

Chapter 8

Step 6 - Operational Implementation of Calibration Results

Introduction

A properly performed historical data analysis and model calibration should have significant benefits for the operational application of the models and procedures as mentioned in Chapter 1. To repeat, these are:

- Short term river forecasts should more closely track observations, thus improving forecast accuracy and lead time and requiring fewer adjustments by the forecaster.
- Models can be used to generate reliable extended probabilistic predictions.

However, these benefits can be substantially diminished if the operational system is applied in a manner that produces biased results compared to the calibration. There should be differences in the random variations between operational and calibration simulations (ideally the operational results will exhibit less random error), but there shouldn't be a bias.

The primary purpose of historical data analysis is to generate unbiased estimates of the model input data, as well as the output data that will be used to validate the simulations. The aim is also to have the minimal amount of random noise given the available historical data networks. These data are then used to calibrate the hydrologic models to simulate streamflow and other variables throughout a river basin. The primary purpose of calibration is to obtain the parameter values for each model that will generate the most unbiased simulation possible. Calibration also aims at producing parameter values that properly represent the physics that are occurring and are consistent from one watershed to another. The purpose of model calibration is not to produce the best possible value of some "goodness of fit" statistic, but instead to determine physically realistic and consistent parameter values that will allow the models to be used to generate the best possible results in an operational application.

The primary purpose of operational forecasting is to produce an unbiased forecast of what will occur in the future with minimal random variation. In order to use the calibration results to accomplish this goal, several things must occur.

- The operational data estimates must be unbiased compared to those used for calibration.
- The models are defined such that the basic simulation (i.e. the simulation produced using the same input data types and model configuration used for calibration with no alternative data analyses, modeling schemes, or real time adjustments included) will be unbiased compared to that obtained during calibration.
- Real time analysis procedures, that in many cases involve data not available or examined

uring the historical processing step, are applied in a way that reduces the random error in the data estimates, without producing a bias.

- Adjustments are applied to the models that reduce the random variations and unavoidable trends that were in the calibration simulations in a manner that doesn't produce bias in the forecasts.

This chapter describes how to use an operational forecast system so that the real time results will be unbiased compared to calibration simulations. There is some discussion of ways to reduce random error, but the emphasis is on minimizing bias. As with the rest of this manual, the concepts might apply to any forecast system that relies on conceptual hydrologic models, but the discussion is focused on the application of the NWSRFS Operational Forecast System (OFS). It should be clearly stated that this chapter is not intended as a complete guide on how to produce operational river forecasts, but only to describe what must be done to insure that the benefits of calibration are preserved.

The chapter will be divided into two main parts. The first part will indicate what needs to be done to define the basic operational system so that it will generate results that are unbiased compared to those produced during calibration. The amount of random variation will differ from that experienced during calibration primarily due to differences in the density and distribution of the data networks, primarily precipitation. The second part will discuss various procedures and techniques that can be used to reduce the random errors in the data input and model results. The focus will be on applying these procedures and techniques in a manner that will minimize adding any bias.

Operational Definition using Calibration Results

Introduction

In NWSRFS various information needs to be specified in order to use the Operational Forecast System (OFS). This includes the definition of :

- all stations with their available data types and related parametric information,
- all of the areas for which mean values of precipitation, temperature, and evaporation are to be computed along with their parametric information, and
- all segments with the sequence of operations (models and procedures needed to generate simulations of streamflow and other variables of interest) and the parametric information for each operation.

Most of the parametric information needed when defining the operational system are obtained directly from the calibration results or are based on values derived during the calibration. The material presented in this part of the chapter describes guidelines for defining the operational system

so that it should very closely mimic the results obtained during calibration. In this definition no changes are made to the time and space scale that the data analysis procedures and hydrologic models are applied or to the types of data used during calibration. Recommendations for using different data types and analysis methods or varying the time and space scales will be mentioned in the second part of the chapter. These may involve making changes to the basic definition of the operational system described in this part. Except for variations caused by differences between the density and distribution of the real time and historical networks, the operational system as defined by following the guidelines in this part of the chapter should produce, with no run time modifications applied, essentially the same simulation results as the calibration system.

Operational Station Definitions

Station definitions can be one of the major reasons for operational results to be biased as compared to calibration simulations. This is primarily due to differences in the real time and historical data networks. It is common for stations to be available in real time that were not included in the historical records and vice versa. Bias can occur due to an improper specification of parametric information for stations that are only available operationally, equipment at sites with only operational data being incompatible with the gages used for calibration, inconsistencies that occurred after the historical data period, and differences in the reporting characteristics of the networks.

The best case scenario would be for the operational and historical networks to be comprised of exactly the same stations with no changes in location, equipment, or reporting characteristics, however, this is seldom the case. Thus, it is critical to make sure that all the real time data are compatible with the station data used for calibration.

When data for which monthly mean values must be provided operationally (mountainous area precipitation stations and all temperature stations) are checked for compatibility, the checks are always made based on the full period of record of the historical data, i.e. the period for which the data were analyzed and processed, not the period used to calibrate models at any specific location.

This can present a dilemma when different historical periods of record are used for different portions of an RFC area. Generally the same historical data period is used for an entire river basin. Ideally all the river basins that utilize the mountainous area precipitation analysis procedure or use temperature data within the RFC area of responsibility would have the same historical data period. However, this is likely not the case due to such factors as differences in networks over time (e.g. the SNOTEL network record length might only impact a portion of the RFC area - see Chapter 5) and the length of time involved in calibrating an entire RFC area (i.e. last basin calibrated would have more historical data available than the first basin). Stations that are used only for computing mountainous area MAP and any MAT time series within a single river basin are not affected. The dilemma occurs for those stations that are near the divide between two or more basins that have different historical data periods. Mean monthly values for these stations can vary from one historical period to another. Hopefully the differences are not too great. This will likely be the case if the historical data periods are quite long. The best that probably can be done when operationally defining these stations is to use the average, or a weighted average of the mean monthly values for each historical period depending on which basin the station has the most influence.

Though there are similarities in the methods of checking for compatibility from one data type to another, there are also differences, thus each major data type will be discussed separately.

Precipitation Data

Precipitation is the single most important input to the snow and soil moisture accounting models. As shown in Chapter 6 any bias in precipitation will be magnified in the runoff computations. Thus, it is critical that the precipitation observations used operationally be as unbiased as possible when compared to that used in calibration. In order to avoid such bias, all station precipitation data used operationally must be compatible with the historical station data.

The methods for doing this depend on whether a given station is part of a mountainous or non-mountainous area procedure for computing MAP. Stations that can be used in both types of procedures should be checked using each method.

Some RFCs use either non-mountainous or mountainous area procedures over their entire area, however, in some cases the procedure used to process precipitation data varies from one part of an RFC area to another. This is especially true when there is a clear demarcation between the relatively flat terrain and mountains. During calibration either one procedure or the other is used for any group of stations in a precipitation analysis (i.e. either monthly means are defined for all stations or they are not). However, during operations, stations in the transition zone from flat to mountainous terrain could be used for areas that use both types of procedures. For operational precipitation station definitions, any station that can be assigned weight, or used to estimate missing data for a station that can be assigned weight, for an MAP area that uses the mountainous area procedure should have mean monthly precipitation amounts (often referred to as characteristics) included in its definition. Typically this means that all stations in the mountains will have mean monthly values defined, as well as any stations in the adjacent flat lands that can be used to estimate missing data at a mountain station. The mean monthly amounts should be those that are appropriate for the historical period of record used for the mountainous areas. If there is not a clear demarcation between the flat terrain and the mountains, it is usually easier to include mean monthly values for all stations.

- Non-Mountainous Areas - As defined in Chapter 6, a non-mountainous area is an area where the long term average values of the variable being analyzed are essentially the same at all locations. The compatibility checks to be performed for stations in non-mountainous areas depend on whether the station was part of the historical analysis. For real time stations that were part of the historical analysis, the user needs to verify that there have been no station moves or equipment changes since the end of the historical analysis period. Equipment changes would involve substituting a different type of precipitation gage or adding or removing a wind shield. If this is so, then it is reasonable to assume that the real time data should be compatible with that used for calibration. If the station was moved or the equipment altered, then the consistency of the record should be checked by doing a double mass analysis against a group of other precipitation stations in the vicinity. If an inconsistency is found, the appropriate consistency correction should be applied.

n the operational station definition (i.e. a correction that will make the real time data consistent with the historical period - e.g. if the gage now catches more than during the historical period, the operational consistency correction would be less than 1.0). In the OFS real time precipitation corrections can be applied on an annual or seasonal basis.

For real time stations that were not part of the historical analysis, a check should be made that the long term average precipitation for those stations is essentially the same as the stations that were available historically. It might be assumed that since long term average precipitation is the same, or a general trend of slowly increasing or decreasing values exists, over the region, that any new station will have a catch that is consistent with the gages used in the historical analysis. This may not be true due to site exposure or the type of equipment used. Windy, open sites or locations with nearby obstructions might under or over catch precipitation. Especially in regions where snow is important, windy, open sites should be avoided unless the gage is properly shielded. In addition, various studies have shown that the catch from certain types of gages are not compatible with the catch from standard climatological stations. For example, it is recognized that heated tipping bucket gages catch less precipitation, especially snowfall, than a universal weighing gage. Different types of gages can also catch varying amounts of rain. A study by the USGS west of Chicago [*Straub and Parmar, 1998*] compared the catch from 5 NOAA climatological gages (8 inch nonrecording and universal weighing gages) and 10 USGS tipping bucket gages. Using only periods when rain occurred, the USGS gages caught only 86% of the amount recorded by the NOAA gages (the tipping bucket gages were adjusted during periods of intense rain as per manufacturer's specifications). When one gage of each type was installed at the same location for a short period, the USGS gage recorded 91% of the catch of the NOAA gage. The inclusion of a number of such gages in the operational network could result in long term MAP values being biased by 5-10%, depending on the weights assigned to each station, as compared to those used for calibration. This could result in a 10-25% bias in runoff computations.

Before using any stations operationally that were not part of the historical network, it should be verified that the catch from these stations is compatible with the historical data. The method used to do this is to compute the ratio of the catch at the new station to one or more nearby stations that were part of the historical network which haven't had any relocations or equipment changes. This should be done over some period of time in order to remove the effect of spatial variations within individual storms. Typically the ratios are computed using monthly totals from the gages. Generally in a non-mountainous area one or two years of data should be sufficient to obtain reasonably stable values of the ratios. The ratios should be very close to 1.0. If not, either the new station shouldn't be used as part of the operational network or it should be corrected so that its precipitation measurements are compatible to the historical data. Hopefully a couple of years of past data can be obtained from the new station in order to do the comparison. If not, sufficient data should ideally be collected so that the data can be checked before defining the station as part of the operational network. In many regions there is a real deficiency of real time precipitation reports. In these cases, it may be judged that any new gages should im

mediately be used to compute areal values even if their use may create a bias. This may be a reasonable decision, but at least the data should be analyzed after a couple of years to determine if an adjustment is needed.

In the western parts of the United States the RFCs have always recognized that there are tremendous variations in amount of precipitation from one location to another in the mountains and have thus always taken this into account when generating model input. In the east, however, there has been a tendency in the past to ignore the effect of the mountains.

Station weights were computed for both calibration and operations using weighting schemes such as the Thiessen method that only considers the station location. In many cases the historical and operational networks were somewhat different, though for many years most of the gages were in the mountain valleys and thus the overall effect on MAP computations wasn't too great. In addition, empirical rainfall-runoff models, such as API type models, used in the past were not as sensitive to precipitation bias as conceptual soil moisture accounting models such as the Sacramento model. With the switch to conceptual moisture accounting models and the advent of the IFLOWS network that covers much of Appalachia with gages on the mountain ridges, the continued use of the non-mountainous area procedure for mountain areas will likely cause a considerable bias between the operational and historical MAP values and result in over forecasting most flood events. The mountainous area procedure should be used in all such regions, both for historical and operational precipitation processing.

- Mountainous Areas - In a mountainous area mean monthly precipitation amounts must be defined for all stations. The monthly means used in the operational definition should be those that correspond to the period of record used for the historical data analysis. For or real time stations that were part of the historical analysis, the mean monthly values determined using the PXPP program during the historical analysis should be used when the station is defined in the operational system. As with non-mountainous areas, these stations should be checked for consistency if any moves or equipment changes have occurred in the years since the historical analysis period ended. If there are any inconsistencies, corrections should be applied in the operational system so that the real time reports are adjusted to be consistent with the historical analysis period.

For operational stations that were not part of the historical network, mean monthly values must be determined prior to the stations being defined as part of the operational network. The procedure for estimating mean monthly values for each of these new stations for the historical period of record is as follows:

- compute the ratio of precipitation at the new station for each month to that at one or more nearby stations that were used in the historical analysis and have well defined monthly means (i.e. the monthly means are based on a long period of observed data), and then
- multiply these monthly ratios by the mean monthly amounts for each of the nearby s

tations used in the historical analysis to obtain estimates of the mean monthly amount for the new station -- then average or weight the estimates if more than one nearby station was used.

Typically monthly totals are used to compute the ratios rather than daily or storm totals.

The length of the period needed to get a stable estimate of the ratio for each month varies somewhat with the spatial variability of the precipitation in the region. In regions with strong orographic patterns during most events, such as the Pacific Coast, 3 years of data should be sufficient to obtain a good estimate of the ratios. In regions with considerable convective activity at times during the year, such as the intermountain west, experience with the PXPP program has shown that around 5 years of data are needed before a reasonably stable estimate of the mean monthly values are obtained. The data needed to compute the monthly means might already be archived or it may be that such data will need to be collected before the station can be defined and used in the operational system. If the user wanted to start using the new station immediately, estimates of the mean monthly precipitation could be obtained from the isohyetal analysis after it is adjusted to the historical period of record (see Section 6-3). This should only be done if the user has considerable confidence in the isohyetal analysis. Even then local exposure and equipment peculiarities could cause the isohyetal estimate to deviate from the actual value enough to cause a significant bias in the resulting MAP values. Operationally it is best to only use stations that were not part of the historical analysis after sufficient data are available to get a good estimate of the mean monthly precipitation at these sites.

For all types of areas, differences in how historical and real time precipitation data are reported can also produce a bias between the resulting MAP time series. In the past many non automated stations in the real time network used a 0.5 inch reporting criterion, i.e. the observer only was to begin reports after more than that amount of precipitation occurred during a storm. Especially during days with spotty convective activity, there was a definite tendency to significantly overestimate the missing amounts at those stations that didn't report using the few stations that did receive more than 0.5 inches of rain. Undocumented studies at the Southeast RFC and HRL showed that the average bias during summer months in regions with a lot of this type of convective activity and many criterion reporting stations could reach nearly 100%. It could not even be assumed that when a station didn't report that it must have had less than 0.5 inches of precipitation as surveys at two RFCs indicated that such stations only reported about half the time when they should.

The 0.5 inch reporting criterion is no longer used, however, there still are differences, some perhaps very subtle, in how historical and real time data are reported that could produce biased operational precipitation estimates. A Forecast Systems Laboratory (FSL) study [Tollerud, 2001] and conversations with personnel from some RFCs indicate that there can be substantial differences in the number of reports from observers on days with light or scattered amounts as opposed to days when most stations receive some precipitation. It appears that many manual observing sites only report on days when precipitation occurs. The FSL study showed that there were dramatically different frequency distributions of the percent of rain days bet

ween daily reporting stations and automated sites. There were even some differences between automated real time sites and stations in the hourly climatological network. The study also showed that a sizable number of the automated stations have 1-4 missing hourly observations per day. Such real time reporting characteristics, especially having many stations not reporting on days with scattered convective activity, could easily result in biased estimates of station precipitation amounts. This is clearly something that needs to be carefully monitored and evaluated as a significant bias could be produced in many cases when there are substantial differences in how real time and historical precipitation data are reported.

Temperature Data

Biased temperature estimates will have a significant effect on the timing of snowmelt runoff as shown in Chapter 6. They will also have a negative effect on the determination of the form of precipitation. Since for all temperature stations the mean monthly maximum (max) and minimum (min) values must be defined in the operational system, the key, as far as the station definitions, is to make sure that the mean monthly values provided are compatible with the period used to perform the historical data analysis. It doesn't matter whether a given station is to be used for a mountainous or non-mountainous area or both.

For operational stations that were included in the historical analysis, the mean monthly max and min temperatures provided for the operational definition should be exactly the same as those used in the final MAT run for the historical period of record. If any of these stations have been relocated since the end of the historical data period, then the consistency of the record should be checked and if necessary, corrections applied in the operational definition to adjust the current data so that it is compatible with the data for the historical period.

For operational stations that were not part of the historical network, mean monthly max and min temperatures for the historical period need to be computed. The procedure for estimating these values for each new station is as follows:

- compute the difference in max and min temperature for each month between the new station and one or more nearby stations that were used in the historical analysis and have well defined monthly means (i.e. the monthly means are based on a long period of observed data), and then

- add these differences to the mean monthly max and min temperatures for each of the nearby stations used in the historical analysis to obtain estimates of the mean monthly max and min temperatures for the new station -- then average or weight the estimates if more than one nearby station was used.

The differences should be computed using monthly average max and min temperatures to avoid the effects of different observation times and frontal passages. Only months with complete records should be used. Monthly differences, rather than long term, should be used because the relationship between stations can vary seasonally, even in non-mountainous areas.

Typically a couple of years of data should be sufficient to determine stable temperature differences since temperature can be observed every day as opposed to precipitation which occurs only periodically and because temperature is not as spatially variable as precipitation.

Evaporation Data

Evaporation data only need to be defined in the operational system for stations where Potential Evaporation (PE) is computed on a daily basis from meteorological factors. If the models only use climatological estimates of mean evaporation rates, these values are supplied with the other model parameters in the segment definition. The primary cause for daily operational PE estimates to be biased compared to the values used in the historical analysis are differences in how solar radiation data are obtained. In addition, an incorrect specification of the anemometer height in the operational definition will also produce a bias. The user needs to make sure that the correct anemometer height is provided as part of the operational station definition and that this value is updated whenever the height is changed.

The only currently available historical daily PE estimates from NHDS are based on solar radiation being derived from manual observations of sky cover. Historical daily PE estimates can be obtained for a selected group of synoptic stations scattered across the country up until the time that each station's ASOS system was commissioned. Once ASOS is commissioned at a station, manual sky cover observations are no longer available. Some automated sky cover values are included in the variables observed by ASOS, but these are not compatible with the method used to estimate solar radiation from manually observed sky cover. In order to estimate solar radiation from manual sky cover observations a parameter, referred to as the B3 parameter, must be supplied. Also a correction factor is needed that adjusts the computed PE values so that they give the same annual lake evaporation as shown in NOAA Technical Report 33 for the location of the station (see section on "Determination of PE" in Section 6-5). These values are automatically provided when historical estimates of PE are produced using the NHDS. If manual sky cover observations were still available at a synoptic station, both the B3 parameter and the appropriate correction factor should be supplied in the PE section of the OFS station definition.

There are other ways of obtaining daily solar radiation data to use for operational daily PE computations. As mentioned in Section 6-5, percent sunshine observations have been shown to produce estimates of solar radiation that will generate unbiased estimates of PE as compared to the Technical Report 33 annual lake evaporation. Thus, operational estimates of PE using percent sunshine to estimate solar radiation should be unbiased as compared to historical estimates of PE available from NHDS. Solar radiation can also be measured directly. Such data are not readily available at many locations. There is also a method of obtaining daily solar radiation from satellite data. Whatever data are used operationally to obtain solar radiation estimates for use in computing daily PE, the user needs to make sure that over the long term the resulting annual PE values are essentially the same as those shown in Technical Report 33. If not, the PE values being used operationally are likely biased compared to those that were used for calibration. As discussed in Chapter 6, biased PE values will result in a bias

s in runoff computations. Such a bias will especially have a significant effect on the soil moisture deficits that are produced after a dry period and thus, can have a major effect on the runoff produced by a subsequent storm event.

Operational Area Definitions

Generally bias in the operational estimates of areal inputs for the models is the result of problems with station data or incorrect information supplied when defining the stations. However, bias can also be produced if the areas are not defined properly. As with station definitions, each data type will be discussed separately.

Mean Areal Precipitation (MAP)

For non mountainous areas the operational MAP time series should be generally unbiased as compared to the historical values as long as the stations involved are checked for consistency and compatibility and the reporting characteristics of the network don't result in biased station estimates, especially on days with spotty convective activity. A slight bias is difficult to completely avoid when there are differences in the composition of the two networks. However, if the station data are properly selected and corrections applied when needed, the resulting bias in the MAP values should be no more than a couple percent as long as the reporting characteristics of the network don't cause a more serious problem.

For mountainous areas, the operational MAP values shouldn't be biased if the appropriate mean monthly precipitation is determined for the stations that were not part of the historical analysis, the stations are checked for consistency when moves or equipment changes occur, network reporting characteristics don't result in biased station estimates on days when precipitation is not occurring over most of the area, and the areas are properly defined. For each MAP area that uses the mountainous area procedure, the exact same average mean areal precipitation that was used for computing historical station weights with Eq. 6-3-4 should be used to compute operational station weights. This can be a seasonal or annual average depending on what was used during the historical analysis. Just as for the historical analysis, Eq. 6-3-4 is used to compute the operational station weights. If the operational network differs from the historical network, the relative weight assigned to each station may be different, but this will not cause a bias. The guidelines for selecting relative weights mentioned in Section 6-3 should also be used when determining relative weights for the operational system.

The amount of random error in the areal precipitation estimates will vary based on differences in the density and distribution of stations in the historical and operational networks. However, if the stations and areas are properly defined and the network reporting characteristics don't create a problem, this shouldn't result in a bias between operational and historical MAP values. As mentioned earlier, if either estimate is only based on a single station, the noise will typically be so great that proper application of the models is not possible. Also, in mountainous areas if there are great discrepancies in the distribution of the gages, especially in terms of elevation, from one network to the other, there may not be a long term bias, but the lik

ely statistical differences in the variability of the estimates will cause problems.

Comparisons should be periodically made between the operational MAP values and what would be obtained using the historical network due to the importance of precipitation estimates and because of all the factors that could cause deviations between operational and historical MAP amounts. This requires that the operational MAP values be archived. Such a comparison can be done anytime that there is a sufficient overlap between the archived operational MAP estimates and the available climatological data needed to compute historical MAP values. An ideal time to perform an MAP comparison is when extending the historical data record (record extension discussed at the end of Chapter 6). If archived operational MAP data are available for the extension period, it would be very easy to compare the values. Besides bias comparisons, both annual and seasonal, other statistics could be computed (e.g. based on precipitation amount, frequency, and variability) to determine what differences exist between the operational and historical estimates. Each time series could also be used to produce simulations of streamflow and other variables to assess how any differences affect model performance. Unfortunately such comparisons have seldom been done in the past.

Mean Areal Temperature (MAT)

For non mountainous areas grid point weights are typically used for both historical and operational MAT computations. For any MAT areas where the station weights differ from the operational to the historical network, areal mean monthly temperatures should be computed and compared. This is done by multiplying the weight for each station by its mean monthly max and min temperature to get the weighted mean monthly max and min temperature for the area. This should be done for all months that the MAT values impact model computations, typically months that can have snow. The weighted mean monthly temperatures should be essentially the same for both networks. If not, either the weights could be modified by using predetermined weights operationally or stations whose mean values deviate significantly from the other stations could be removed from the computations. It is essential that the weighted areal mean temperatures are basically the same for both networks to avoid biased temperature data used in the model computations.

For mountainous areas, the typical procedure is to define a synthetic or “dummy” station that represents the average temperature over the area or the portion of the area that is generally affected by snow. The synthetic station is then assigned a weight of 1.0. When defining mountainous MAT areas in the operational system, it is critical to use the exact same mean monthly temperatures for the synthetic station as were used in the historical analysis. The location of the synthetic station can vary from the historical to the operational system, but not the mean max and min temperatures assigned to the synthetic station. The location of the synthetic station in the operational system should be based on obtaining the best possible estimate of the areal value based on the available real time temperature data. That is, the same logic is used for locating the synthetic station operationally that was used for the historical analysis (see Chapter 6-4).

The historical MAT time series are generated by using only max and min temperature data.

As shown on Figures 6-4-1 and 6-4-2 in Section 6-4 the use of just max/min data can produce erroneous MAT estimates not only on days when the diurnal variation differs from the assumed pattern, but also on other days based on when the max and min values are observed. Even though the problem only occurs on certain days, it can cause the historical MAT data to have an overall bias as compared to what would be computed using both instantaneous and max/min values. The NWSRFS operational system uses both types of temperature data. For stations with an a.m. observation time, this causes the historical estimates to be somewhat greater than the operational values on the affected days. For stations with a p.m. observation time, the result is just the reverse. The overall difference for any area will depend on whether a.m. or p.m. observation times prevail in the historical network. In addition, as mentioned in Section 6-4 the diurnal variation used to obtain 6 hour means from max and min values when computing historical MAT is based on sites in northern Vermont and the central Sierras of California. This assumed typical daily temperature pattern might not be representative of all regions, especially Alaska where the daylight period can be quite a bit longer during melt periods than in the lower 48. It would be very difficult to operationally correct for the bias caused by using only max/min data historically. The real solution to this problem is to modify the historical MAT preprocessor so that it uses the same data, i.e. both instantaneous and max/min, as the operational MAT preprocessor.

Just as with precipitation, comparisons should periodically be made between the operational MAT estimates and those that would be generated from the climatological network to insure that there is minimal bias. Again the ideal time to make such comparisons is when the historical data record is being extended.

Mean Areal Potential Evaporation (MAPE)

The operational MAPE definition should use exactly the same synoptic stations and weights as used for the historical estimates. The average monthly PE values required for the operational definition should be those that were computed by the calibration MAPE program for the historical data period (table of monthly averages displayed at the end of the calibration MAPE program output). Though the calibration MAPE program allows for the use of daily pan data to be used for MAPE computations, this is seldom done. The NWSRFS OFS doesn't include the option of using daily pan observations. If pan data were used historically, any bias between the operational and historical estimates should be minimal if all the computed station PE values are adjusted to conform to the lake evaporation given in NOAA Technical Report 33 (see Section 6-5).

While it would be good to compare archived operational MAPE estimates to values computed from climatological data, this is currently not an option since the data used historically are no longer available with the advent of ASOS. The best that can be done is to verify that the operational estimates of PE at each station are compatible to the annual lake evaporation from report 33.

Forecast Data

Estimates of precipitation, temperature, and evaporation out into the future are clearly important for river forecasting. Any bias in future data will not have a cumulative effect, but it will decrease forecast accuracy. Bias in observed data estimates affect the model state variables (in NWS terminology carryover) that are saved in order to initiate subsequent forecast runs, thus not only affecting the current period, but also later events.

- **Future Precipitation** - Future estimates of MAP are input directly into OFS and are not computed based on any defined information. In the past Quantitative Precipitation Forecasts (QPF) that were used to determine areal estimates of future precipitation for the next few days were generally biased on the high side, especially during convective storm periods. Much work has been done over the past decade to reduce this bias. QPF is now routinely used at most RFCs for short term forecasting. It is still important to verify that the future MAP values being used are unbiased compared to the data used for calibration. In order to properly do this, future MAP should be compared to MAP values generated from climatological data, i.e. the same data used for calibration. Comparisons between future MAP and operational MAP values don't prove that the QPF estimates are unbiased since, as discussed earlier, the operational MAP estimates can be biased.
- **Future Temperature** - Future estimates of MAT are generated in OFS using the station and area definitions that are used for the observed data period. The major difference is that for the future period the computations are based only on forecast max and min temperatures, typically for only a subset of the observational network. Even though this appears to be the same procedure that is used to generate historical MAT estimates, there are some differences which could result in a bias. First, the equations used to compute 6 hour means from max and min values, i.e. represent the typical diurnal temperature pattern, are somewhat different than those used in the historical MAT program. The future MAT equations are also based on Z time rather than local time, thus they produce a slightly different diurnal pattern for each time zone. Second, at least some of the forecast temperatures used by the RFCs are the mid-day and early morning predictions. Thus, the "max" could be lower than the "min" if the temperature was forecast to drop throughout the day. Also, the problem described in Section 6-4 of the wrong max or min being used on days when temperatures are generally increasing or decreasing should not occur with forecast values. Third, the meteorological forecasts may not result in the same long averages used in the historical analysis since they are produced using a different procedure. For all these reasons it is important to compare the future and historical MAT values periodically. It would also be more consistent if all the MAT computations in NWSRFS used the same method. The most accurate estimates should come from a procedure involving the use of both max/min and instantaneous temperature values.
- **Future Evaporation** - Future MAPE values are calculated in OFS from the average monthly values defined for each area. The last computed value for the observed data period is blended over a specified number of days to the average curve. The average curve is determined by linear interpolation between the average monthly values. Thus, if the average monthly values are properly defined, the future MAPE values shouldn't create a simulation bias.

- Data for Extended Predictions - The historical time series of precipitation, temperature, and evaporation are currently used as the input for extended predictions. Near future estimates, as just described in the preceding paragraphs, are typically used for the first few days and then blended into the historical data in an attempt to minimize the effect of events in the historical record that are very unlikely to occur based on the current weather forecast. Since the same input is being used for extended predictions as was used for calibration, there shouldn't be a bias introduced other than that contained in the model states for the start of the run and in the forecast data for the first few days. However, given the meteorological outlook, it could be that the data traces for each of the historical years are not equally likely to occur this year. Methods to include long range weather outlooks in ESP runs have been subjective in the past.

Operational Segment Definitions

All parametric information contained in the active part of the basic segment definitions (i.e. the operations that perform the actual simulation computations as opposed to operations that convert or display data) should be exactly the same in both the operational and calibration systems. This includes the areas that are being modeled and the time interval of all data and model computations. Changes may need to be made to some of this parametric information if the space and time scales used operationally are modified in an attempt to improve forecasts by using new data types and processing methods that were not available historically. Such changes will be discussed in the second part of this chapter. This section will concentrate on the proper definition of the segments to reproduce the results obtained during calibration utilizing similar data with no real time modifications to input values or model states.

Operational segment definitions need to reflect the different data types that are typically used for real time computations, such as river stage observations that need to be converted to discharge with the STAGE-Q operation as opposed to historical streamflow data that has been already converted by the agency responsible for collecting the data. Operational segments will contain direct input of reservoir and diversion data, that were used historically to compute natural flow conditions, when operational forecasts need to reflect what is actually occurring within a river basin.

In other cases, especially in the west where water supply is the major concern, the operational segments will continue to generate natural flow as during calibration. The operational segments will also include different operations to produce and display the output needed for forecasting instead of the WY-PLOT and PLOT-TS displays used to verify simulation results during calibration. However, the bottom line is that the basic segment definitions should conform to the same areas and time intervals that were used for calibration and all the parameters for the models that are performing the basic hydrologic computations should be exactly the same as those determined during the calibration.

When first defining a segment, initial values of the state variables need to be provided for many of the models. In many cases, such as for channel response, routing, and reservoir models, these values can easily be estimated from current river and pool levels or default values can be used if

or operations where the values will only affect the computations for a short period. For the snow model, the initial values of the model states can be defaulted to zero when no snow cover exists. If a significant amount of snow is present when the model is initiated, then it is critical to input a realistic estimate of the current mean areal water equivalent. If a greater amount of snow existed earlier in the season, then the maximum water equivalent since snow began to accumulate also needs to be provided. Both of these values can be obtained from available snow measurements as described under the .WECHNG MOD in Section 8-1. The heat deficit, antecedent temperature index, and liquid water content are estimated based on the temperature pattern in the recent past. For the Sacramento model, estimates of the lower zone free water contents should be able to be derived from the hydrograph by determining the current supplemental and primary flow contributions. These runoff amounts are then divided by the appropriate withdrawal rates to get the current contents. Upper zone free water contents are typically set to zero since the effect of this variable will only last for a few days. Tension water contents can be reasonably estimated by comparing the general precipitation pattern over the last few months with a similar period from the calibration simulation. The initial value of ADIMC is set to the sum of the upper and lower zone tension water contents.

Real Time Forecasting using Calibration Results

Introduction

The major objective during forecasting is to utilize all available information so that simulations closely mimic observations and forecasts have maximum lead time with minimal uncertainty. Random fluctuations that were tolerable during calibration, in that they didn't drastically interfere with the determination of the proper values of the model parameters, now need to be reduced to a minimum to provide the best possible operational forecasts. There are several methods for improving real time forecasts over what would be obtained by just using the simulated results generated from the basic OFS station, area, and segment definitions (i.e. the definitions just described in this chapter utilizing the same types of station data and modeling schemes used for the historical analysis and calibration).

- Use data analysis procedures that incorporate the dynamics of what is occurring, both by new measuring techniques and physical modeling, to improve the input to the hydrologic models over that provided by the historical techniques which often rely on climatological patterns. Such procedures are capable of providing better spatial and temporal resolution as well as improved areal estimates.
- Apply the models in a distributed manner at a finer spatial and temporal resolution than was possible during calibration by utilizing the output from the new data analysis procedures.

- Adjust the model states and computations, either automatically or manually, based on real time observations or the output from external modeling systems.

Besides these methods for improving river forecasts which will be discussed in more detail in this chapter, forecasts can also be improved for many users by quantifying the uncertainty in the predicted values. This has been done in the past for extended predictions. In the future probabilistic forecasts will be possible for all lead times. In order to make accurate probability statements for short term forecasts, the uncertainty in the models themselves and the model states must be taken into account, in addition to the uncertainty in future precipitation, temperature, and evaporation. The ensembles used to generate probabilistic predictions will need to incorporate the trends and uncertainty in meteorological forecasts from short term QPF to long range outlooks. Whatever is done in terms of producing the ensembles, it is important that the statistical properties of the historical data and the historical model simulations based on the calibrations be preserved, including subtle features like the amount of snow versus rain and the frequency of surface runoff.

Before discussing each of the listed methods for improving forecasts a couple of general items should be noted. First, in most cases there is not a one-to-one relationship between new data estimates and the historical values used for calibration or between observations or values computed by external modeling systems and values produced by conceptual models. Thus in order to avoid any bias being introduced, the relationship between the estimates must be determined and corrections applied when necessary. Second, most new data analysis or hydrologic modeling systems evolve over some period of time. Even though a new method is tested thoroughly during the development phase, further modifications to the algorithms and data processing techniques are frequently needed when the procedure is put into operational use. Thus, the relationship between values produced by the new method and the historical calibration may change for some period of time before becoming stable.

The intention of the following discussion of each of the methods for improving forecasts is to indicate how such methods can be applied in a manner that preserves the benefits of calibration. The emphasis is on how to use such procedures without creating a bias compared to the calibration simulations. It is not the intention of this manual to describe these methods in any detail or to offer guidelines for their use other than how they can be applied in an unbiased manner.

Improved Data Analysis Procedures

The only procedure that is currently available operationally for improving model input utilizes radar, raingage, satellite, and other data to produce hourly estimates of precipitation on a 4x4 km grid. This procedure is currently referred to as the Multisensor Precipitation Estimator (MPE) [Soro et al., 2000]. Since the MPE procedure generates estimates with a finer spatial and temporal resolution than generally can be obtained from just raingage data, it opens up the possibility of not only providing improved precipitation values for use by lumped applications of hydrologic models, but of allowing for the real time use of distributed models for river forecasting.

Other possible real time data analysis procedures that might be developed in the future to improve

e operational input include:

- mountainous area precipitation analyses that utilize some combination of orographic modeling, gage reports, radar information, and other data to account for the storm dynamics,
- mountainous area temperature analyses that utilize a combination of meteorological modeling, temperature observations, upper air soundings, and other data to account for the dynamics of how temperature varies with elevation, and
- direct estimates of ET-Demand that account for vegetation dynamics on evaporation rates.

Experience with radar-based estimates of precipitation have indicated that in many cases the estimates are biased and not consistent over time [Johnson *et al.*, 1999]. This is partly due to changes that have been made to the algorithms and data processing techniques. There have especially been difficulties with obtaining reasonable precipitation estimates from radar-based methods in mountainous areas and when snow is occurring. Thus, before using these data as either lumped or distributed input to the hydrologic models for real time applications, it is essential that the estimates are carefully evaluated in a systematic manner. This evaluation should determine whether radar-based precipitation estimates can be used to produce unbiased simulation results and if so, how they can be applied in order to improve operational forecasts. The application could be to provide lumped 6 hour values or lumped hourly estimates of precipitation. Beyond that radar-based precipitation estimates could allow the models to be applied in a distributed manner.

One possible procedure of evaluating whether improvements are possible by using radar-based precipitation estimates is described in this subsection. In the discussion of the procedure 3 types of mean areal precipitation time series are utilized. These are referred to as:

- MAPX - areal mean precipitation produced from the radar-based estimates by averaging the grid points that fall within the area being modeled.
- MAPH - areal values produced from climatological data, i.e. values generated using the same precipitation gage data and methods as used for the historical analysis.
- MAPO - areal values produced using real time precipitation reports with the OFS MAP preprocessor function based on the station and area definitions discussed previously in this chapter.

The steps in the procedure for a given watershed are as follows.

1. Calibrate the watershed following the methods described in this manual. This will insure that a sufficiently long record is used so that a stable set of parameter values can be obtained considering the noise in the input data.
2. Evaluate the archived radar-based estimates of precipitation for consistency and potential

bias. There probably needs to be in the order of 5 years of reasonably stable radar based estimates to obtain a good evaluation. One way to do this is to compare the monthly MAPX values to overlapping MAPH estimates. The pattern of the ratio of MAPX to MAPH values should be consistent over time. There will be scatter since each method will compute different values for each storm. Trends in the ratio with time generally would indicate changes in how the radar based values were processed. Ideally the average ratio should be reasonably close to 1.0. For a non mountainous area this would indicate that the radar based procedure is producing values over the long term at each point that are the same as would be measured by a rain gauge. However, it is more critical that the ratio is reasonably consistent over time than being equal to 1.0, though typically if there is a significant bias, there may also be a lot of variability associated with the radar-based estimates.

This would also be a good time to compare archived MAPO values with overlapping MAPH estimates. If the station and area definitions for OFS are compatible with those used for the historical analysis, the monthly and annual ratios of these time series should be very close to 1.0. If not, it indicates that either the OFS station or area information is not correct or something else, such as different reporting characteristics, causes these time series to differ.

3. If the evaluation in step 2 looks reasonable, then generate 6 hour MAPX time series (assuming that was the time interval used for calibration) for the watershed and use these as input to the hydrologic models. Compare the results to those obtained using the MAPH estimates as model input. If the long term ratio of MAPX to MAPH is 1.0, the overall simulation bias should be similar from both time series. If the ratio is not 1.0, then the MAPX values need to be adjusted, typically by a multiplying factor (PXADJ in either the SNOW-17 or SAC-SMA operation). It is preferable at this point to use PXADJ to correct for any bias as opposed to modifying model parameter values because the simulation bias is caused by differences in the MAPX and MAPH amounts. If the MAPX estimates are not consistent over time, this will result in the model simulations using those data to have more of a time trend than those produced by the MAPH time series.

Once the MAPX values are adjusted so that the overall bias is the same for both precipitation time series, an evaluation can be made of the results. Ideally the MAPX values will contain less random error and thus provide an improved simulation capability. If so, it indicates that at least the radar based precipitation can be used to provide a lumped 6 hour input for operational forecasting that should improve results. If the comparison of the simulation results indicates that there is a pattern, perhaps seasonal, in terms of which MAP produces the best results, then rules can be devised for when to use the radar based values operationally and when to use MAPO estimates (assumes that the MAPO values are compatible with the MAPH estimates). Such a situation will likely exist in regions where snow predominates during the winter months. In this case the MAPX values may provide improved input during warm months, but the MAPO estimates would produce better results during the cold season (the MERGE-TS operation is used in OFS to switch between alternative time series). The MAPX and MAPH time series for the test period could be merged based on the proposed rules to determine how this should affect the operational simulation results.

4. If the use of MAPX time series indicates an improvement in step 3, then generate 1 hour MAPX time series, adjust model parameters for the change in the time scale if necessary, and evaluate the results. Models that are non linear, such as the Sacramento model, are scale dependent, i.e. the results can vary depending on the spatial and temporal scale that the model is applied [Finnerty *et al.*, 1997]. Thus, in order to get similar results at different scales, some of the parameter values will likely need to be altered.

Hourly data should have the most chance of improving simulation results for fairly fast responding watersheds that have significant amounts of surface runoff and variations in rainfall intensity during storm events. For these watersheds instantaneous streamflow data, preferably hourly, is needed in order to make the parameter adjustments and to properly evaluate the results. The PXADJ value determined in step 3 should continue to be used to correct for the overall bias in the MAPX time series. The most likely Sacramento model parameter modifications are to the percolation rate and UZFWM. The percolation curve would typically need to be adjusted slightly upwards. This should be done by changing the LZFSM and LZFPM parameters by the same percentage. This will change the PBASE value, but retain the shape of the curve and not alter baseflow recession rates. UZFWM should need to be increased somewhat. The unit hydrograph would also need to be converted to a one hour runoff duration. It is also likely that there would be a need to modify the shape of the unit hydrograph somewhat based on the finer temporal resolution of the computations and the observed flow data.

If doing the model computations on a hourly basis improves the simulation results, it should primarily be noticeable in the simulation of storm peaks. This would especially be the case for major surface runoff events when high intensity rainfall occurs for just a few hours within the 6 hour intervals. Watersheds that don't generate much or any surface runoff or have fairly uniform rainfall rates during major storms, shouldn't show any real improvement in simulation results by using hourly precipitation input. During seasons when snowmelt runoff predominates, it is unlikely that using hourly precipitation input will improve the results. The melt rates, especially when using a temperature index model, will not vary significantly from hour to hour and the snow cover will have a damping effect during rain-on-snow periods.

By going through such an evaluation for several watersheds scattered throughout an RFC area, it could be determine whether the radar-based estimates would provide unbiased values of precipitation for use with the calibrated conceptual models. A determination could be made of whether radar-based estimates could be used all the time or only for certain seasons or types of storms. It could also be determined whether adjustments were needed to the model parameters in order to avoid a simulation bias, especially when a change is made to the temporal scale of the model computations. Radar-based estimates of precipitation clearly have the potential of improving this critical model input, at least in non mountainous areas. However, before they are used operationally it should be verified that a simulation bias is not be introduced.

Distributed Modeling

The Hydrology Laboratory of NWS has had a major research effort for a number of years involving the use of distributed models for river forecasting [Smith *et al.*, 1999]. This effort is aimed at evaluating the data needed for such an application, examining various options for how models can be applied in a distributed manner, investigating the appropriate models to use, and ultimately to come up with a set of guidelines for how to use distributed models and under what circumstances they are needed and should produce improved forecast products.

The use of hydrologic models in a distributed mode offers the potential for improvements in simulating the response from basins with spatially varying precipitation patterns and allows for creating forecast products at a finer spatial resolution. Forecasts could not only be generated at the watershed outlet, but possibly also at interior points. This would allow for significant improvements in specifying the locations of possible flash flooding. The greatest potential for benefits from distributed modeling is in basins where runoff typically only occurs over a portion of the drainage area and in areas where the location of intense rainfall and thus fast response runoff varies from event to event. In such watersheds the magnitude and timing of the hydrograph response is highly dependent on where the runoff is being produced. Distributed modeling should allow for the successful simulation of streamflow in regions where lumped applications provide unsatisfactory or marginal results (see Figure 1-1). Even in regions where lumped applications generally provide reasonable results, distributed models could improve the forecasts for certain events. The most improvement should be noticed for events with considerable surface runoff concentrated over only a portion of the drainage. Elongated watersheds are more likely to have a different hydrograph response in such cases than a more oval basin. The more dampening of the response either due to infiltration and percolation, snow cover, or the channel system, the less likely that a distributed application of the models will result in improved simulation results.

In order to apply models in a distributed mode it is essential that quality precipitation estimates are available at the appropriate spatial and temporal scale. This is why the validity of gridded precipitation estimates, such as those from radar-based procedures, must first be verified before attempting to use the models in a distributed manner. In addition to gridded estimates of precipitation, some of the rainfall-runoff model parameters will have to be modified when going to a finer spatial resolution. This is because of the scale dependency of many such models, including the Sacramento model [Finnerty *et al.*, 1997]. These factors imply that some period of stable gridded precipitation estimates need to be available for testing the distributed application of the model and to determine what parameter changes are needed to insure that the results are unbiased compared to the lumped calibration. This could be done by applying the model in a distributed mode for a period with stable precipitation estimates and comparing the simulation to the lumped calibration results for the same period (calibration simulation would most likely use the areal average of the gridded precipitation, but could use MAP computed from historical gage data). Parameters should be adjusted so that the overall bias of both simulations are the same. Also such a comparison would allow for an evaluation of the potential benefits of distributed modeling.

For watersheds in regions where the lumped application of conceptual models generally yields unsatisfactory results (see Figure 1-1), the comparison and parameter adjustment procedure just d

described is not really relevant. In such regions it is impossible to obtain appropriate model parameter values by applying the models in a lumped mode and the climatological records do not contain a sufficient gage density to attempt a distributed calibration. In such regions reasonably reliable calibrations and operational application are only possible after a sufficiently long record of stable radar-based precipitation values have been archived. These data would then be used in an attempt to calibrate the models in a distributed mode.

Distributed modeling could also be attempted for snow accumulation and ablation. Most likely this would be done by using an energy budget approach for computing the snow cover energy exchange rather than by trying to figure out how to realistically vary melt factors from one subarea to another. There are potential difficulties in determining how precipitation and the energy balance input vary in mountainous areas as a function of elevation, aspect, slope, and vegetation cover. Even though a energy budget based distributed snow model will hopefully remove the trend of under simulating major melt events, a comparison of distributed model simulation results with a lumped application of a temperature index model should be able to determine if the distributed results are reasonably unbiased and provide an overall improvement compared to the historical calibration.

It also should be noted that changes to the spatial and temporal scales that the models are applied operationally has an impact on ensemble streamflow prediction (ESP) techniques used to generate probabilistic forecasts. Short term ESP applications will most likely use ensembles that are generated based on uncertainties in QPF and other meteorological forecasts. These ensembles should be able to be generated for the same spatial and temporal scale as the models are being applied operationally. However, for extended ESP applications the ensembles have been the historical precipitation, temperature, and evaporation time series that were produced during the historical data analysis step of the calibration process. These time series are at the spatial and temporal scale that was used for model calibration. Thus, in order to make extended ESP runs, the models must, at least currently, be defined operationally at these same scales. It would be quite difficult for an RFC to maintain multiple sets of compatible operational segment definitions at different spatial and temporal scales for short term forecasting and extended predictions. Thus, until some solution to this dilemma is developed, basins for which extended ESP applications are important will need to be defined operationally at the same spatial and temporal scale as used during calibration.

Adjustment Procedures

One of the most common methods for reducing the variability of the operational simulations is to adjust the input data, model computations, or model state variables based on a variety of observations or the output from an external modeling system. Adjustments can be made manually by the forecaster, generated by an external analysis of relevant data, or through the use of an automated procedure. The aim in all cases is to reduce variability and remove trends so that the models can be used to generate an unbiased forecast with minimal uncertainty. It is important when applying such adjustments that a bias is not introduced as compared to the simulation that would result if no adjustments were applied, other than the removal of some of the modeling deficiencies

s mentioned in Chapter 7.

Automated updating procedures include Kalman filter based procedures such as that utilized by the SS-SAC operation in the OFS [Georgakakos *et al.*, 1988] and a variational assimilation approach [Seo *et al.*, 2002]. The application of these automated procedures have thus far been limited to non-snow, lumped, single area headwater basins. The procedures use streamflow observations to adjust precipitation and evaporation inputs to the Sacramento model and state variables for the Sacramento and channel response models. The procedures are primarily intended to make adjustments during runoff events. The SS-SAC updating procedure uses a state-space version of the Sacramento model. Since this version doesn't exactly duplicate the regular Sacramento model computations, there is some potential for bias to be introduced. Both of these procedures have primarily been tested by comparing short lead time forecasts using the updating procedure to forecasts produce using a simple blending or persistence technique. In evaluating such procedures it would be important to determine which variables are being altered and by how much. It is possible that state variables could be changed to physically unrealistic values based on current streamflow observations that would not impact a short lead time forecast, but could produce a substantial bias for a longer lead time or especially for an extended prediction.

The primary method for making real time adjustments for all types of watersheds and modeling configurations with the OFS has been the use of run time modifications (MODs). The MODs in the OFS allow the user to modify time series values, model state variables, model parameters, and rating curves. Input data can be changed, as well as time series generated by any of the operations in the segment definitions. The user can specify where in the operations table the changes are made and which values are modified. Model state variables can be altered at the beginning of a run or at a specified time within the run. MODs that alter model parameters are primarily used to adjust the computations on a temporary basis for such things as abnormal snowmelt rates or particular precipitation patterns that affect the channel response. Rating curve modifications are intended for locations with constantly changing ratings.

MODs can be specified by the forecaster via the Interactive Forecast Program (IFP) in NWSRFS at run time or they can be generated externally. When making MODs with IFP, the forecaster should not only rely on deviations between observed and simulated streamflow, but on a variety of other information such as: radar data, meteorological data and analyses including forecasts, observer reports, snow cover information, recent climatic trends (above or below normal conditions), satellite measurements, and river ice reports. MOD images are generated externally when an objective analysis procedure is used to compute the adjustments (e.g. procedures are available to compute snow water equivalent estimates based on snow course or areal gamma flight line measurements).

The main concern from the point of view of this manual is that MODs are not applied in a manner that will result in a biased forecast compared to what would be generated without any modifications. For example, it is difficult to determine the cause of any deviation between simulated and observed values and the magnitude of any needed adjustments early in a storm event. After the event is over, it is much easier to decide the cause (e.g. a volume problem can clearly be distin

guished from a timing error after the event, however, early in the event the hydrograph response may look the same in both cases). The sooner the proper adjustment can be made, the longer the forecast lead time. However, if the wrong adjustment is made, it will typically result in the need to make further modifications which will alter the forecast and decrease the time available for action. In addition, the wrong adjustment might cause the state variables to take on values that could result in a considerable bias during the subsequent recession period or the next event and could especially impact extended predictions.

The chance of making a modification that could bias subsequent forecasts is even greater when there are several sources of error. For example, during the widespread flooding in the northeastern United States in late January 1996 [*Office of Hydrology, 1998*], the meteorological situation clearly indicated extreme snowmelt conditions, but there was also evidence of an underestimation of precipitation due to pronounced orographic effects. In addition, there was uncertainty in the initial snow cover and soil moisture conditions. If all the error was attributed to a single factor, it would be quite likely that some of the model states would not be correct and would result in subsequent forecasts being biased. In such cases it would be extremely valuable for the forecaster to have access to tools that would provide insight into potential errors and estimate a reasonable magnitude for each likely problem.

Besides storm events, there are other times when MODs are used to modify model states. During low flow conditions adjustments can be made to baseflow storages based on recent deviations between simulated and observed streamflow. Prior to the snowmelt season, model estimates of areal water equivalent are modified based on measurements of the snow cover. Soil moisture tension water deficits could be altered based on soil moisture measurements or model response after a dry period. Such updates can not only have an effect on the next significant runoff event, but more importantly they will alter the results of any extended predictions and can produce a substantial bias if not applied properly.

If the models being used for forecasting are well calibrated and properly applied operationally, two things should occur in regard to real time adjustments. First, there should be less need to make updates as the model simulations should more closely track observed conditions. Except for trends caused by model limitations or the spatial scale of the application (see Chapter 7), the errors from a well calibrated and properly applied model should be random. Models that are poorly calibrated or improperly applied will produce biased results and thus will need to be adjusted more frequently. Second, the proper adjustment to make and its magnitude should be able to be based on a logical assessment of potential sources of error. For example, a scientific evaluation of rainfall volume or snowmelt rates should be able to be used directly to determine if adjustments to these values are needed and the appropriate magnitude. Whereas, with a poorly calibrated or improperly applied model, there may be no logical explanation for the adjustments needed to cause the simulated conditions to reasonably match observations.

Section 8-1 provides some guidance for using the MODs directly associated with the SNOW-17 and SAC-SMA operations in a manner that reduces the chance of generating biased forecasts. General recommendations for using MODs are more in line with an operational forecasting, rather

er than a calibration manual, and are thus not included.

Operational Implementation Summary

If the objectives are met, i.e. unbiased simulation with minimal random error, parameters that properly reflect the function of each component of the models, and spatially consistent parameter values, then calibration results should significantly improve operational forecasts. The models should more closely track observations, fewer real time adjustments should be needed, and the choice of which modifications to apply and their magnitude can be based on logical decisions using all available information. In addition, besides producing unbiased short term forecasts with minimal uncertainty, the models can be used to generate accurate extended probabilistic predictions. However, all of these benefits are only possible if the calibration results are properly implemented in the operational environment. This requires that the results produced with the operational parametric definitions must be unbiased compared to the calibration results. It also requires that alternative sources of model input, changes to the spatial and temporal scales that the models are applied, and real time adjustments be applied in a manner that while reducing random error and model deficiencies doesn't generate biased results.

While the guidelines and recommendations offered in this chapter should produce forecasts that contain the benefits of a proper calibration, the user needs to periodically verify the results and carefully evaluate alternative inputs and methods of applying the models. Comparisons and consistency checks should be made every few years to verify that real time data are indeed generating operational input and simulation results that are unbiased compared to the historical input and calibration output. New data analysis procedures, changes in the spatial and temporal scale that the models are applied, and real time adjustments need to be carefully evaluated to make sure that their use results in not only unbiased simulations, but produce real forecast improvements.