Chapter 5

Step 3 – Select Flow Points and Period of Record

Introduction

The next step in the calibration of a river basin is to select the locations at which streamflow will be modeled and the period of record to use. The locations to model streamflow are primarily selected based on the location of historical daily flow data, the location of forecast points, the real time operational data network, the spatial variability of hydrologic conditions, and the location of control structures. During calibration streamflow may need to be modeled at more locations than just forecast points in order to properly handle the variation in hydrologic response and the effect of reservoirs, diversions, etc..

There are two periods of record involved in the calibration process. The first is the period to use when processing the historical input data. This period is dependant on the length of the available historical precipitation, temperature, and evaporation records, the consistency of the historical network over time, and the type of forecast products to be generated. The second is the period of record to use when calibrating the models. This period is dependant on the length of the historical streamflow records, the variability and noise in the streamflow data, and changes in vegetation and land use.

It is important to carefully assess all these factors and make a deliberate decision regarding which locations to model and the periods of record to use before processing the data and beginning to calibrate the models. Making these decisions at this point will save a considerable amount of time and effort in the long run.

Selection of Points to be Modeled

The following steps can be followed to determine which river locations should be modeled during the calibration of a river basin.

1. All forecast points with historical streamflow data should be modeled. This includes both existing forecast points and new or potential forecast points. In the case of reservoirs and lakes where inflow or water level is to be forecast, historical pool elevations and possibly outflows will be needed. This is an excellent time to review forecast needs in the river basin and, with input from WFOs and users, decide if all existing forecast points are still needed, whether there are new locations for which forecasts should be provided, and try to anticipate future needs. This includes the consideration of all types of forecasts, including floods, low flow, water supply, recreation, navigation, and reservoir operations. It also includes both short term and extended forecasts. If there are forecast locations that don't have any historical data, then typically a procedure is later developed for operational use that takes the results from a location with historical data and extrapolates those results to the forecast point.
2. River locations that have real time data available and a historical record should in many ca

ses be modeled even if they are not forecast points. Such locations can be valuable in that operationally computations can be verified and adjustments made to improve the lead time and accuracy of forecasts at downstream points. However, if the drainage area reflected at such a location is small and especially when data are sparse, the combination of streamflow measurement and data input errors may make it very difficult to make realistic adjustments. Errors need to be averaged over a sufficiently large area in order to make real time adjustments that will improve the forecast results at a downstream location.

3. Locations at the outlet of reservoirs or lakes that have a significant effect on movement of water through the river system and locations of diversions into or out of the stream network need to be considered for modeling during calibration. Typically there are two options at such locations. First, the control structure can be modeled. In this case historical data such as pool elevation, storage, outflows, or diverted quantities are needed so that a comparison can be made between simulated and observed values. Second, the outflow from a reservoir or lake or the amount of diverted water can be obtained from the agency that operates the structure. Operationally this must include actual, as well as forecast values, so that the operation of the structure will not need to be modeled. During calibration the historical streamflow record will need to be modified to reflect natural flow conditions (described in Section 6-6).

4. Other river locations with historical streamflow data also need to be considered for modeling during calibration based on the variability of hydrologic conditions over the river basin and the available data networks. These gages may be upstream from a headwater forecast or real time data point or they maybe at intermediate locations within the river system. If these gages are at locations where there is likely a significant variation in hydrologic response based on hydrograph comparisons or other information such as soils or vegetation data, then these gages should be considered as possible sites for calibration. Also if the hydrograph response for the area above or between previously determined calibration locations can vary considerably based on the spatial variability of rainfall, then these additional sites may need to be included during calibration. Another factor to consider is whether there is sufficient historical and real time input data, primarily precipitation, to adequately define the areal input if the drainages are subdivided in order to model these additional locations.

After considering all these factors the flow points to be modeled as part of the calibration process can be selected. In many cases, especially in areas with significant variability in hydrologic conditions or rainfall amounts, calibration simulations and comparisons with observed values will be produced at more locations than just forecast points. The inclusion of these extra locations should improve the simulation results at downstream locations and result in more accurate and timely operational forecasts.

Selection of Periods of Record

As mentioned in the introduction for this chapter there are two periods of record that need to be determined. First, is the period to generate input time series such as precipitation, temperature, a

nd evaporation. This is referred to as the historical data period. Second, is the period used to determine the parameters for the hydrologic models. This period is referred to as the calibratio n period and is typically a subset of the historical data period.

1. Historical Data Period of Record – This period depends on the length of available historica l data, the consistency of the network over time, and the type of forecast products to be gener ated. An appropriate period is based on the influence of each of these factors.

a. Generally in the United States digital records of climatic variables are only available st arting in 1948. This is when NCDC started to put the data for all climatic stations in dig ital form. Records prior to that time have been added to the digital archives when additi onal funding was available, however, in most parts of the country these earlier records ar e only for selected locations and do not result in a network that is dense enough for hydro logic modeling.

b. While in most parts of the country the density of the historical network and the represe ntativeness of gage locations are fairly similar over time there are some regions where sig nificant changes occur in network coverage. A prime example is in the western states w ith the advent of the SNOTEL network maintained by the NRCS. The SNOTEL networ k consists primarily of high elevation sites where a sizeable snow cover can exist. In so me portions of the west there were a reasonable number of high elevation sites prior to th e advent of the SNOTEL network, but in other parts of the region there were few, if any, high elevation locations during the earlier years. Most of the runoff in much of the west is generated at the higher elevations. It is generally impossible to produce realistic tim e series of precipitation at high elevations based only on data from low elevation sites. One can adjust for the annual or seasonal difference based on an isohyetal analysis, but th e time distribution and frequency can't be accounted for properly. Much of the precipit ation at high elevation locations occurs when no precipitation is falling at lower elevation s.

When there are significant changes in the network during the period of available historica l data, it may be that only a portion of the data can be used to generate input time series f or use in calibration. This is certainly the case in the west for basins where there were li ttle or no high elevation data sites prior to the SNOTEL network. In these basins only t he period of record after the advent of the SNOTEL network should be used to generate i nput time series. In other areas of the west there are a reasonable number of high elevati on sites prior to SNOTEL. In some of these areas the entire period of available data can be used to generate time series that have similar properties. In other areas there may b e statistical differences in the time series produced with data before and after SNOTEL su ch that parameter determination and simulation results will differ significantly from one p eriod to the other. In such cases it may be necessary to do some comparisons at selected locations to determine if the different periods of record are compatible. This can be do ne by generating input time series for the entire period of available data and then calibrati ng the models using the period with the best data coverage, typically the period with SN

OTEL data. After this is done, the parameters determined from this period would be used to simulate the earlier period and a comparison made of the simulation results. If the results are similar for both periods then the entire length of available data can be used to generate input time series, whereas if there are significant differences in results then the two periods will not produce compatible time series and only the later period should be used during calibration.

c. If extended probabilistic forecasts are to be generated anywhere in the river basin or at locations further downstream using flows routed from the current basin, then the longest possible period of record that will produce consistent values should be used to generate input time series. In this case the time series will not only be used for determining model parameters, but will be used as inputs to generate possible streamflow traces with the ESP procedure. The more traces available to statistically analyze after an ESP run, the higher the confidence in the results. Extended probabilistic predictions produced using ESP are primarily of value in river basins where the current state of the system can have a significant effect on what happens in the future.

On the other hand, if extended applications of the ESP procedure will never be used in the river basin or at downstream locations, then only a period sufficient to determine model parameters needs to be used to generate input time series. The considerations in determining this period are discussed in the next section. Due to differences in periods of streamflow data at different locations and the effects of vegetation and land use changes on the period used to calibrate the models, the calibration period for one location within the river basin may not be the same as for others. The period used to generate input time series should overlap the calibration periods to be used for all locations.

2. <u>Calibration Period of Record</u> – This period is based on the length of historical streamflow records, the variability and noise in the streamflow data, and vegetation and land use changes. An appropriate period is selected for each point to be modeled after considering the influence of each of these factors.

a. At a given location within the regions where lumped, conceptual models produce satisfactory results (see Figure 1-1), typically about 10 years of streamflow data are adequate for determining model parameters. This length of record is needed so that one can have reasonable confidence that the noise in the data signals is random. The shorter the period of record the greater the chance that the data noise will not be random, thus resulting in biased parameter values. Parameter sets based on only a few years of data will typically vary from one time period to another. For these regions a calibration period of about 10 years should result in fairly stable parameter values. In areas where the model results are generally marginal, a longer period of record likely is needed in order to have enough events and to minimize the noise caused by the typical large spatial variability of rainfall that occurs in these regions. This longer period is needed so that at least a reasonable degree of confidence can be assigned to the model parameter values. In regions where model results are typically unsatisfactory, it is unlikely that any length of record will be s

ufficient to determine parameter values with any significant degree of confidence. For t he initial headwater basin to be calibrated (see chapter 7 for more details), it is recommen ded that besides the period used to calibrate the model parameters, that another period be available to verify the calibration results. This period should ideally be at least as long a s the calibration period.

b. The period of record selected for model calibration should contain a variety of hydrolo gic conditions. There needs to be wet periods with high flows, as well as dry periods wi th low flows. If snow is a factor, there should be years with below normal snow cover a nd years with above normal cover, as well as near normal years. In regions where clima tic indices are proven to have a significant effect, the calibration period should include a f ull range of values of the index (e.g. if the Pacific Oscillation is important, years with a st rong El Nino effect, as well as years when La Nina dominates should be included). At l east for the initial headwater calibration, it may be a good idea to exclude the flood of rec ord and the lowest flow period from the calibration and use these events to verify the extr apolation capabilities of the calibration.

c. In areas where physiographic factors are essentially constant over the period used to ge nerate the input time series, any portion of the period that contains sufficient variability i n hydrologic conditions can be used to calibrate the models. However, in areas where t here are significant changes in vegetation or land use, the portion of the period of record t hat should be used to calibrate the models is that which most closely reflects current cond itions. This is typically the most recent period though, for example, in the case where a watershed has just recovered from the effects of a large forest fire, the period prior to the fire would better reflect the current state of the drainage. In agricultural areas that have seen large changes in crop type or farming practices or for watersheds near metropolitan areas that have undergone considerable suburbanization, the most recent period of record should be used to calibrate the models.

In many river basins the periods of historical streamflow data vary from one location to the n ext. Also vegetation or land use changes may only affect certain portions of the basin. Th e aim is to try to find an appropriate period to calibrate the models that is as common to as m any flow points as possible. There are a couple of reasons for this. First, when calibrating downstream locations it is best to have observed data at the immediate upstream sites so tha t errors in the upstream calibrations are not propagated downstream. Second, having observ ed flows at each point within the basin during the period used will allow for a comparison wi th total simulated flows to mimic forecast and extended ESP conditions.