conducting this process produces the root-mean-square-normalized error (RMSNE) of daily streamflow. The RMSNE is defined for a given site as:

$$E = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{S_i - O_i}{O_i}\right)^2}$$

where S is the simulated streamflow and O is the observed streamflow on day i out of n days. The distributions of RMSNE are shown in figure C4. Because the RMSNE is still a mean value, the errors can appear large. Still, the message is the same: The annual methods are weakest; NN outperform MC; the best methods are the nearest-neighbor implementations of the drainage area ratio, QPPQ, monthly standardizations (SMS12 and SM12), and PRMS and NN-AFINCH.

Average Percent Error of Estimated Daily Streamflow

The average percent error of daily streamflow gives an indication how accurate each PUB method is on a given day. As can be seen in figure C5, all of the PUB methods here show a positive bias and vary widely. The annual methods show the greatest range and magnitude of error. The bias in the map-correlation results is of a greater magnitude than the nearest-neighbor results and shows a wider range of variability. The average percent error is distinct from but similar to the percent bias, which is discussed as the average error in the mean below.

Pearson Correlation between Observed and Estimated Daily Streamflow

The Pearson correlation between observed and estimated streamflow quantifies the agreement between the two records. If the estimated series is a perfect prediction of the observations, the couplets of observed and estimated flow would fall along a one-to-one line and the Pearson correlation would be one. The distribution of the observed Pearson correlations for each PUB method is shown in figure C6. It is immediately apparent that the nearest-neighbor approaches are superior to the map-correlation implementations. From this measure, the annual methods appear more competitive than before. From this result, the choice of index gage seems to drive the Pearson correlation most. The PRMS model tracks closest with the nearest-neighbor approaches, while the two AFINCH group align with their index-gage cohorts.

Spearman Correlation between Observed and Estimated Daily Streamflow

The Spearman correlation between observed and estimated streamflow captures the agreement between the ranks of the observed and estimated flows. Unlike Pearson, the magnitude of error plays little part in Spearman, which relies only on the relative ranking of magnitudes. By this metric, all of the methods perform well (fig. C7). There is still a distinction between nearest-neighbor and map correlation, but the annual methods do not show the same degree of weakness seen in other metrics. This result suggests that the relative timing of the streamflow record is effectively transferred by all methods, but the magnitude of accumulation of errors drives the low values of Nash-Sutcliffe efficiency.

Ability to Reproduce Observed Storage-Yield Curves

In addition to overall goodness of fit, a reliable PUB method should also reproduce the cumulative properties and signatures of the streamflow record. The storage-yield curve (SYC) is one such signature and is used here to assess the cumulative impact of prediction errors on the possible applications of the estimated records. Here, the SYC is estimated using a constant-yield, no-fail, daily sequent peak algorithm (Thomas and Burden, 1963). See the report body for more information.

Nash-Sutcliffe Efficiency of Estimated Storage-Yield Curves

Because the SYCs are generated from a uniform distribution of yield fractions, data do not exhibit the level of skew seen in the original streamflow records. Accordingly, the Nash-Sutcliffe efficiency of the SYCs (SYC-NSE) is more reliable than the NSE of streamflow. From figure C8, which shows the distributions of the SYC-NSE, all of the methods reproduce the SYC well. These values of NSE are even greater than was seen in the overall goodness of fit. The standardization of flows with an annual mean and standard deviation is the only method that fails to reliably reproduce the SYC. There is less of a distinction between nearest-neighbor and map correlation here, but the NN methods remain slightly superior. Some of the methods, such as the drainage-area ratio, are highly sensitive to the index gage.

Nash-Sutcliffe Efficiency of Logarithms of Estimated Storage-Yield Curves

As before, in the case of skewed data, the NSE of the logarithms of the nonzero realizations of a quantity might be more reliable than the original NSE. There is little skew present in the SYC, so the SYC-NSEL—the distributions of which are shown in figure C9—gives nearly the same message as the NSE-SYC. Still, the SYC-NSEL suggests an interesting weakness of NN-SMS12R.

Root-Mean-Square Error of Estimated Storage-Yield Curves

Figure C10 shows the distribution of root-mean-square error in the SYC for each PUB method. The best methods are NN-QPPQ, NN-SMS12L, NN-SM1, NN-SM12, and MC-SMS12L. In general, the nearest-neighbor methods are better than the map-correlation methods, but this is not uniformly the case. By this measure, the more process-based methods (PRMS and AFINCH) do not perform as well as the transfer-based methods using nearest-neighbor selection criteria.

Root-Mean-Square-Normalized Error of Estimated Storage-Yield Curves

Because there is not much skew in data, normalizing the RMSE does not greatly alter the conclusions. (See above for calculation of the RMSNE; here the RMSNE was calculated based on observed (*O*) and simulated (*S*) storage.) Figure C11 shows the distribution of the root-mean-square-normalized error of the SYC. The RMSNE-SYC does show an improved performance in the PRMS. This, in comparison to previous results, suggests that the error is greatest in the high end of the

SYC for the PRMS, but that the error is of a similar proportion along the SYC. The PRMS may be misrepresenting a portion of the distribution of flows.

Average Percent Error of Estimated Storage-Yield Curves

The distribution of the average percent error of the estimated storage is shown in figure C12 for each PUB method. Most of the methods show a significantly positive error on average in the SYC. The PRMS and AFINCH show a strong underestimate of the SYC. There is only a slight distinction between the nearest-neighbor set and map-correlation group. NN-DAR produces the smallest median error. The percent error is interesting because it can be interpreted as resulting in overdesign or underdesign. In this case, the process-based models lead toward underdesign, while the transfer-based models result in only a slight overdesign.

Pearson Correlation between Observed and Estimated Storage-Yield Curves

The Pearson correlations between observed and estimated SYCs are all quite high (fig. C13). There is only a slight difference between the nearest-neighbor and map-correlation implementations. This uniformly high performance suggests the errors associated with the daily records of streamflows are smoothed out, to some degree, when aggregated and sequenced to produce a simplification like the SYC.

Spearman Correlation between Observed and Estimated Storage-Yield Curves

Because Spearman correlation only considers the ranks of a series and the SYC is a simple, monotonic curve, all of the Spearman correlations are nearly perfect for almost every PUB method shown in figure C14. It is, therefore, remarkable that only the annual standardization of flows with a mean and standard deviation produces such dissimilar results. All of the other methods perform well.

Ability to Reproduce Observed Flow Statistics

Coefficient of Variation of Annual Streamflow

The distribution of percent error in the estimate of annual coefficient of variation calculated from each estimated flow series is shown in figure C15. The transfer-based methods show general unbiasedness. The best of these methods are the nearest-neighbor implementations of the drainage area ratio, and standardizing by annual or monthly means. The more process-based methods (PRMS and AFINCH) underestimate the annual coefficient of variation. The coefficient of variation of the annual flows is calculated by aggregating the daily flows to annual flows and then taking the ratio of the mean and standard deviation of these annual flows. This is slightly different than the coefficient of variation of daily streamflow and the L-CV of daily streamflow, which are discussed below.

Coefficient of Variation of Daily Streamflow

All of the PUB methods show more variability in the prediction of the daily coefficient of variation than in the annual coefficient. The transfer-based methods all show approximate unbiasedness, but some have a much greater variability—especially the annual methods (fig. C16). There is little difference between the nearest-neighbor and map-correlation cohorts. The same methods remain strongest: the nearest-neighbor implementations of the drainage area ratio, and standardizing by annual or monthly means. The more process-based methods (PRMS and AFINCH) continue to produce a strongly negative median bias. The coefficient of daily flows is the ratio of the mean and standard deviation of daily flows; for related statistics, see the discussion of the coefficient of variation of annual streamflow and the L-CV of daily streamflow.

10th Percentile of the 7-Day Average Annual-Minimum Streamflows

The error in low-flow statistics is much more variable than in other metrics. The distribution of errors in the 10th percentile of the 7-day average annual-minimum event is shown in figure C17 for each PUB method. Except for the PRMS, most of the competitive methods show a negative error. The SMS12R method produces the greatest magnitude of error, on average. NN-SMS12L produces the best balance between median bias and variability of bias. The 10th percentile of the 7-day average annual-minimum streamflow is based on the empirical distribution of 7-day average annual minimums; this is related to but distinct from the 10-year, 7-day average annual minimum, which is typically calculated by fitting a log-Pearson distribution.

50th Percentile of the 7-Day Average Annual-Minimum Streamflows

The 50th percentile of the 7-day average annual-minimum event shows slightly more variability across PUB methods than the 10th percentile. The distributions of these percent errors are shown in figure C18. Only the NN-DAR and NN-SM12 show minimal bias on average. NN-SM12 offers a median error of -3 percent with a relatively small amount of spread. NN-SMS12L shows smaller variability, but with a greater magnitude of median error (-16 percent). As with the 10th percentile, the 50th percentile of the 7-day average annual minimum is closely related to the 2-year, 7-day average annual minimum. The 10th percentile was calculated using only an empirical distribution of 7-day average annual minimums rather than fitting a specific distribution.

90th Percentile of the Annual-Maximum Streamflows

Figure C19 shows the distribution of errors in the 90th percentile of the distribution of the annual-maximum events for each PUB method. All of the transfer-based methods produce a relatively unbiased estimate of this percentile, on average, with medians ranging from -5 percent to 8 percent. The variability of bias is greatest for the annual methods with the nearest-neighbor approaches showing slightly less variability than the map-correlation applications. The unbiasedness of this annual-maximum event, when compared to the annual minimum presented earlier, suggests that all the PUB methods are better at predicting high flows than low-flow events. Both THE PRMS and both iterations of AFINCH underestimate the flood event: medians errors of -19 percent, -20 percent, and -29 percent. The 90th percentile of annual-maximum events is closely related to the 10-year flood. (As with annual minimums above, the 90th percentile was estimated with an empirical distribution rather than a fitted distribution.) In light of this relationship, the biasedness of the more process-based approaches ties back to what was seen with the SYC: these methods lead towards underdesign of water-resources structures, planning smaller reservoirs than needed, or preparing for smaller floods than might be expected.

Daily, 90-Percent-Exceedance Streamflow

Considering the 90-percent-exceedance flow again shows that the low flows are not estimated as well as the high flows. The distribution in the percent error of the 90-percent-exceedance flow is shown in figure C20 for each PUB method. NN-DAR and NN-SM12 and MC-AFINCH all show general unbiasedness. The variability of bias is greatest in the annual transfer-based methods. The PRMS drastically overestimates the low-flow events. The wide variability of performance pertaining to this statistic demonstrates that the low-flow events are not well represented by most PUB methods.

Daily, 75-Percent-Exceedance Streamflow

As compared to the error seen in the 90-percent-exceedance flow, the 75-percent-exceedance flow shows significantly less variability in performance (fig. C21). For all of the transfer-based methods, except the annual method, unbiasedness (±5 percent) is shown with a reasonable amount of variability. Map-correlation causes only a slight exaggeration of the variability of errors. The process-based methods show a strong level of bias; the PRMS in particular continues to overestimate this flow quantile by a median of 51 percent. NN-SM12 balances variability and median error quite well.

Daily, 50-Percent-Exceedance Streamflow

As one moves along the flow duration curve to consider the median, the bias in the transfer-based methods continues to improve. The distribution of the percent error in the median is shown in figure C22 for each PUB method. The annual methods continue to produce a widely variable bias, but the other methods show a reduction in the variability of error. Both the PRMS and AFINCH continue to show a distinct positive error.

Daily, 25-Percent-Exceedance Streamflow

The distribution of the errors in the 25-percent-exceedance flow is not very different from that seen in the median, except that variability is further reduced here. The distributions for all of the PUB methods are shown in figure C23. The drainage area ratio, QPPQ, and the monthly standardizations all show median unbiasedness with a much smaller variability than seen previously. The annual methods overpredict this high-flow event. The more process-based methods continue to overpredict, though the median is much less egregious here, especially for the AFINCH realizations. This begins to suggest that the more process-based methods are strongly driven by high-flow fitting.

Daily, 10-Percent-Exceedance Streamflow

The distribution of errors in the 10-percent-exceedance event is nearly identical to the distribution of errors in the 25-percent-exceedance event. The distributions for every PUB method are shown in figure C24. The transfer-based methods remain unbiased, with the annual methods continuing to show a positive bias with a high degree of variability. AFINCH is nearly unbiased, while the PRMS continues to show a slightly greater amount of negative bias. As the bias in the process-based methods is smaller here than in the 25-percent-exceedance event, the process-based methods appear to be tied closely to reproducing high-flow events and the cost of overestimating low flows.

Ability to Reproduce Fundamental Daily Streamflow Statistics

It was recently shown that seven statistics can be used to characterize the distribution of daily streamflow (Archfield and others, 2013). These Fundamental Daily Streamflow Statistics (FDSS) include the mean daily streamflow, the coefficient of variation of daily streamflow (L-CV), the skew of daily streamflow (L-skew), the kurtosis of daily streamflow

(L-kurtosis), the lag-1 autocorrelation coefficient of daily streamflow, and the amplitude and phase of the sinusoidal seasonal trend of daily streamflow. Reliable PUB methods should accurately reproduce the distribution of daily streamflow. The following section presents the errors of each PUB method in reproducing the FDSS.

Mean Daily Streamflow

Nearly all of the PUB methods considered here reproduce an unbiased estimate of the mean daily streamflow. The distribution of the error in the mean is shown for each PUB method in figure C25. This statistic is equivalent to the percent bias, a common statistical tool for assessing an estimator. The same methods continue to succeed: nearest-neighbor implementations of DAR, QPPQ, and monthly standardizations (SM12 and SMS12). The annual methods are not competitive. The PRMS produces a distinct positive error in the mean, while AFINCH slightly underestimates the mean.

Coefficient of Variation of Daily Streamflow (L-CV)

The L-CV of daily streamflow is a more reliable estimate of the variability of daily streamflow than the normal CV (Vogel and Fennessey, 1993). The distributions of errors in the L-CV are shown in figure C26. The transfer-based methods produce a slight positive bias, but are generally unbiased. The least biased methods are NN-DAR, NN-SM1, and NN-MS12. The PRMS significantly underestimates the variability in daily streamflow (median error of -17 percent), a problem that could have design and management implications. The AFINCH methods show a slightly negative median error (near –10 percent) as well.

Skewness of Daily Streamflow (L-skew)

The skewness of the daily record represents the third moment of the distribution. The distribution of the error in the L-skew for each PUB method is shown in figure C27. Except for NN- and MC-SMS12R, all of the transfer-based methods produce unbiased estimates of the L-skew. The PRMS and NN- and MC-AFINCH continue to underestimate these higher order moments. The discrepancy between process-based and transfer-based model behavior suggests that the error structure is distinctly different between the two approaches; it may be that the transfer of information solely from an index gage carries some intrinsic natural behavior.

Kurtosis of Daily Streamflow (L-kurtosis)

The L-kurtosis represents the fourth moment of the distribution of daily streamflow. As can be seen in figure C28, the transfer-based methods continue to produce relatively unbiased estimates of the L-kurtosis while the more process-based models produce a distinct underestimate. This is further evidence that the error structure in the predicted records of these two broad approaches is inherently different. The natural flow information from an index gage, whereas AFINCH only transfers relative timing, improves the reproducibility of the parameters of the distribution of daily streamflow.

Lag-1 Autocorrelation of Daily Streamflow

All of the transfer-based methods are nearly identically unbiased in their ability to reproduce the lag-1 autocorrelation of daily streamflow (fig. C29). The daily disaggregation of AFINCH, which relies on an index gage to downscale flows, also produces unbiased estimates of the lag-1 autocorrelation. The PRMS, on the other hand, underestimates the lag-1 autocorrelation. Again, this difference could be driven by the use of an index-gage to map the timing of events: by using an index gage, a “natural” estimate of lag-1 autocorrelation is explicitly transferred, while a process-based model must attempt to reproduce that autocorrelation from scratch.

Amplitude of the Sinusoidal, Seasonal Trend of Daily Streamflow

All of the methods, except AFINCH, produce an unbiased estimate of the amplitude of the sinusoidal seasonal trend in daily streamflow (fig. C30). The PRMS is surprisingly in line with all of the transfer-based methods. AFINCH, which is at its heart a monthly model, overestimates the amplitude of the seasonal trend. This is true for both the NN and MC implementations. For more information on the estimation of the amplitude of the sinusoidal seasonal trend, see Archfield and others (2013).

Phase of the Sinusoidal, Seasonal Trend of Daily Streamflow

The phase, or timing, of the sinusoidal seasonal trend is unbiasedly represented by all of the PUB methods considered here (fig. C31). The PRMS slightly underestimates the timing of the seasonal trend and produces a wider range of variability. As this is again an issue of timing, like the lag-1 autocorrelation, this difference may be the result of the mechanism of prediction: the transfer-based models (and AFINCH) rely on an index gage for daily timing, while the PRMS must work from climate inputs alone and reproduce those processes. For more information on the estimation of the phase of the sinusoidal seasonal trend see Archfield and others (2013).

Root-Mean-Square-Normalized Error across All Fundamental Daily Streamflow Statistics

The root-mean-square-normalized error (RMSNE) can be used to quantify the average error across a set of statistics that significantly differs in scale or units. This statistical set is calculated using the same formula as the root-mean-square-normalized error of daily streamflow, except that the seven FDSS are used in the formula instead of daily streamflow values. The result is that each error in each statistic is standardized to be a percent error. By standardizing errors, all seven FDSS can be combined into a single statistic that sums up how well each PUB method reproduced the distribution of daily streamflow. The distribution of the RMSNE of FDSS is shown in figure C32. The annual methods are clearly inferior to all others. The map-correlation applications show only a slightly greater RMSNE than their nearest-neighbor counterparts. The best methods are the nearest-neighbor implementations of DAR, QPPQ, and standardizing by monthly moments (SMS12 and SM12). The PRMS and MC-AFINCH show a greater RMSNE than these methods, but the magnitude is still relatively small.