Section 6-2

Consistency Checks for Precipitation, Temperature, and Evaporation

<u>Basic Method</u>

Consistency of precipitation, temperature, and evaporation data from individual stations is checked using a double mass analysis. NWSRFS contains consistency check options within each of the programs that are available for processing these data types, i.e. the PXPP and MAP programs for precipitation, the MAT program for temperature, and the MAPE program for evaporation. In addition a graphical user interface program, IDMA (Interactive Double Mass Analysis), is available to display the consistency plots generated by the processing programs and to interactively make consistency corrections. When using these processing programs the first step is to check the consistency of the data and make any adjustments needed prior to computing mean areal values for use in model calibration.

When using a double mass analysis technique it is essential to include enough stations so that inconsistencies in individual stations can't have a significant effect on the consistency of the computed group average. Typically when performing the data analysis for an entire river basin or on a regional basis, there are more than enough stations, however, in case the technique is being applied to a smaller network it is recommended that at least a minimum of 10 stations be included.

For precipitation and evaporation the consistency checks involve plotting the data for one station against the average of the data values from a large group of stations. This is because the relationship between precipitation and evaporation stations is normally expressed as a ratio, i.e. it is said that one station typically has a certain fraction of the value observed at another station, the average of a group of stations, or an areal averaged value. The accumulated value for the station is plotted against the accumulated value of the group average. In NWSRFS in order to more clearly see changes in the relationship between one station and the group, the deviation of the accumulated station average from the accumulated group average is plotted against the accumulation of the group average. The group average is computed based on all the stations in the group minus the station being plotted against the group. Thus, the accumulated group average generally changes slightly from one station to another and therefore there is not a direct relationship between the accumulated group average and a specific date in time.

For temperature the consistency check involves plotting the deviation between the accumulated average temperature for a single station and the accumulated average temperature for a large group of stations against time. This is because the relationship between temperature stations is normally expressed as a difference, i.e. it is said that one location is a certain amount colder or warmer than other locations on the average. Separate plots are generated for maximum and minimum temperatures.

For all these data types the record for the individual station is consistent over time if the relationship between it and the group plots generally as a straight line. However, there is a lot of variabi lity in nature such that the relationship between an single station and a group of stations will not be perfectly straight. These random variations in the relationship will cause the plotted lines for each station to wobble. As long as the general trend of line is straight and there are not significant deviations from the general trend line, the data are consistent. Inconsistencies are indicated by changes in the slope of the line that are of a significant magnitude and last for a reasonably long period of time to be caused by something other than the natural variability in the data fields.

The trick is to be able to recognize inconsistencies and adjust the appropriate data periods without removing the natural variations that exist in the data. Precipitation data typically are much more spatially variable than temperature and evaporation, thus consistency plots for precipitation generally exhibit more wobble than plots for the other variables.

NWSRFS Program Options

The NWSRFS programs with consistency checks contain options for specifying the number of groups, the stations assigned to each group, the stations to be included on each plot, and whether to do precipitation checks on a seasonal basis.

Generally all the available stations within the river basin are included in a single group. In this case each station is then plotted against the average for all the other stations. NWSRFS does contain the option to have multiple groups. It was initially thought that it would be better in mountainous areas not to put all the stations in one group, but instead group the data by some factor such as elevation or mean annual value. This feature still exists, but experience has shown that typically nothing is really gained by trying to use multiple groups when making consistency checks. Also, generally all the stations in the group are used to compute the group average even though the option exists to remove some stations from this calculation (option not in PXPP). This option is seldom needed since normally there are enough stations involved in the group that the effect of any one station is minimal. If the data for a station are so questionable that it should not be part of the group average, the station probably shouldn’t be included in the analysis in the first place.

The consistency plots generated by the NWSRFS data processing programs display a maximum of 5 stations per plot. When a large number of stations are being included for a river basin, it is very important as to how the stations are grouped for plotting, especially in the case of precipitation. As indicated earlier, inconsistencies typically show as a change in the slope of the double mass plot that persists for a reasonably long period of time. For precipitation data such changes in slope can also occur due to shifts in storm tracks or storm types over an area like a river basin and can persist for a number of years. For example, changes in storm tracks can cause one side of the basin to catch more precipitation relative to the other side than normal for a period of several years or changes in prevailing storm types can change the relationship between high and low elevation stations for a period of time. Thus, the selection of which stations are included on each plot is important. Selection should be made so that stations in the same geographical part of the basin or with similar elevations are displayed together. When this is done, then real shifts that occur over the basin due to climatic changes can be recognized and not inferred to be inconsistencies in the data. When many of the stations in the same portion of the basin or at similar ele vations exhibit a change in the slope of their consistency plots at the same point in time, this is an indication that this is a real shift that occurred in nature and should not be adjusted. When only a single station exhibits a change in slope, then it may indicate an inconsistency. Temperature and evaporation stations should also be grouped by location or elevation for plotting purposes though such pattern shifts are less likely to occur with these data.

With precipitation data the option exists to perform consistency checks on a seasonal rather than an annual basis, i.e. separate plots are generated for the winter and summer seasons. The decision to do consistency checks on a seasonal basis is independent of whether station weights are specified on a seasonal basis though the NWSRFS programs do force one to use the same definition of the seasons for both options. Consistency checks for precipitation should be done on a seasonal basis when snowfall dominates some months during the winter. This is because the effect of gage exposure or equipment changes is generally much larger when the precipitation is in the form of snow rather than rain. Thus, if a station is moved or a wind shield added, the existence and magnitude of an inconsistency are highly dependent on the form of the precipitation. The NWSRFS programs allow only for the separation of precipitation into different consistency plots based on month of the year, not on an event or time interval basis. Thus, the winter season should include those months when snowfall predominates and the summer season when the majority of the precipitation is rain. Seasonal precipitation consistency plots should be used whenever snowfall is significant over the river basin.

<u>Guidelines for Making Consistency Corrections</u>

There are a number of factors to consider when analyzing the consistency plots and making corrections for apparent inconsistencies in the data.

- Consistency corrections can be applied only to periods with observed data, not to periods of estimated values. Thus, it is first important to delineate those periods that contain missing values and are thus estimated from those periods with observed data. The PXPP program works only with monthly precipitation totals. A month is treated as missing and thus estimated whenever any data are missing during the month. The other programs estimate only those days or hours that are missing, thus a month can contain a mix of observed and estimated data. The IDMA program tries to designate when data are observed and when the values are estimated by changes in the color of the plotted line. Estimated data periods may not be consistent with the observed data periods for some stations, i.e. the slopes of the plots during these periods may vary. However, once the inconsistencies have been corrected for all the stations, the estimated data periods for all stations should plot at the same slope as the observed data periods, i.e. the trend of the entire plot should be straight.

- Inconsistencies should occur only when a station is moved or there is an equipment change that can affect the measurement. Thus, it is important to have the meta data that indicates the type and time of modifications to a station. Actual moves and equipment changes, such as adding a wind shield or converting from a weighing gage to a heated tipping bucket gage at a precipitation observation site, should be marked on the plots as possible times when inconsi stencies can occur. Inconsistencies seldom occur in any network that is stable, i.e. there are no site relocations or equipment changes. The network that is most susceptible to inconsistencies is the national climatological network maintained by NCDC. This network relies on volunteer observers for the most part and station relocations occur quite frequently. In some cases when a station is relocated, a new station number and name are assigned and thus you have a new record. In other cases when a move occurs, the station number remains the same, though the name will be modified by changing the distance and direction from the nearest town, e.g. ANYTOWN 1N may be changed to ANYTOWN 5SW. When a move occurs there are not clear rules as to when a new station is established and when the number remains the same and just the name is modified. Sometimes a site can be moved only a couple miles and a new number is assigned, while in other cases a station can be moved 15-20 miles and the number remains the same. In one case a station was not moved at all, but the number was changed when the two word name was reversed thus changing is alphabetical position. Frequently when a station is discontinued and a new station is established just a short distance away, the records are merged prior to being used in the data analysis programs. Obviously the time of this move should be examined for a possible inconsistency in the record.

- Data should generally be made consistent with the current location of the instrumentation. This is especially true for a real time reporting site. However, if the data for the current location seems unreasonable when compared to nearby gages, then either the data should be made consistent with another portion of the period of record or the data for the current location should be ignored by setting it to missing. When using climatic stations that have been discontinued and are no longer in use, the data can be made consistent with any portion of the period of record that you choose. If a correction is applied to the current location of a real time reporting station, this correction should also be applied in OFS.

- For temperature data the mean monthly max and min temperatures at each station are used to estimate missing data. These mean monthly values are computed from the observed data for the period of record being used. When corrections are applied to portions of the observed record, the mean monthly values for that station are altered. For precipitation the computation of monthly means is done automatically within the PXPP program, however, no preliminary processing program currently exists for temperature, thus the mean monthly temperatures must be manually adjusted whenever a consistency correction is applied to a station. The MAT program contains the option to compute and output the new monthly means, but the user must manually take this output and use it to edit the input file. This should be done after every run that involves new consistency corrections. Since the revised monthly means are not computed until after the corrections are input, the program is typically using the monthly means from the previous run which are not correct whenever new adjustments have been entered. Thus, it is necessary at the end to make one final run with no new adjustments

- Consistency plots will show random wobbles due to the natural variability of the data. Precipitation data will generally show more of this variability than temperature or evaporation. Consistency plots can also exhibit a seasonal wobble caused by variations in the relationships between stations, especially due to elevation. Especially in the intermountain west the p recipitation relationship between high and low elevation sites varies considerably from one time of the year to another. There is also typically a seasonal variation in lapse rates for temperature. The extreme case is in portions of Alaska where temperature decreases with elevation generally in the summer, but in the winter inversion conditions can persist even for maximum temperatures at interior locations during the months with the least sunlight. These effects can cause natural seasonal wobbles in the consistency plots.

- Sudden jumps in a consistency plot for a station indicate that there may be some bad data in the record. In the case of precipitation, the jump could occur because that station received much more precipitation during a given month than any of the other stations, typically from an intense thunderstorm directly over the gage. However, in most cases, discontinuities in the plots are due to improper values being entered into the data record for a station. The bad data at one station can cause bad estimates at nearby stations therefore causing jumps in their plots also. A little detective work should find the culprit. There have been some problems, at least in the past, with hourly precipitation data not being able to be decoded properly, resulting in a number of months of data that should be missing being set to all zeros. If the problem period is long enough, this should cause a shift in the slope of the consistency plot. Such a period will not respond to a correction factor since zero times any factor is still zero.

- There is a greater chance of inconsistencies occurring in mountainous areas than in flat terrain. This is because in mountainous areas the amount of normal precipitation can vary by a considerable amount over a fairly small area due to orographic effects. Thus, when a station is relocated there is good chance that the average amount of precipitation being caught will change. There is also a greater chance for inconsistencies to occur in regions with substantial snowfall than in regions where most or all of the precipitation is rain. This is because snowfall catch is greatly affected by the exposure of the gage to wind. Station relocations will frequently change the exposure of the gage. In general, one will find a greater need to make consistency corrections in mountainous areas with substantial snowfall where there are periodic site relocations. In a region of flat terrain with only rainfall, inconsistencies should seldom occur.

- The aim is not to see how many "inconsistencies" one can find, but to preserve the variability that occurs in nature and correct only those periods that clearly need to be corrected. It is also not rational to go to the other extreme and assume that consistency adjustments will just distort the observed data and therefore avoid making consistency checks. There are clearly times when the measurements at a station are modified by a relocation, equipment change, or alteration of the site surroundings. These changes to the observed data need to be adjusted so that model parameters can be reliably determined during calibration and that these parameters will be applicable for forecast applications. One should adjust those periods that clearly need to be corrected, but if there is any doubt as to whether an adjustment is needed, it is best to not make that correction.

By taking all of these factors into account, one should be able to adjust the data record for any significant inconsistencies without removing the variability that occurs in nature.

Mechanics of Computing Consistency Corrections

Even though the IDMA program will compute consistency corrections for you, it is a good idea to understand how the corrections are calculated. The calculations can be done in several ways. Following is one method.

1. For each station determine the line segment that the data should be made consistent with. As mentioned earlier this generally corresponds to the most recent location of the gage.

2. Draw a line that is parallel to the line segment from step1 that originates at the beginning of the period of observed data for the station. This is referred to as the “Corrected Line”. Note that this line will start after the beginning of the period being analyzed when measurements aren’t initiated for a given station until a later date, thus the data for the first part of the analysis period are estimated.

3. Compute corrections as:

Precipitation or evaporation:

(6-2-1)

Temperature:

(6-2-2)

where: A = Adjustment, precipitation or evaporation (multiplying factor), temperature (degrees/month)

S = Accumulated precipitation or evaporation

C = Magnitude of correction (difference between current location and Corrected Line; positive if increase, negative if decrease)

ΔN = number of months

t’ = start time of adjustment

t = end time of adjustment

Figures 6-2-1 and 6-2-2 illustrate the computation of the consistency factors. These figures show how all the corrections for the stations can be computed at one time, however, when first learning to adjust data for inconsistencies, it is probably easiest to work with one station at a time. It is probably also best in this situation to make and verify one correction at a time, beginning with the earliest.

Summation (Station minus Group Average)
negative zero positive

Summation of Group Average

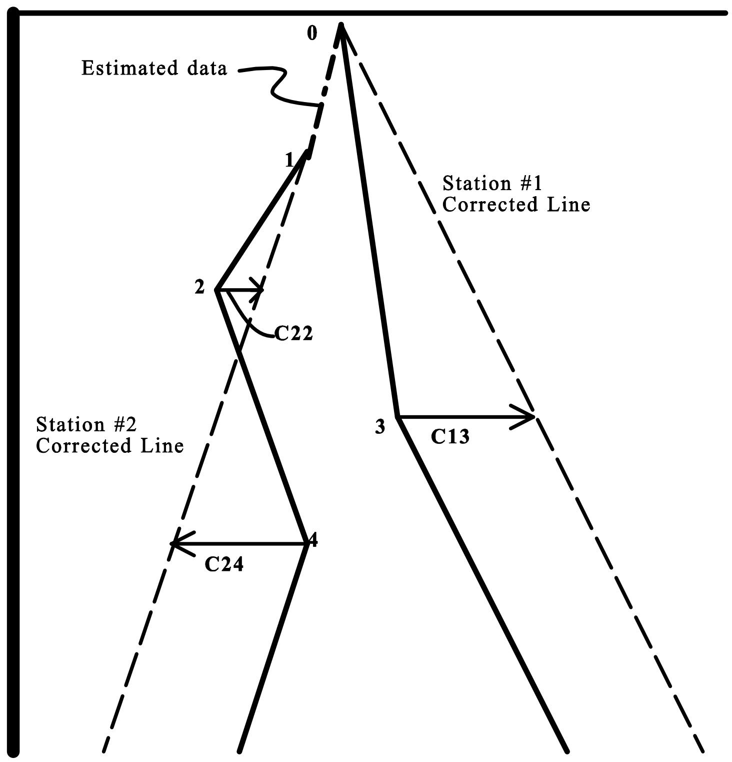

Figure 6-2-1. Illustration of pr

For figure 6-2-1 the adjustments would be compu ted as:

For station 1:

(greater than 1.0 since $C1_3$ is positive)

If the data record begins with October 1948 and the inconsistency for station 1 occurs in May 1975 and if $S1_3$ = 860 inches (i.e. accumulated precipitation for station 1, not the group aver age -- when using the PXPP, MAP, or MAPE programs this value is obtained from the tables that accompany the consistency plots using the date corresponding to when the inconsistenc y occurred) and $C1_3$ = 260 inches (computed from the plots using the station minus group ave rage summation scale), then the adjustment is 1.30. This adjustment is applied starting in O ctober 1948 and remains in effect through April 1975. In May 1975 the adjustment factor n eeds to be set back to 1.0 since the remaining data plot parallel to the Corrected Line.

For station 2:

(greater than 1.0 since $C2_2$ is positive)

(less than 1.0 since $C2_4$ and $-C2_2$ are both negativ e)

For figure 6-2-2 the adjustments would be computed as:

For station 1:

(adjustment is negative since $C1_2$ is negative)

If the data record begins in October 1977 and the inconsistency for station 1 occurs in June 1 984 (i.e. 80 months after the start of the observed record) and if $C1_2$ = -120 °F (computed fro m the plot using the deviation of accumulated means scale), then $A1_{2,0}$ = -1.5 °F/month. Th is adjustment is applied beginning in October 1977. The adjustment is then set back to 0.0 i n June 1984.

For station 2:

(adjustment is positive since $C2_3$ is positive)

(adjustment is negative since $C2_4$ and $-C2_3$ are both negative)

Deviation of Station Accumulated Mean from Group Accumulated Mean
negative zero
positive

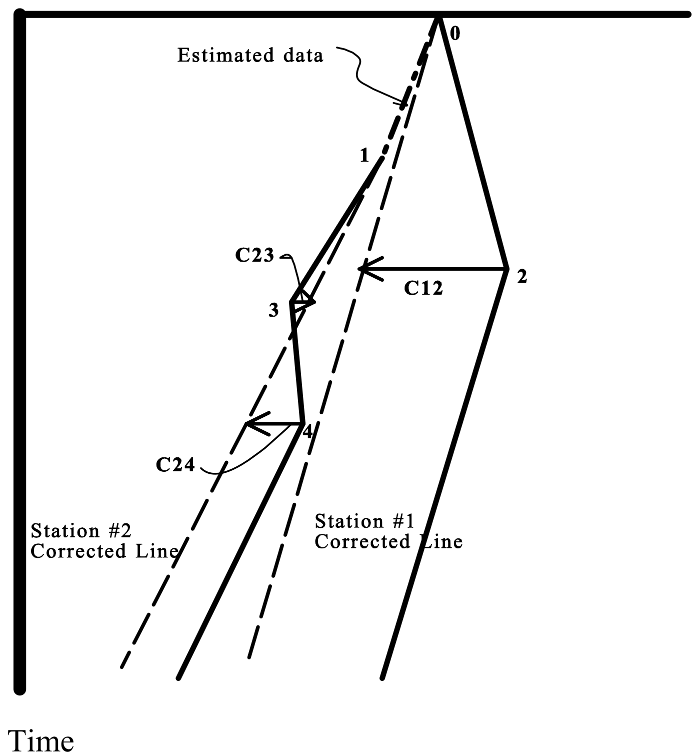

Figure 6-2-2. Illustration of consistency adjustments for temperature