CONDUCTING REGIONAL ANALYSES WITH L-RAP

6-1 SEQUENCE OF COMPUTATIONAL PROCEDURES

When the L-RAP program starts (Screen Shot 6-1), a series of tabs are visible across the top of the screen. The L-RAP interface is designed to guide the user through the steps in the regional analysis by progressing from left to right through the various tabs, as follows:

- Control
- Data Management
- Data Filter
- Data Screening
- Regional Analysis
- Quantile Estimates
- L-Moment Calculator (Single Station Computations)

The procedures for working through the various steps in the analysis are described in detail in the following sections. It should be noted that the examples in the following sections are for multi-month precipitation. Similar procedures are applicable to other data series with minor changes to accommodate the type of data series.

6-2 CREATING A PROJECT FILE

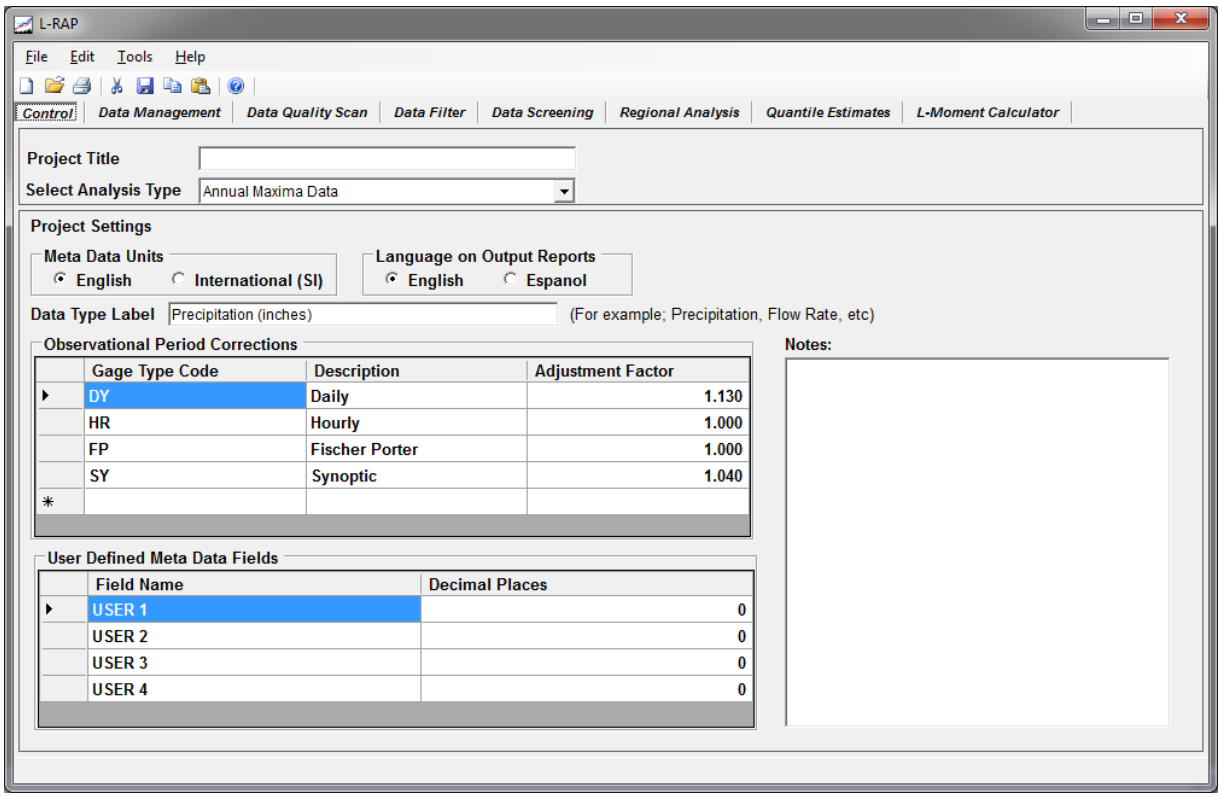
A project file should be created that allows for easy access to data imported to the program and the program settings. The project file is created as follows:

- Click on the *File* icon on the L-RAP toolbar
- Click on the *Save As* command
- Navigate to the folder where the project file is to be stored, or create a folder using the normal Windows procedure for creating a folder
- Save the project file with a user-specified name (Screen Shot 6-2). The file will be stored with an RGA extension in a standard ASCII format. These files can be opened and viewed in any text editor, such as Notepad.

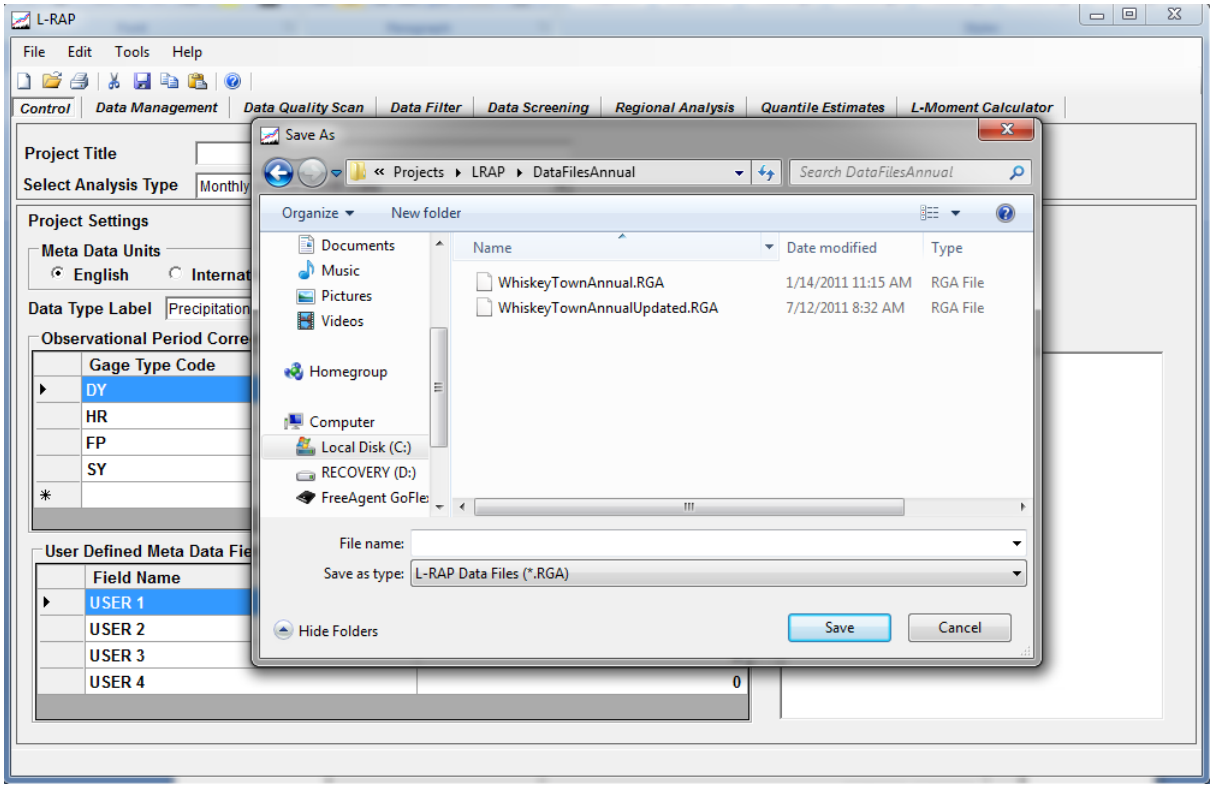
6-3 OPENING A PREVIOUSLY CREATED PROJECT FILE

If a project file has previously been created, it can be opened using standard Windows procedures, specifically:

- Click on the *File* icon on the L-RAP toolbar
- Click on the *Open* command
- Navigate to the folder where the project file is to located, highlight the desired project file and click the *Open* command; the file will automatically be loaded.



Screen Shot 6-1 – Control Tab Screen



Screen Shot 6-2 – Saving the Project File

6-4 CONTROL TAB

The *Control* tab provides the first step in conducting the regional analysis. The ***Control*** screen has features that are used to identify: the preferred language for interaction with the program; the type of data series; and a means to provide any additional information that is required for the type of data series that has been selected.

Complete the ***Control*** screen as follows:

- Click on the *Language* radio button to set the desired language for the report output (English, Spanish).
- Identify the project by entering a title in the *Project Name* field
- Click on the *Analysis Type* List button to set the type of data analysis (Screen Shot 6-1)
- Enter the units associated with the data being analyzed. This label will appear on subsequent graphics and reports
- Fill out the data entry grid relating the gage type to the adjustment for the number of observational periods (Screen Shot 6-2). The data entry grid will be used to set the relationship between the type of instrumentation (gage type) and the adjustment for the number of observational periods
- Up to four user-defined metadata fields may be included in the analysis. The user defined fields are included on the Data Management tab for each station analyzed. Meta data fields can be defined at any time during the analysis.

6-5 DATA MANAGEMENT

The *Data Management* tab contains features for: importing data series; editing of data; and editing of station metadata. At the current time, importing of data is accomplished from Excel spreadsheets to provide a familiar, easily used format for data transferal. An alternative method of importing data would be to edit the project files with the RGA extension, which are ASCII formatted.

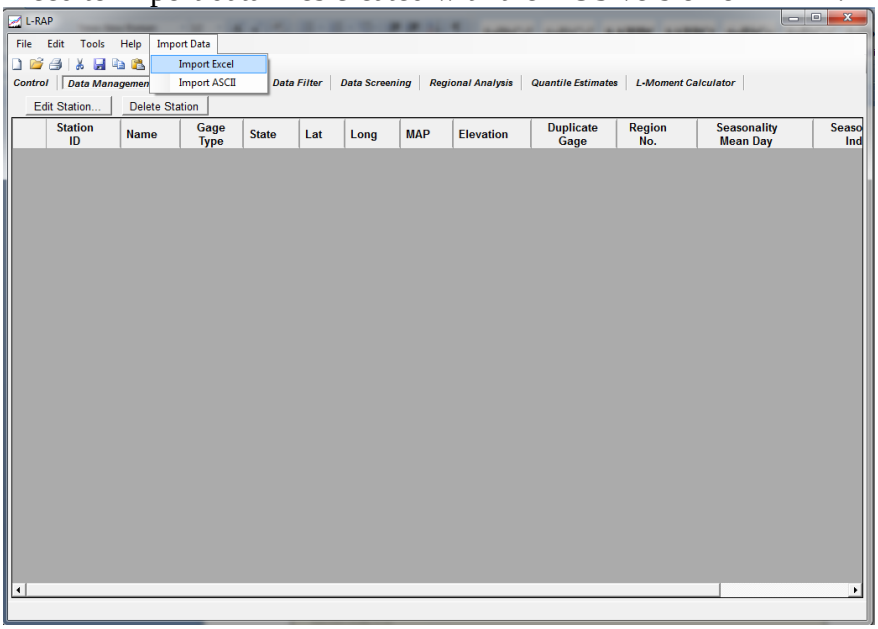
6-5.1 Excel Templates for Data Entry

Excel templates have been created for a variety of data series types for use in importing of data series and station metadata. Chapter 5 contains information about assembling datasets using Excel templates that are provided as part of L-RAP.

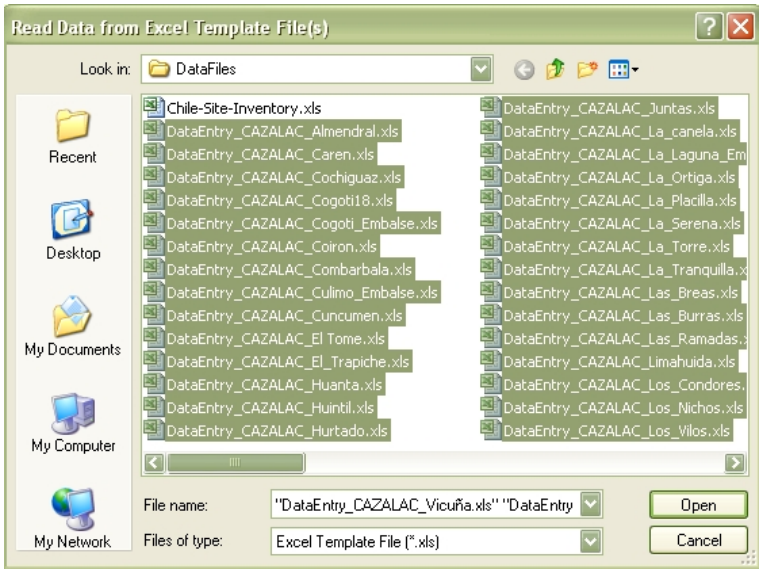
6-5.2 Importing Data

Station data are imported as follows:

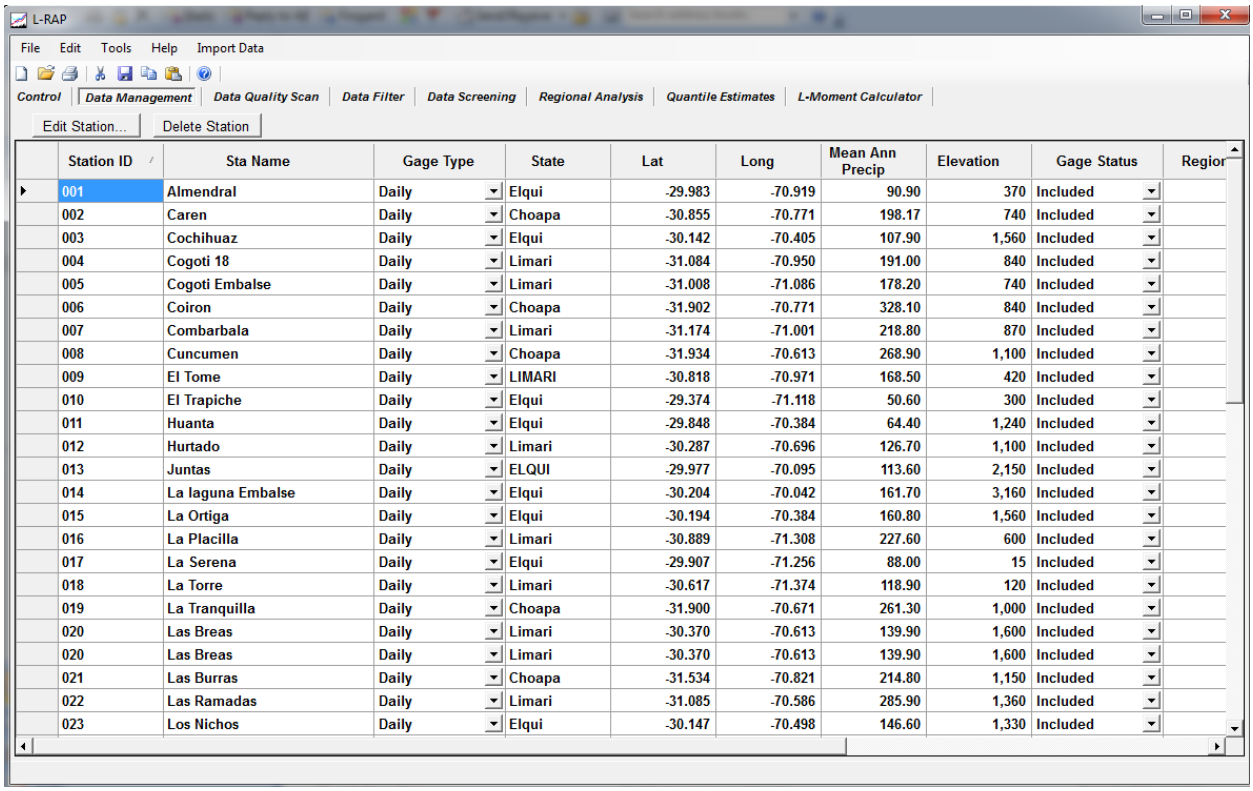
- From the *Data Management* tab, click the *Import Data* menu and then *Import Excel*. (Screen Shot 6-3a)
- Navigate to the folder where the Excel data files are located for the stations of interest
- Select one or more Excel files by holding the control key and clicking the desired file names. (Screen Shot 6-3b)
- Click on the *Open* button to begin the process of importing station data
- Once the import is complete, the imported stations metadata will appear in the table at the bottom of the screen. (Screen Shot 6-3c)
- The Import ASCII option is used to import legacy data files. This feature should only be used if you need to import data files created with the DOS version of LRAP.



Screen Shot 6-3a – Data Management Screen



Screen Shot 6-3b – Highlighted Files for Loading into L-RAP



Station ID	Sta Name	Gage Type	State	Lat	Long	Mean Ann Precip	Elevation	Gage Status	Regior
001	Almendral	Daily	Elqui	-29.983	-70.919	90.90	370	Included	
002	Caren	Daily	Choapa	-30.855	-70.771	198.17	740	Included	
003	Cochihuaz	Daily	Elqui	-30.142	-70.405	107.90	1,560	Included	
004	Cogoti 18	Daily	Limari	-31.084	-70.950	191.00	840	Included	
005	Cogoti Embalse	Daily	Limari	-31.008	-71.086	178.20	740	Included	
006	Coiron	Daily	Choapa	-31.902	-70.771	328.10	840	Included	
007	Combarbala	Daily	Limari	-31.174	-71.001	218.80	870	Included	
008	Cuncumen	Daily	Choapa	-31.934	-70.613	268.90	1,100	Included	
009	El Tome	Daily	LIMARI	-30.818	-70.971	168.50	420	Included	
010	El Trapiche	Daily	Elqui	-29.374	-71.118	50.60	300	Included	
011	Huanta	Daily	Elqui	-29.848	-70.384	64.40	1,240	Included	
012	Hurtado	Daily	Limari	-30.287	-70.696	126.70	1,100	Included	
013	Juntas	Daily	ELQUI	-29.977	-70.095	113.60	2,150	Included	
014	La laguna Embalse	Daily	Elqui	-30.204	-70.042	161.70	3,160	Included	
015	La Ortiga	Daily	Elqui	-30.194	-70.384	160.80	1,560	Included	
016	La Placilla	Daily	Limari	-30.889	-71.308	227.60	600	Included	
017	La Serena	Daily	Elqui	-29.907	-71.256	88.00	15	Included	
018	La Torre	Daily	Limari	-30.617	-71.374	118.90	120	Included	
019	La Tranquilla	Daily	Choapa	-31.900	-70.671	261.30	1,000	Included	
020	Las Breas	Daily	Limari	-30.370	-70.613	139.90	1,600	Included	
020	Las Breas	Daily	Limari	-30.370	-70.613	139.90	1,600	Included	
021	Las Burras	Daily	Choapa	-31.534	-70.821	214.80	1,150	Included	
022	Las Ramadas	Daily	Limari	-31.085	-70.586	285.90	1,360	Included	
023	Los Nichos	Daily	Elqui	-30.147	-70.498	146.60	1,330	Included	

Screen Shot 6-3c – Data Management Screen Showing Stations Loaded into L-RAP

6-5.3 Adding New Stations to an Existing Project File

Adding new stations to an existing project file is accomplished as follows:

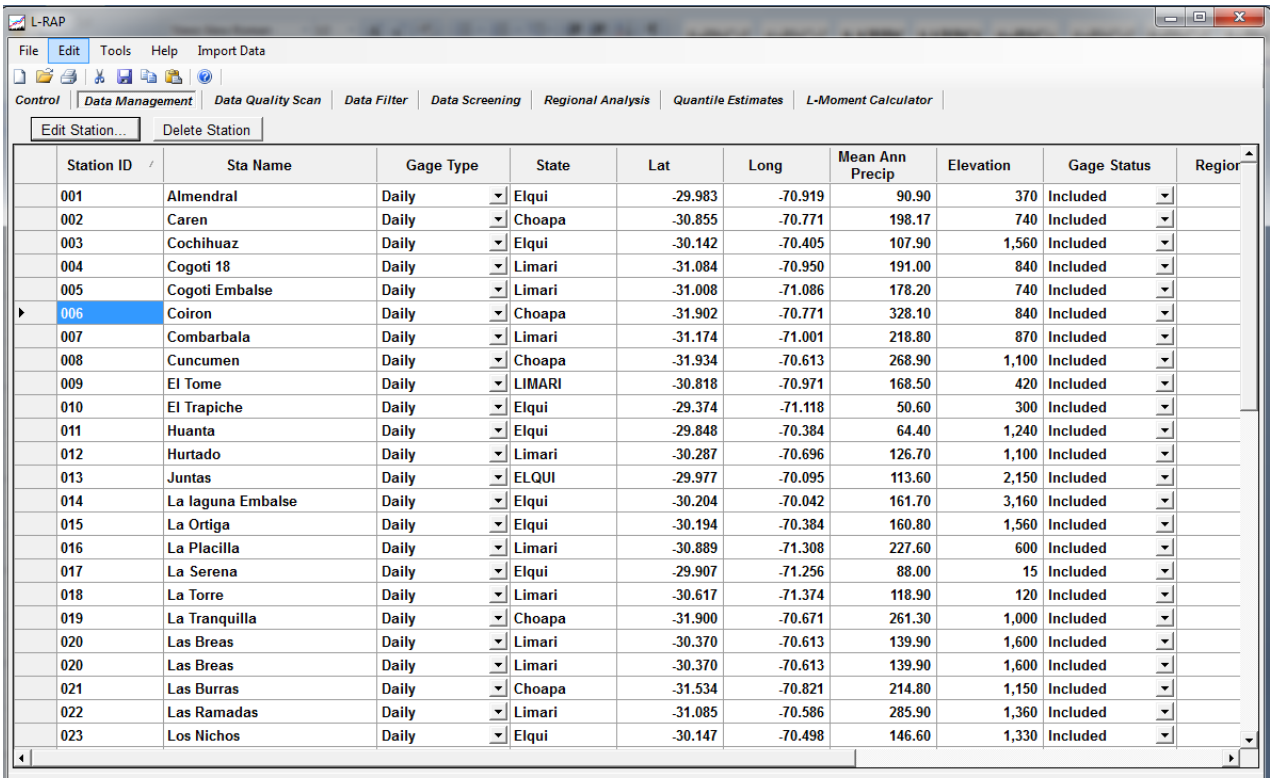
- Import the new stations using the importing procedures described in Section 6-5.2 above. The new stations will be added to the existing project file and will appear in the table at the bottom of the *Data Management* tab.
- Save the project file as described in Section 6-2.

6-5.4 Editing Data Series and Station Metadata

Station metadata and data series may be edited from the *Data Management* tab. Metadata can be changed by editing the data directly in the table at the bottom of the *Data Management* tab. Clicking on any of the column headings sorts the stations by the alpha-numeric values in that column. The data series can be edited as follows:

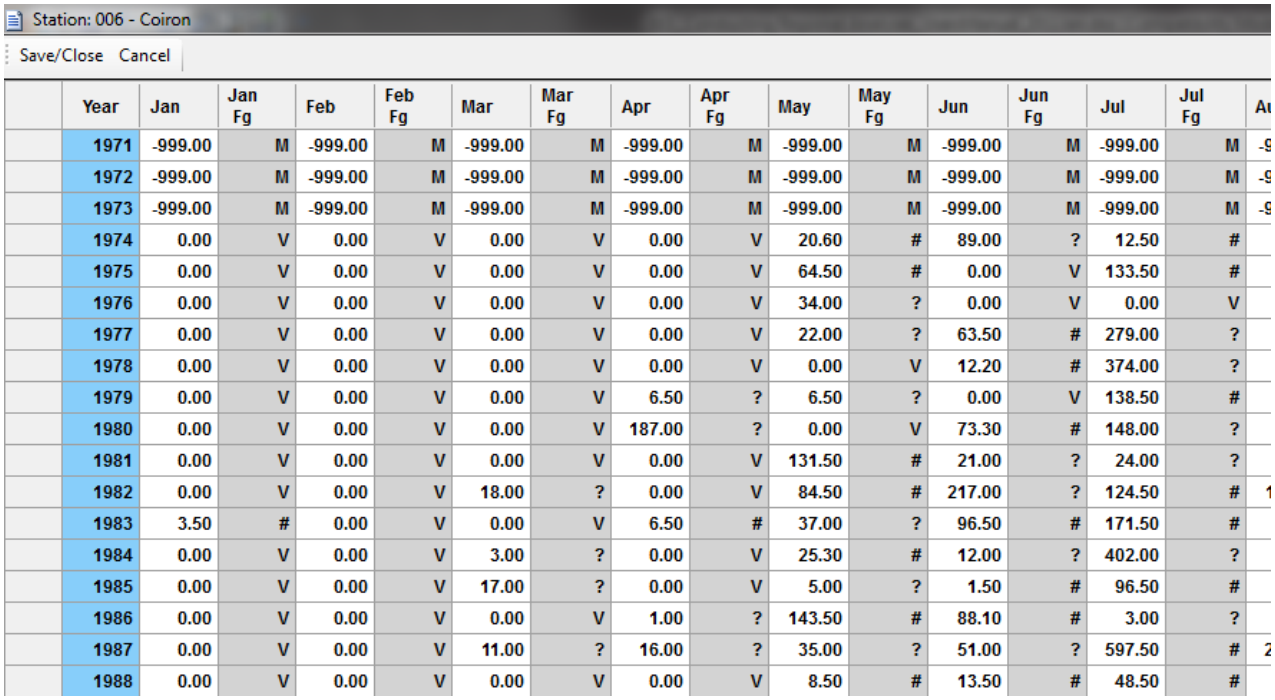
- Click on the row containing the Station of interest (Screen Shot 6-4a),
- Click on the *Edit Data* button and a new editing window will appear with the data series and data quality flags for the selected station (Screen Shot 6-4b)
- Make changes as needed and save the changes by clicking on the *Save/Close* button
- Edit another data series by highlighting another station, clicking on the *Edit Data* button, and repeating the editing procedure described above
- Complete the editing session by clicking on the *Save/Close* button on the *Station Editor* window.

In the process of forming homogeneous regions, reassignment of a station to a different region is accomplished by simply changing the region number for a given station. This is accomplished by editing the station metadata on the grid at the bottom of the Data Management tab (Screen Shot 6-4a).



Station ID	Sta Name	Gage Type	State	Lat	Long	Mean Ann Precip	Elevation	Gage Status	Regior
001	Almendral	Daily	Elqui	-29.983	-70.919	90.90	370	Included	
002	Caren	Daily	Choapa	-30.855	-70.771	198.17	740	Included	
003	Cochihuaz	Daily	Elqui	-30.142	-70.405	107.90	1,560	Included	
004	Cogoti 18	Daily	Limari	-31.084	-70.950	191.00	840	Included	
005	Cogoti Embalse	Daily	Limari	-31.008	-71.086	178.20	740	Included	
006	Coiron	Daily	Choapa	-31.902	-70.771	328.10	840	Included	
007	Combarbala	Daily	Limari	-31.174	-71.001	218.80	870	Included	
008	Cuncumen	Daily	Choapa	-31.934	-70.613	268.90	1,100	Included	
009	El Tome	Daily	LIMARI	-30.818	-70.971	168.50	420	Included	
010	El Trapiche	Daily	Elqui	-29.374	-71.118	50.60	300	Included	
011	Huanta	Daily	Elqui	-29.848	-70.384	64.40	1,240	Included	
012	Hurtado	Daily	Limari	-30.287	-70.696	126.70	1,100	Included	
013	Juntas	Daily	ELQUI	-29.977	-70.095	113.60	2,150	Included	
014	La laguna Embalse	Daily	Elqui	-30.204	-70.042	161.70	3,160	Included	
015	La Ortiga	Daily	Elqui	-30.194	-70.384	160.80	1,560	Included	
016	La Placilla	Daily	Limari	-30.889	-71.308	227.60	600	Included	
017	La Serena	Daily	Elqui	-29.907	-71.256	88.00	15	Included	
018	La Torre	Daily	Limari	-30.617	-71.374	118.90	120	Included	
019	La Tranquilla	Daily	Choapa	-31.900	-70.671	261.30	1,000	Included	
020	Las Breas	Daily	Limari	-30.370	-70.613	139.90	1,600	Included	
020	Las Breas	Daily	Limari	-30.370	-70.613	139.90	1,600	Included	
021	Las Burras	Daily	Choapa	-31.534	-70.821	214.80	1,150	Included	
022	Las Ramadas	Daily	Limari	-31.085	-70.586	285.90	1,360	Included	
023	Los Nichos	Daily	Elqui	-30.147	-70.498	146.60	1,330	Included	

Shot 6-4a – Metadata Grid at the bottom of the Data Management Tab
(Station 006 is selected for Editing)



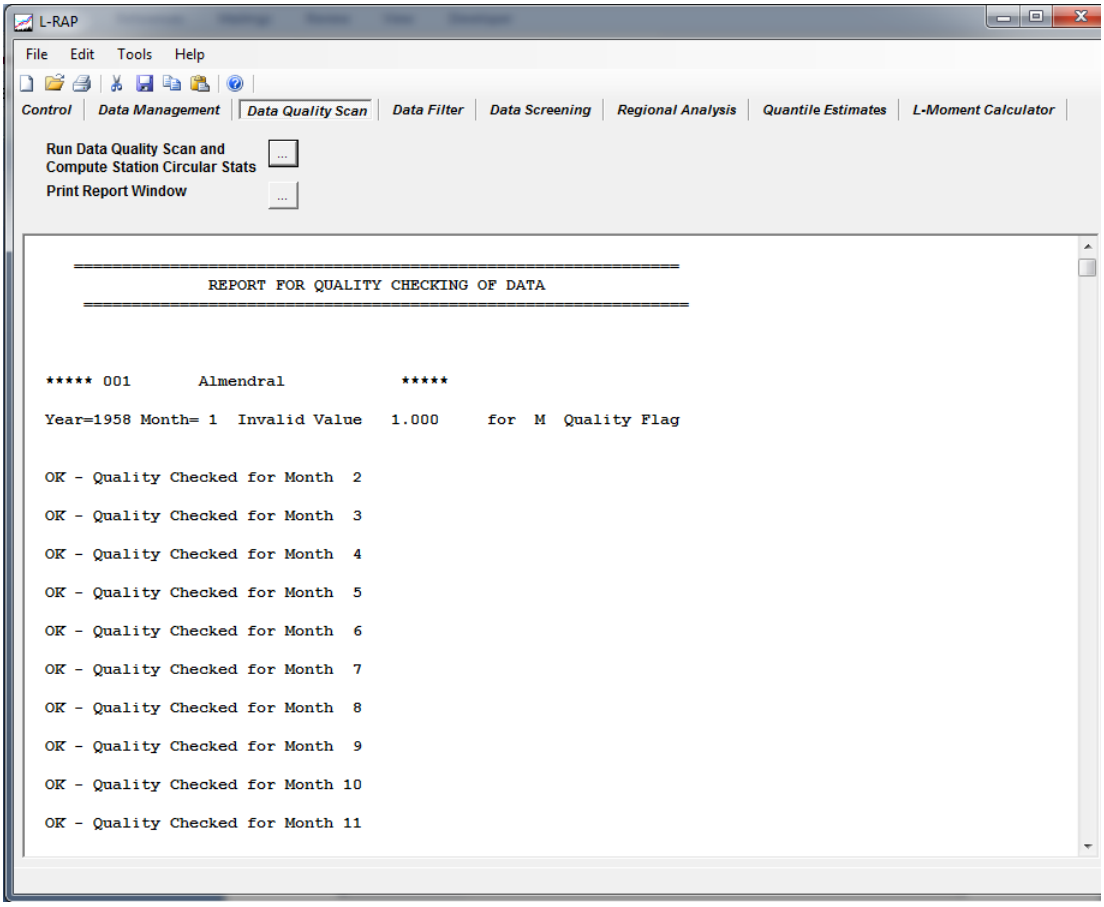
	Year	Jan	Jan Fg	Feb	Feb Fg	Mar	Mar Fg	Apr	Apr Fg	May	May Fg	Jun	Jun Fg	Jul	Jul Fg	Au
	1971	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-9
	1972	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-9
	1973	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-999.00	M	-9
	1974	0.00	V	0.00	V	0.00	V	0.00	V	20.60	#	89.00	?	12.50	#	
	1975	0.00	V	0.00	V	0.00	V	0.00	V	64.50	#	0.00	V	133.50	#	
	1976	0.00	V	0.00	V	0.00	V	0.00	V	34.00	?	0.00	V	0.00	V	
	1977	0.00	V	0.00	V	0.00	V	0.00	V	22.00	?	63.50	#	279.00	?	
	1978	0.00	V	0.00	V	0.00	V	0.00	V	0.00	V	12.20	#	374.00	?	
	1979	0.00	V	0.00	V	0.00	V	6.50	?	6.50	?	0.00	V	138.50	#	
	1980	0.00	V	0.00	V	0.00	V	187.00	?	0.00	V	73.30	#	148.00	?	
	1981	0.00	V	0.00	V	0.00	V	0.00	V	131.50	#	21.00	?	24.00	?	
	1982	0.00	V	0.00	V	18.00	?	0.00	V	84.50	#	217.00	?	124.50	#	1
	1983	3.50	#	0.00	V	0.00	V	6.50	#	37.00	?	96.50	#	171.50	#	
	1984	0.00	V	0.00	V	3.00	?	0.00	V	25.30	#	12.00	?	402.00	?	
	1985	0.00	V	0.00	V	17.00	?	0.00	V	5.00	?	1.50	#	96.50	#	
	1986	0.00	V	0.00	V	0.00	V	1.00	?	143.50	#	88.10	#	3.00	?	
	1987	0.00	V	0.00	V	11.00	?	16.00	?	35.00	?	51.00	?	597.50	#	2
	1988	0.00	V	0.00	V	0.00	V	0.00	V	8.50	#	13.50	#	48.50	#	

Screen Shot 6-4b– Editing of Data Series for a Selected Station

6-6 DATA QUALITY SCANNING

The *Data Quality Scan* tab is used to check the station data for input errors. Error checking must be performed each time new data is imported into the program. The scan takes place automatically each time a project data file (RGA file) is opened. To quality check the data, perform the following steps.

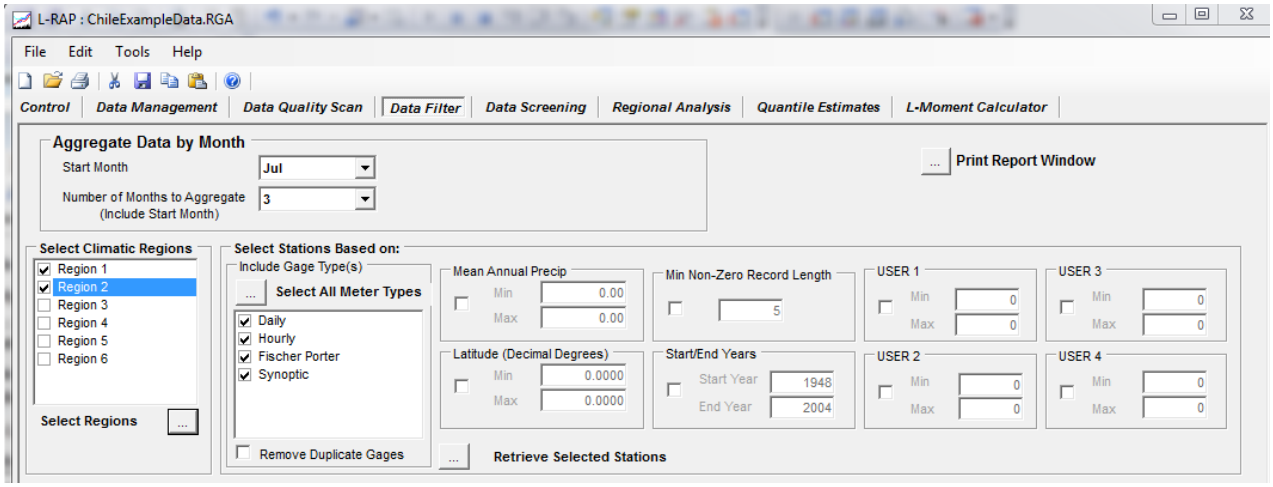
- Click the *Data Quality Scan* tab. Click the *Run Data Quality Scan and Compute Station Circular Statistics* button. A report is automatically generated and appears in the text window on the *Data Quality Scan* Screen (Screen Shot 6-5).
- The error report should be scanned and all errors should be addressed through editing of the data series and station metadata (discussed in Section 6-5.4).



Screen Shot 6-5 – Data Management Screen Showing Error Report for Station Data Series (An Input Error is noted for Station 001, Almendral)

6-7 DATA FILTERING

The ***Data Filter*** screen provides options for selecting regions and setting criteria for selecting sites/stations within the selected regions. Once the group of sites/stations is selected, that group is used in subsequent analyses for data screening and conducting regional analyses (Screen Shot 6-6).



Screen Shot 6-6 – Screen Shot of Data Filtering Screen

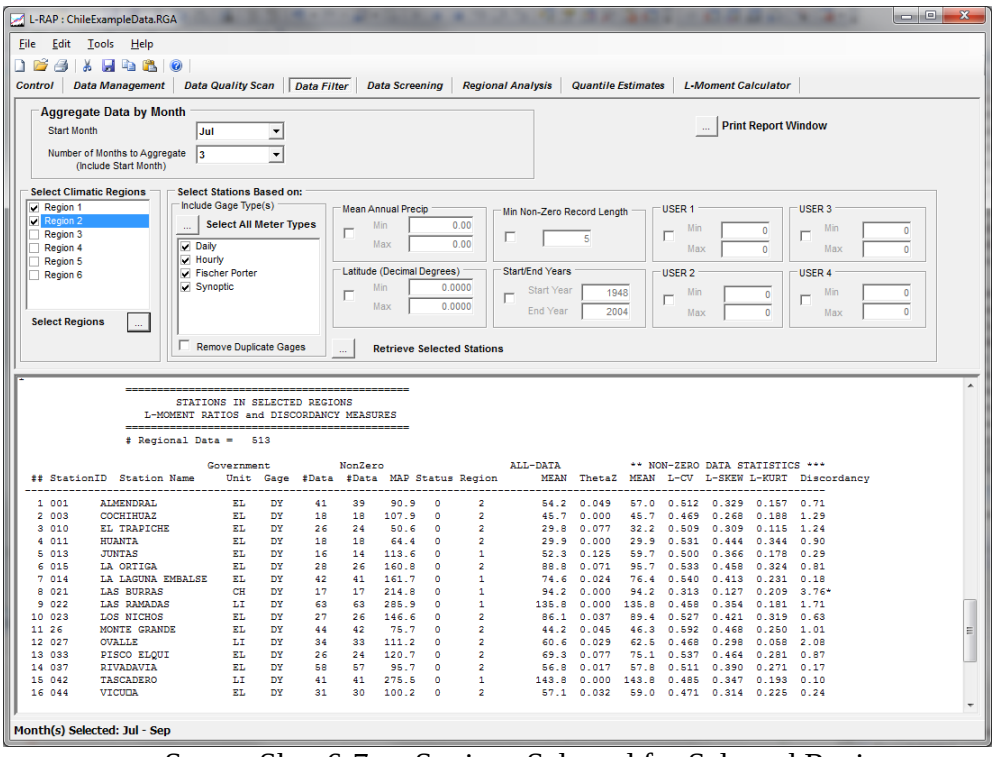
6-7.1 Sequence of Actions for Selecting Stations for Proposed Region

The sequence of actions for selecting stations is as follows:

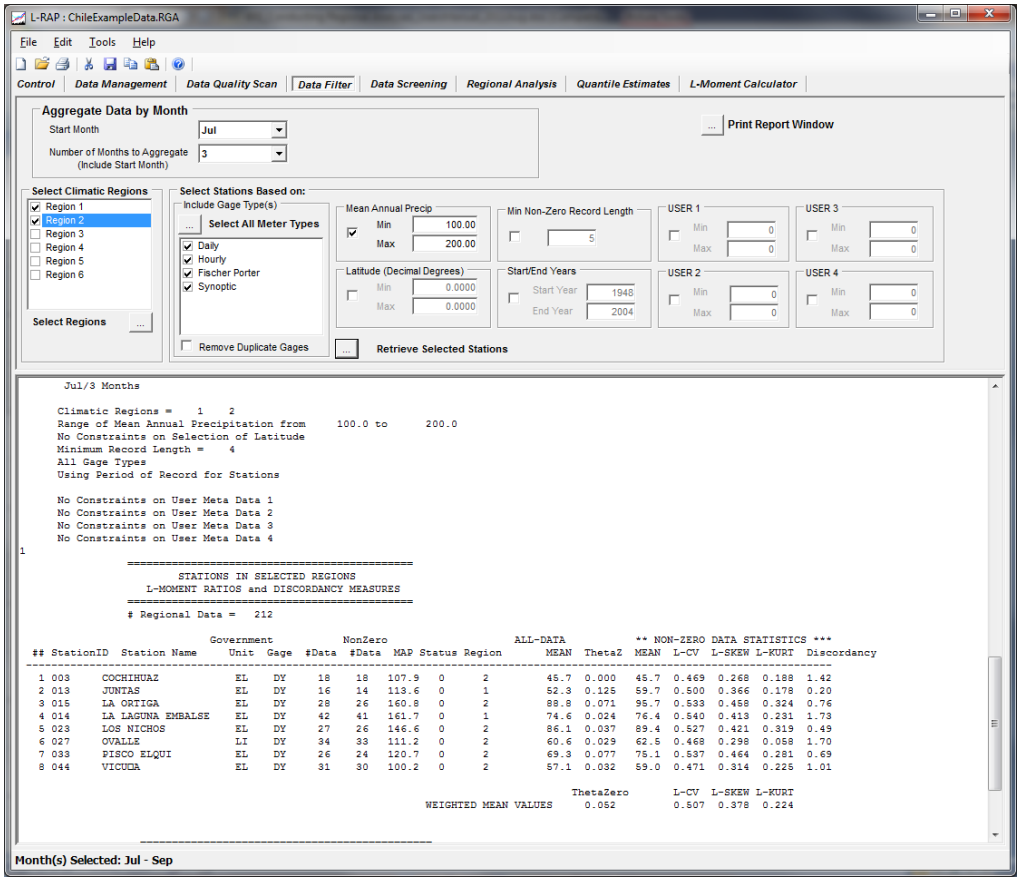
- If monthly or multi-month data are to be analyzed, use the drop-down menu in the *Aggregate Data by Month* box and select the starting month and number of months for analysis. For all other data types, proceed to the next step to select regions
- Use check-boxes to select the regions of interest in the *Select Climatic Regions* box
- Click on the *Select Regions* button to create the listing of stations within the regions of interest, the list of potential stations will appear in the text window (Screen Shot 6-7a). By default, all stations in the selected regions are included in the analysis. The view window now displays all stations in the region along with L-moment statistics and discordancy measures.
- The user can now either proceed with the analysis of the stations in the selected regions or further filter the stations based on metadata fields.

Optionally Filter Stations based on Metadata Fields

- Set any desired station selection criteria, such as minimum record length, range of mean annual precipitation, etc. in the *Select Stations Based On* box.
- Click on *Retrieve Selected Stations* button to filter the stations in the selected regions based on the selected criteria, (Screen Shot 6-7b). The view window now displays the selected stations along with L-moment statistics and discordancy measures.



Screen Shot 6-7a – Stations Selected for Selected Regions



Screen Shot 6-7b – Stations Selected Meeting Selection Criteria

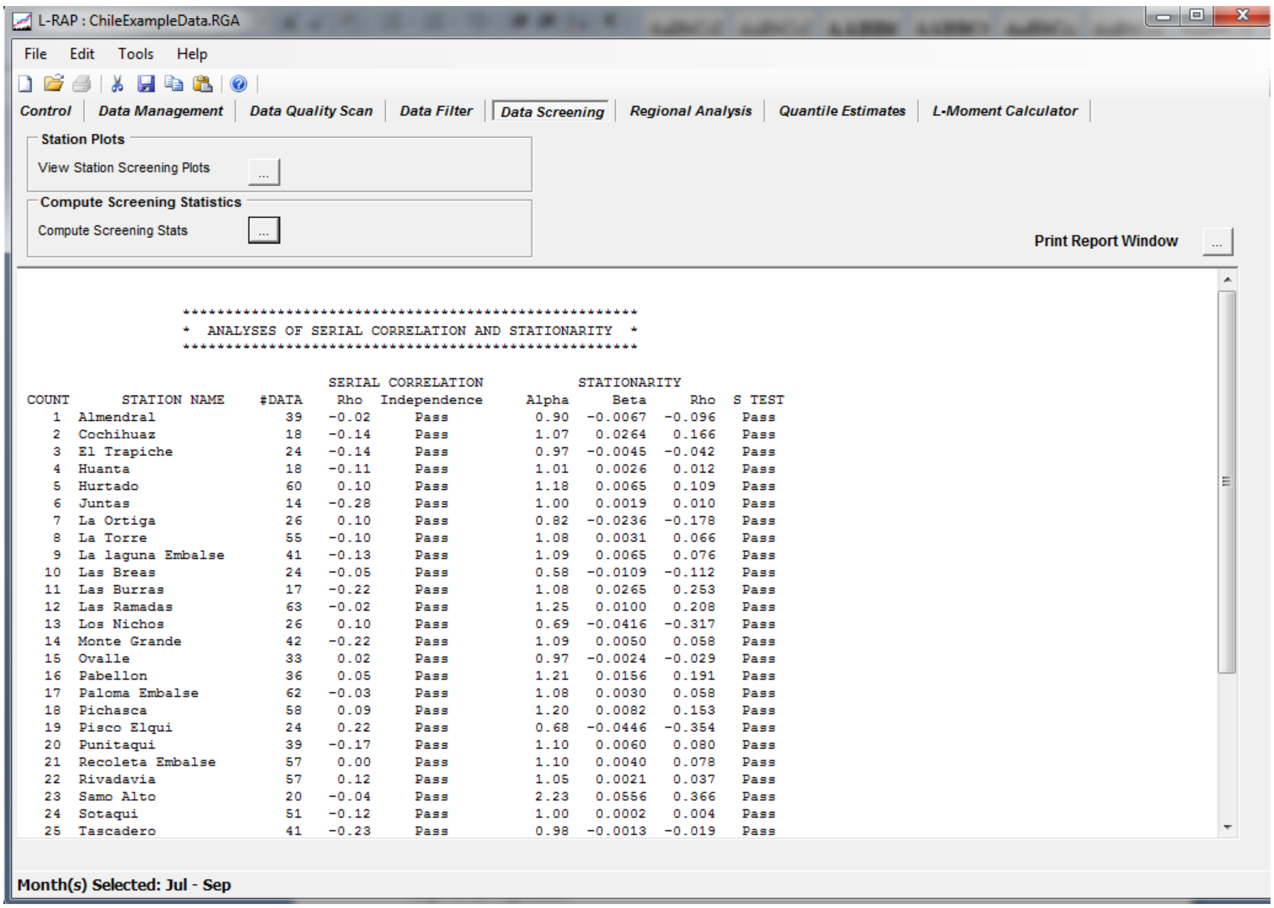
6-8 DATA SCREENING

The ***Data Screening*** tab provides functionality to conduct tests for serial independence and stationarity of the data series. There is also a feature for computing cross-correlation coefficients for the data series and producing a plot that describes the decay of cross-correlation with distance between stations. Lastly, there is a powerful station graphics tool that provides: a probability-plot of the data series; a time-series plot; an L-moment ratio diagram; and a histogram of the seasonality of maxima values. Each of these features is described in the following sections. The layout of the *Data Screening* screen is shown in Screen Shot 6-8.

The stations available for *Data Screening* are set through application of the ***Data Filter***. The *Data Filter* must first be used to select stations for the ***Data Screening*** functions to become active.

6.8.1 Sequence of Actions for Data Screening Selecting Stations for Proposed Region

- Click on the *Compute Screening Statistics* button (Screen Shot 6-8) to generate the statistics for serial independence, stationarity and cross-correlation of the data series
- Click on *View Station Screen Plots* button (Screen Shot 6-8) to view the probability-plots, time-series plots, L-moment diagrams and seasonality histograms for the selected stations



Screen Shot 6-8 – Layout of Screen for Data Screening Showing Report Created After Clicking the Compute Screening Stats Button

6-8.2 Serial Independence and Stationarity of Data Series

Statistical tests for serial independence and stationarity of the data series are standard tests that are conducted as part of the data screening process. The test for serial independence is conducted to confirm that the data series are serially independent, a required condition for conducting frequency analyses. The test for stationarity is conducted to confirm that that there are no significant trends in the data over the period of observation with regard to central tendency. Both of these tests are conducted on the collection of stations, with the findings dependent on the behavior of the group of stations rather than on any particular station.

Serial Independence – The test for serial independence is conducted by computing a serial correlation coefficient (lag-1 auto-correlation coefficient) for the data series for each station. A global weighted-average serial correlation coefficient is then computed with the weightings based on the record length for each station. A standard t-test is conducted to examine if the global serial correlation coefficient is significantly different from zero. The null hypothesis is that the global serial correlation coefficient is zero for a sample size equal to the global average record length for the collection of stations. The results are displayed in the text view window (Screen Shot 6-8).

Stationarity – The test for stationary is conducted by first dividing each element of the data series at a station by the at-site mean and then subtracting 2000 from the year of occurrence. A time-series plot is then assembled and standard linear regression methods are used to compute an intercept and slope (Figure 6-1). A perfectly stationary sample would have an intercept value (Alpha) of 1.00 (at year 2000) and a slope (Beta) of zero over the period of record (Figure 6-1). This approach allows the regression parameters from all stations to be grouped for a statistical test for the global slope of the regression.

A global slope value is computed as a weighted-average of the slope values for each station weighted by record length. A standard t-test is conducted to examine if the global slope value is significantly different from zero. The null hypothesis is that the global slope value is zero for a sample size equal to the global average record length for the collection of stations. The results are displayed in the text view window (Screen Shot 6-8).

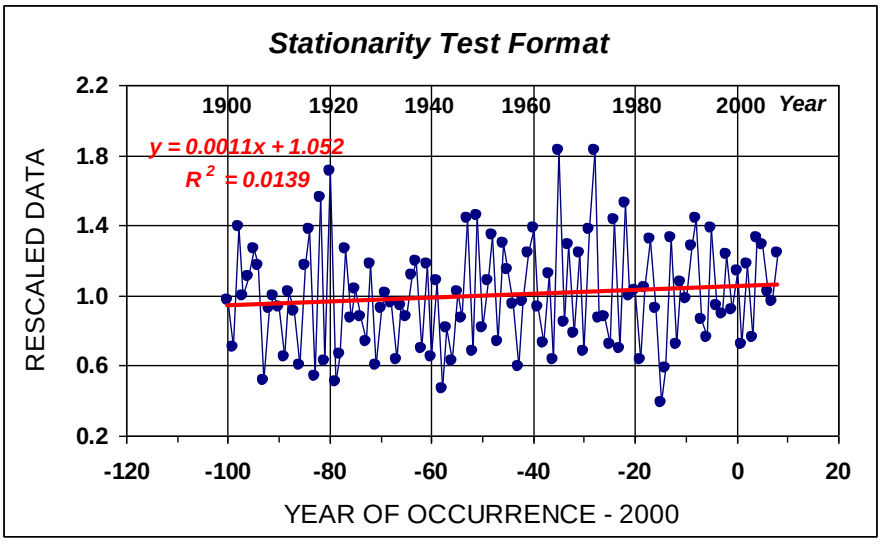


Figure 6-1 – Example of Plot Used for Test for Stationarity

6-8.3 Cross-Correlation of Data Series

One of the goals of a regional analysis is to use the collective information from a grouping of stations to improve the reliability of quantile estimates for all stations. The magnitude of improvement in quantile estimates for a regional analysis relative to at-site estimates is based in-part on the size of the regional dataset. If the data series for the stations used in the regional analysis are independent or have low cross-correlation, then the equivalent independent record length for the regional dataset is nearly equally to size of the regional dataset. Conversely, if the data series for the stations are highly cross-correlated, then the amount of information afforded by the regional dataset is much less that the size of regional dataset.

A plot of the decay of cross-correlation with distance between stations provides an approach to assessing the magnitude and behavior of cross-correlation between station data series. L-RAP produces a cross-correlation decay plot and provides a LOWESS fit (Cleveland[2]) to assist in a qualitative assessment of the level of cross-correlation (Figure 6-2).

All plots/graphics within L-RAP have the capability of being saved as jpg files for import into other documents. Hover the mouse over the upper right hand corner of each graphic to display icons for printing, copying, and saving of images. Just click on the icon for the desired action. Each of the graphics may be modified to better fit user needs. Titles may be changed along with labeling of the axes and plotting colors and symbols. Right click on the image to bring up the menus for altering the graphics.

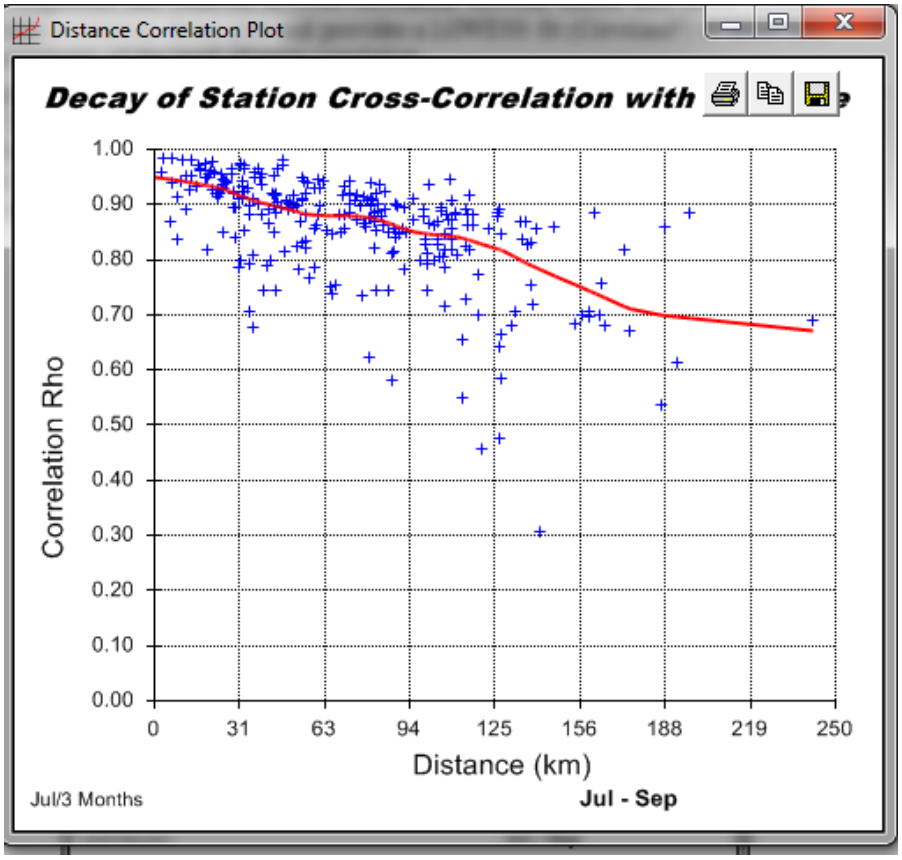


Figure 6-2 – Example Plot of Decay of Cross-Correlation with Distance

6-8.4 Seasonality Statistics

Circular statistics (Fisher[8]) are appropriate for analysis of data that are circular or directional in nature. Months of the year, days of the year (dates), compass headings (wind direction) are all examples of circular data. For example, January (month 1) follows December (month 12). Arithmetic averaging of a group of numerical months or dates is not appropriate with conventional sample statistics because the counting system is circular not linear. In conducting the analysis of the seasonality of annual maxima or extreme storms, the Julian day of the year is used for describing the date of occurrence. The *average day of occurrence* is analogous to the arithmetic mean and the *seasonality index* (Dingman[6]) is analogous to a standardized measure of variation. Specifically, values of the seasonality index range from zero to unity with values near zero indicating wide variation in the dates of occurrence. A seasonality index near unity indicates low variation in the dates of occurrence and strong clustering of dates. Circular statistics for dates of occurrence using Julian day-of-year are computed as follows:

Conversion of Julian day-of-year to compass direction (ϕ_i):

$$\phi_i = 360\ [\ J_i / Days_{total}] \qquad (6\text{-}1)$$

Compute vectors for compass direction:

$$S = \sum_{i=1}^{n} P_i\ [sin(\phi_i)] \qquad (6\text{-}2a)$$

$$C = \sum_{i=1}^{n} P_i\ [cos(\phi_i)] \qquad (6\text{-}2b)$$

Compute Average Day-of-Occurrence (Julian day-of-year J_{mean}):

$$\phi_2 = ArcTan(S/C) \qquad (6\text{-}3a)$$

$$\phi_m = \phi_2 \qquad if\ S>0\ and\ C>0 \qquad (6\text{-}3b)$$

$$\phi_m = \phi_2 + 180^\circ \qquad if\ C<0 \qquad (6\text{-}3c)$$

$$\phi_m = \phi_2 + 360^\circ \qquad if\ S<0\ and\ C>0 \qquad (6\text{-}3d)$$

$$J_{mean} = 365\ \phi_m \qquad (6\text{-}3e)$$

Compute Seasonality Index (*SI*):

$$SI = SQRT(S^2 + C^2)\ /\ P_{total} \qquad (6\text{-}4a)$$

$$P_{total} = \sum_{i=1}^{n} P_i \qquad (6\text{-}4b)$$

where: J_i = Julian day-of-year for given date of interest; $Days_{total}$ is the total number of days in the current year; P_i is the data value for a given date (J_i); n is the total number of data and date pairs; and P_{total} is the sum of all data values for the dataset.

Screen Shot 6-9 shows an example output of seasonality statistics for annual precipitation data. The Seasonality report output is included with the Regional Analysis Report. Figures 6-3a,b,c depict examples of seasonality histograms for annual precipitation and associated circular statistics. Note that concentration of the data in several months results in a high value of the seasonality index, whereas data spread throughout the year results in a low value of the seasonality index.

```
*****************************************************************************
*   SEASONALITY ANALYSIS USING CIRCULAR STATISTICS FOR MONTH OF MAXIMUM VALUE  *
*****************************************************************************

Seasonality Report

INPUT FILE = C:\Program Files\LRAP\Database.txt
```

ID	STATION	GAGE	LATITUDE	LONGITUDE	MAP	REGION	NDATA	JULIAN MEAN DAY	SEASONAL RESULTS SEASONALITY INDEX
001	Almendral	DY	-29.983	-70.919	90.9	7	40	186	0.886
003	Cochihuaz	DY	-30.142	-70.405	107.9	7	17	171	0.867
010	El Trapiche	DY	-29.374	-71.118	50.6	7	25	186	0.889
012	Hurtado	DY	-30.287	-70.696	126.7	7	57	183	0.842
013	Juntas	DY	-29.977	-70.094	113.6	7	15	181	0.793
017	La Serena	DY	-29.907	-71.256	88.0	7	30	181	0.878
018	La Torre	DY	-30.617	-71.374	118.9	7	53	177	0.798
26	Monte Grande	DY	-30.089	-70.493	75.7	7	43	186	0.899
027	Ovalle	DY	-30.601	-71.200	111.2	7	34	177	0.873
032	Pichasca	DY	-30.393	-70.868	124.6	7	58	181	0.836
033	Pisco Elqui	DY	-30.122	-70.493	120.7	7	26	182	0.891
036	Recoleta Embalse	DY	-30.507	-71.100	113.8	7	55	177	0.856
037	Rivadavia	DY	-29.977	-70.561	95.7	7	57	189	0.870
038	Salamanca	DY	-31.776	-70.967	258.8	7	31	181	0.911
039	Samo Alto	DY	-30.410	-70.939	108.9	7	18	191	0.885
041	Sotaqui	DY	30.631	71.120	121.4	7	47	180	0.867
044	Vicuña	DY	-30.057	-70.717	100.2	7	30	188	0.891

```
REGIONAL STATISTICS FOR COLLECTION OF STATIONS

WEIGHTED STATS - MEAN                          182.        0.864
WEIGHTED STATS - STANDARD DEVIATION              5.        0.031

NUMBER OF REGIONAL DATA =    636.
```

Screen Shot 6-9 – Seasonality Statistics for Date of Maxima

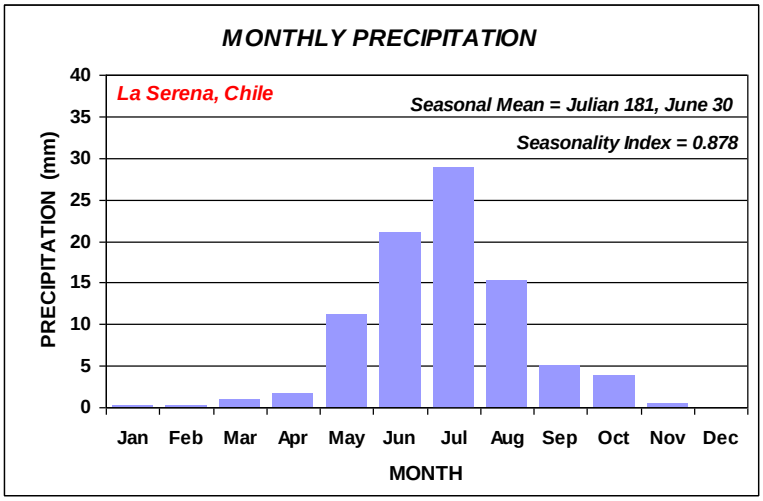


Figure 6-3a – Seasonality Histogram and Circular Statistics for Annual Precipitation for La Serena, Chile

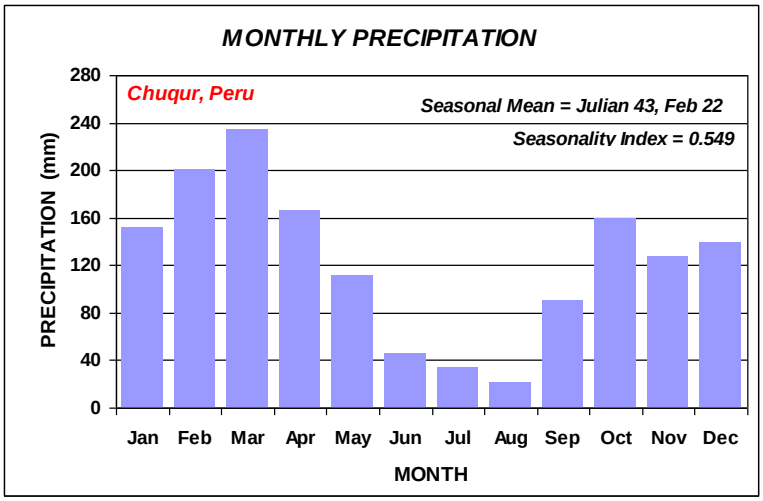


Figure 6-3b – Seasonality Histogram and Circular Statistics for Annual Precipitation for Chuqur, Peru

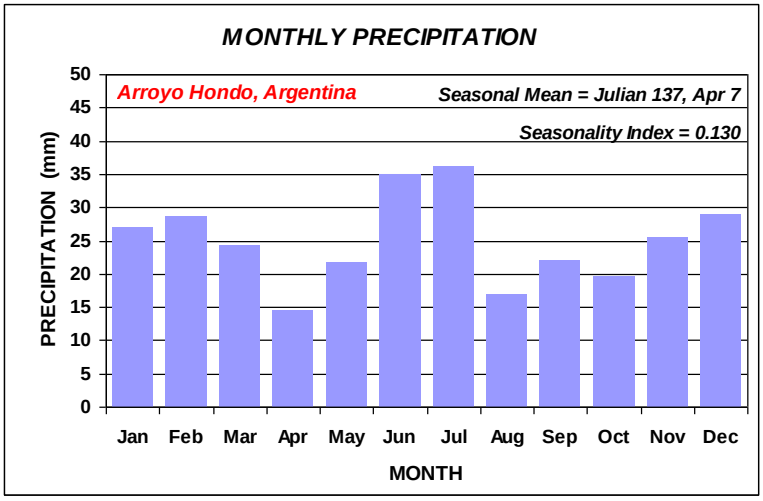
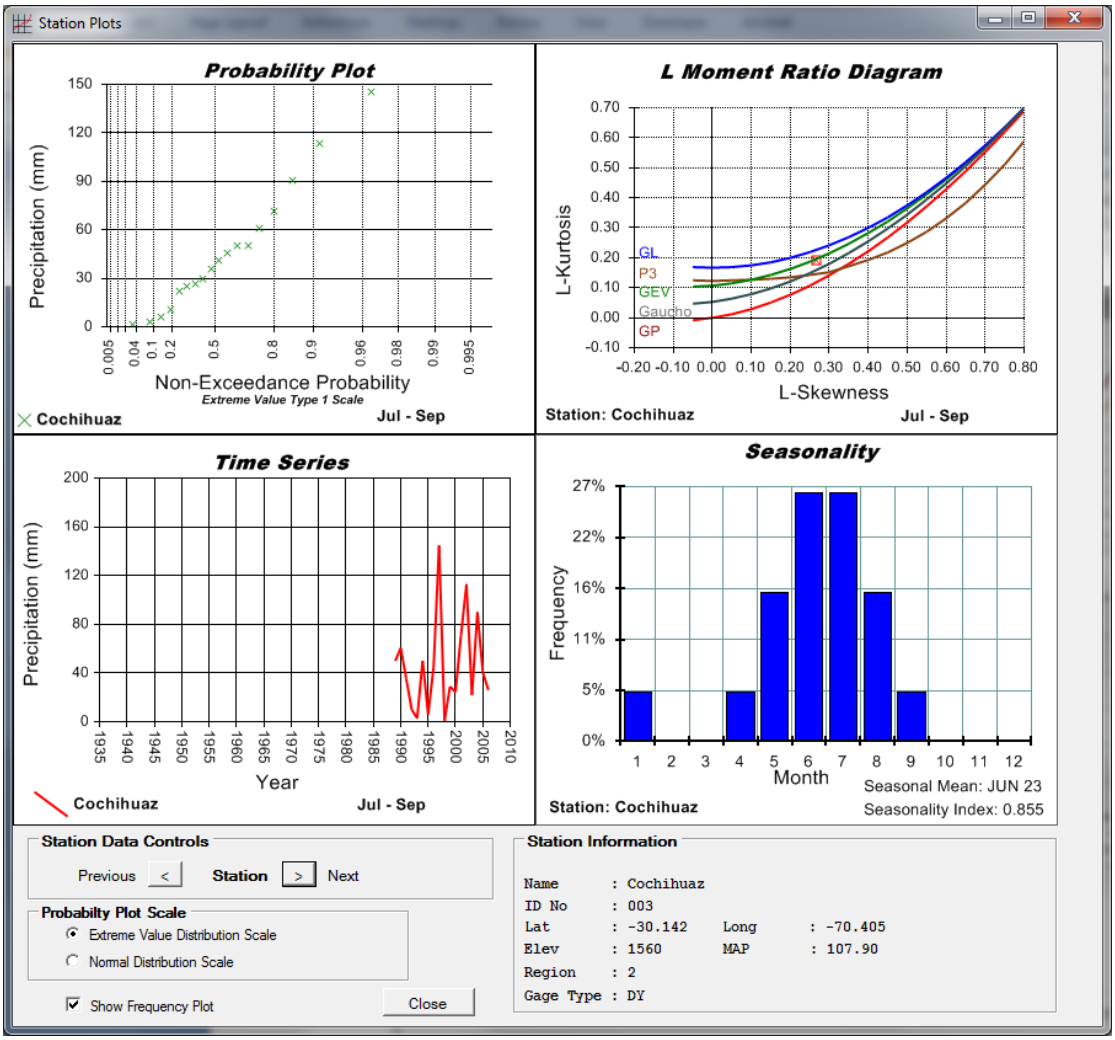


Figure 6-3c – Seasonality Histogram and Circular Statistics for Annual Precipitation for Arroyo Hondo, Argentina

6-8.5 Initial Assessment of Data Series for Data Quality

The station data screening tool provides the ability to quickly examine the behavior of data series. Upon clicking the *View Station Screening Plots* button, the user can toggle through the selected stations using the forward and reverse *Data Control* icons. For each station, a view is provided of the probability-plot, time-series plot, L-Moment Ratio Diagram and a seasonality histogram (Screen Shot 6-10). The user has the option of selecting the probability-plot to be drawn on Normal probability plotting paper or Extreme Value Type 1 plotting paper. The Normal probability plotting paper would be preferred for data series with low L-Skewness. The Extreme Value Type 1 plotting paper would be preferred for data series with moderate to large L-Skewness.



Screen Shot 6-10 – Graphical Analyses Available for Station Data Series

Data Quality Checking – One of the primary uses of the station screening tool is for data quality checking. Errors in the largest and smallest values in the data series would generally have the greatest effect on distorting L-moment sample statistics. Therefore, the probability-plot should be examined for each station and an assessment made of the general behavior (shape) of the plot with particular attention given to the largest and smallest values. Any of the largest or smallest values that markedly depart from the general shape of the plotted points should be subject to further examination. This is accomplished by reviewing the original data and looking for any

corroborating information from other sources. The goal is to confirm the validity of the recorded values. All values that are found to be valid are retained. All values that are found to be erroneous should be removed from the analyses by marking the value as negative and assigning an "R" data quality flag (see Section 6-5.4). Alternatively, if the correct data value can be determined, then edit the data series to provide the correct value and mark the data with a data quality flag of "E". A general guideline is that data values are considered innocent and retained unless proven to be erroneous and then removed.

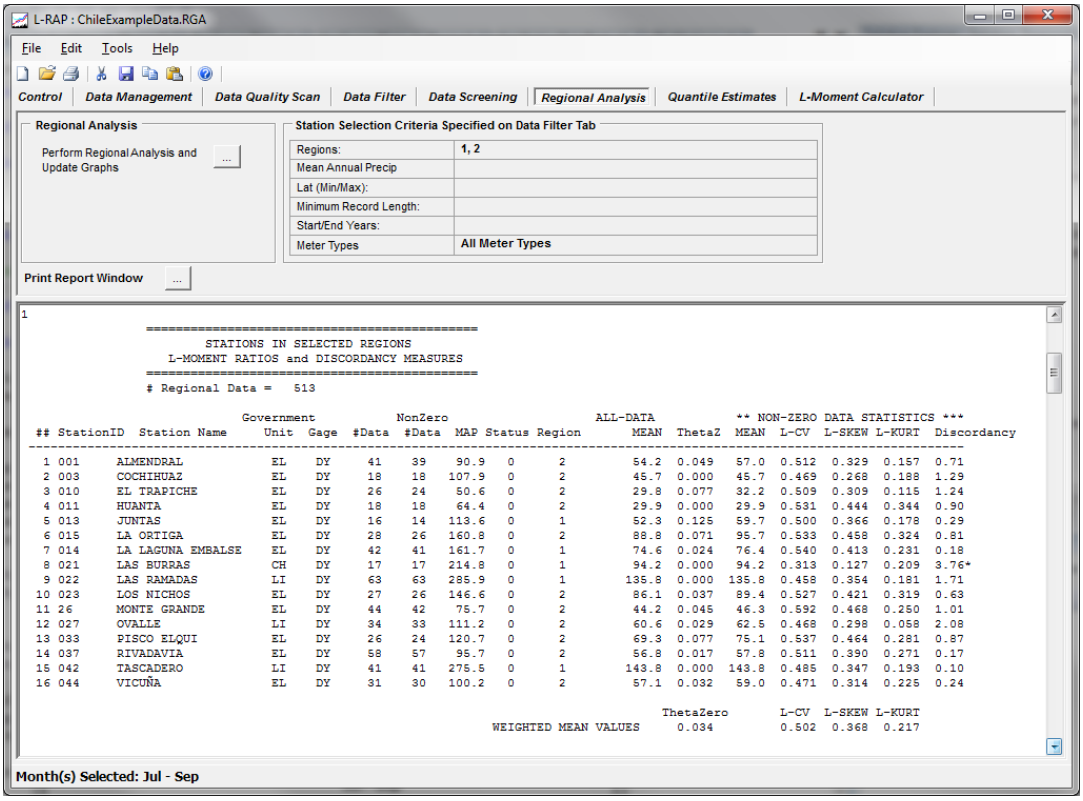
The L-Moment Ratio Diagram can also be used to assist in identifying any L-Skewness and L-Kurtosis pairings that differ markedly from the other stations in the selected group of stations. In addition, any stations that were marked as discordant when the regions were selected during the ***Data Filter*** process should also be examined further to determine if there were any data errors that are the cause of the discordancy. Alternatively, the apparent unusual behavior may simply due to sampling variability and valid high or low outliers.

6-9 REGIONAL ANALYSES

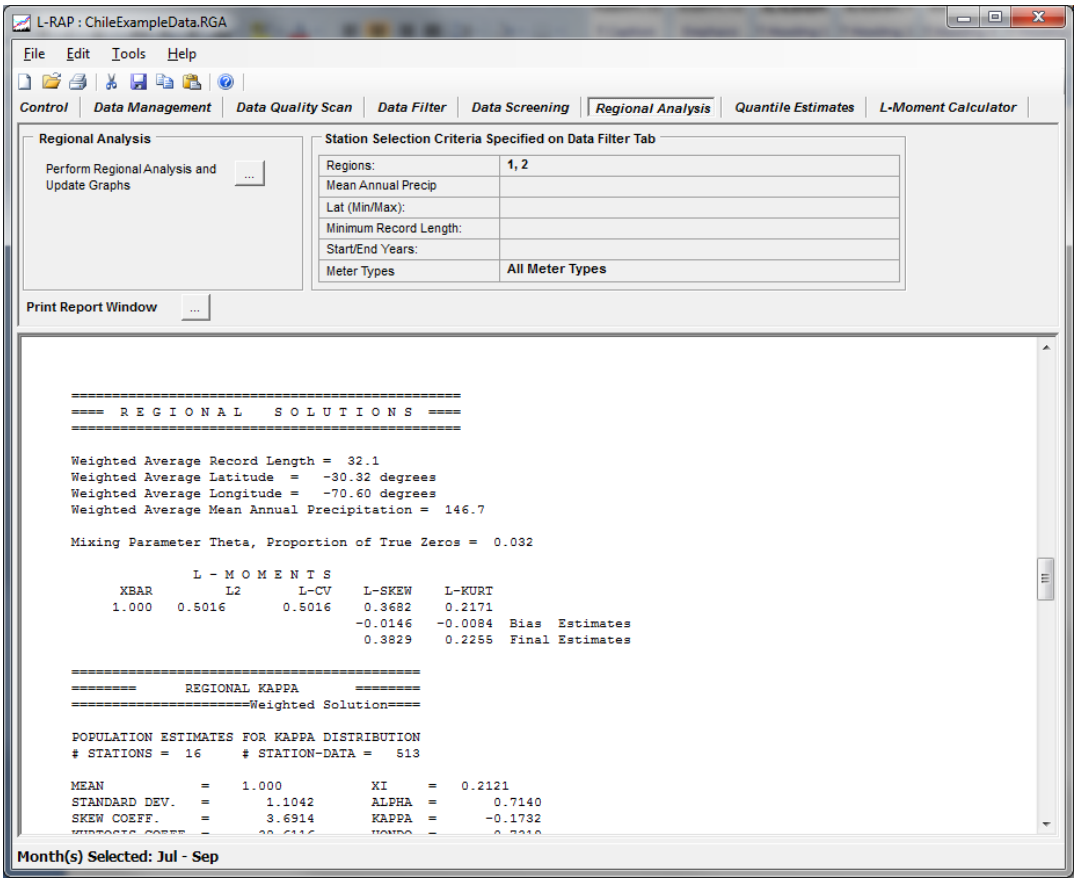
Regional Analyses can be conducted after data series for all the stations have been examined and all data quality issues have been addressed. The Regional Analyses screen shows the data filter selection criteria that are in effect for those stations that were selected in the ***Data Filter*** step. The Regional Analyses is executed by Clicking on the *Perform Regional Analyses and Update Graphs* button (Screen Shot 6-11a). The numerical results are depicted in the view screen (Screen Shots 6-11b,c,d) and graphics are automatically displayed in a separate window (Screen Shot 6-12). For all graphics in L-RAP, right clicking on the graphic allows changes to be made in formatting the graphic, such as changing tiles, label axes, etc.

The numerical results include:

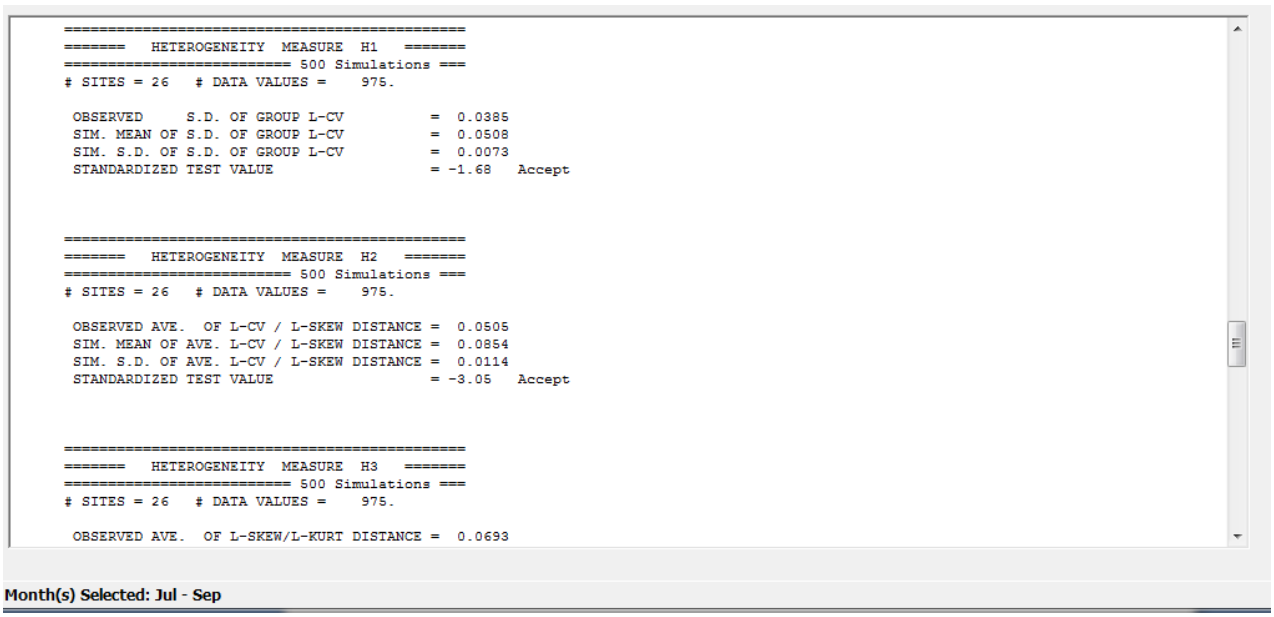
- For all stations, a listing is provided of: mixing parameter (θ); at-site mean; at-site L-Cv, at-site L-Skewness and L-Kurtosis, discordancy measure (Di).
- Regional solution of L-moment ratios weighted by record length.
- Heterogeneity measures H1, H2 and H3.
- Goodness-of-fit measures for all distributions and identification of distributions that are acceptably close to regional L-moment ratio solutions.
- Solution of distribution parameters for accepted probability distributions based on fitting to regional L-moment ratios.
- Quantile estimates for non-zero values for regional growth curve for those probability distributions identified as acceptably close to regional L-moments.



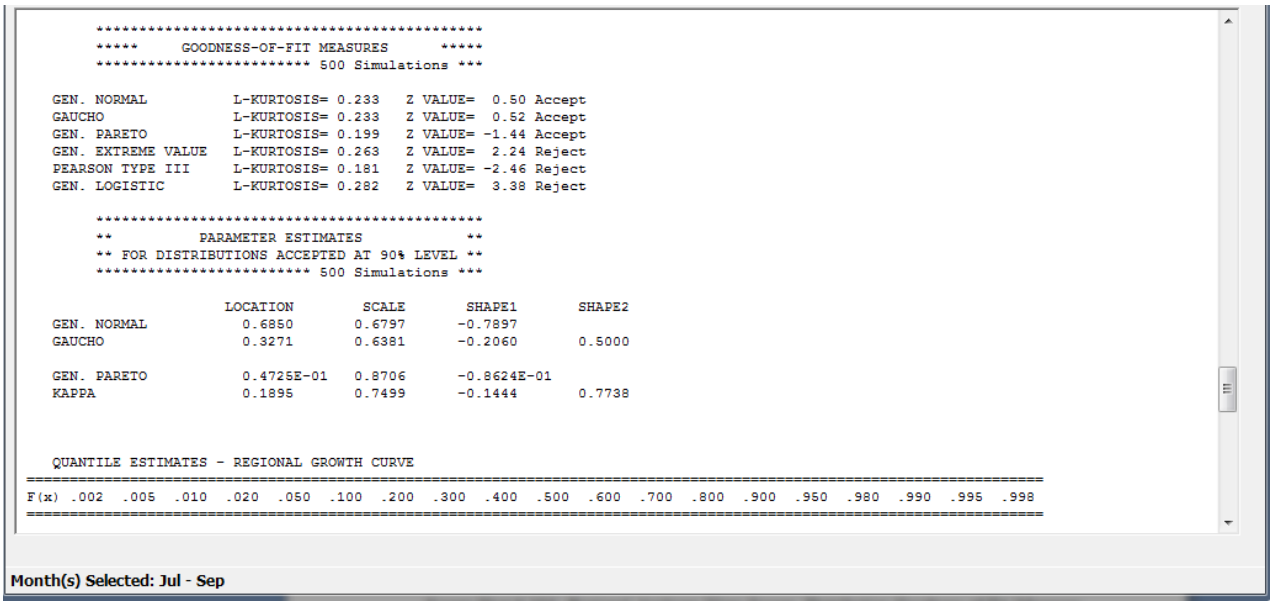
Screen Shot 6-11a – Layout of Regional Analyses Screen



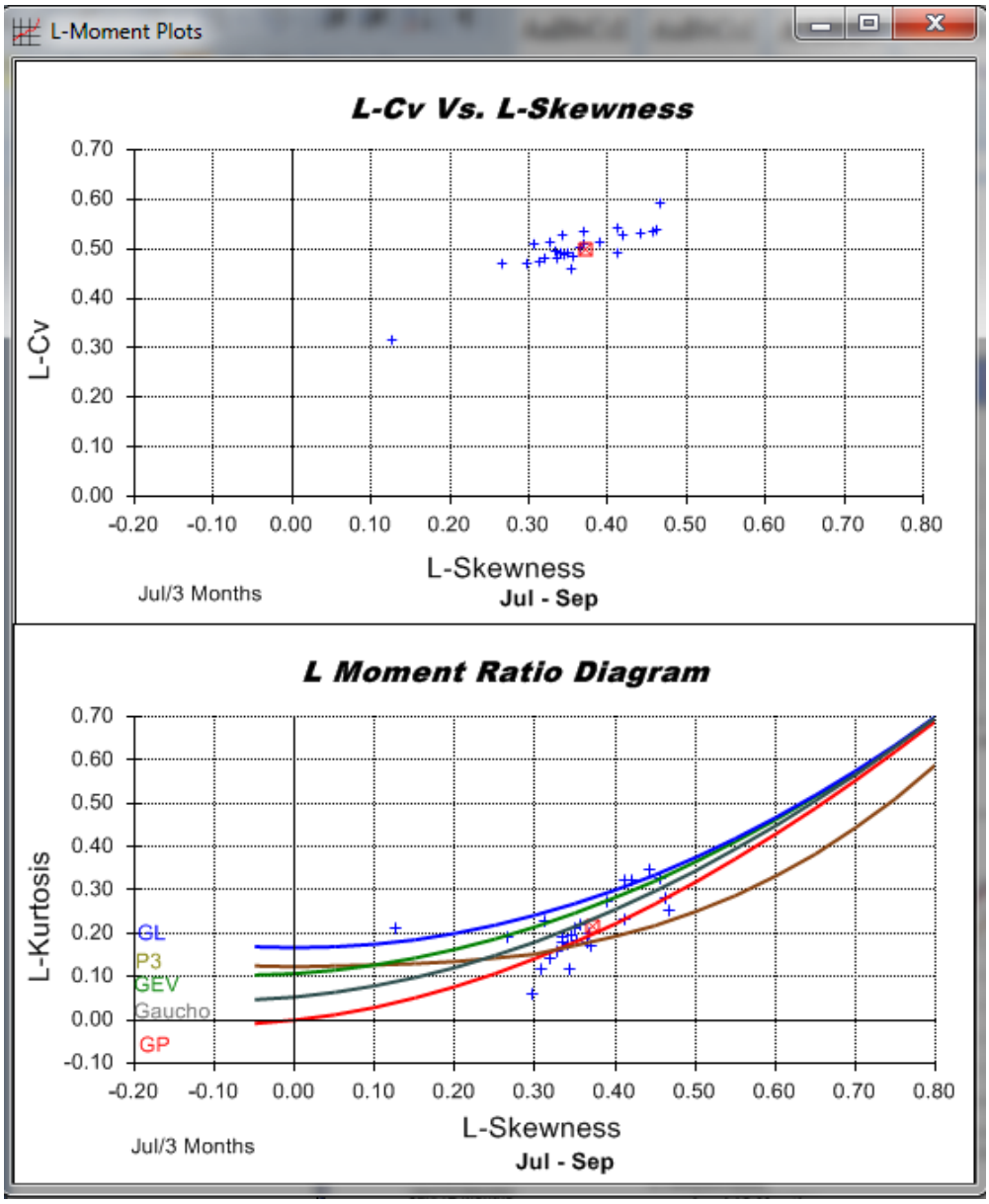
Screen Shot 6-11b –Regional Analyses Screen Showing Regional L-Moments



Screen Shot 6-11c –Regional Analyses View Screen, Heterogeneity Measures



Screen Shot 6-11d –Regional Analyses View Screen, Distribution Goodness-of-Fit Measures

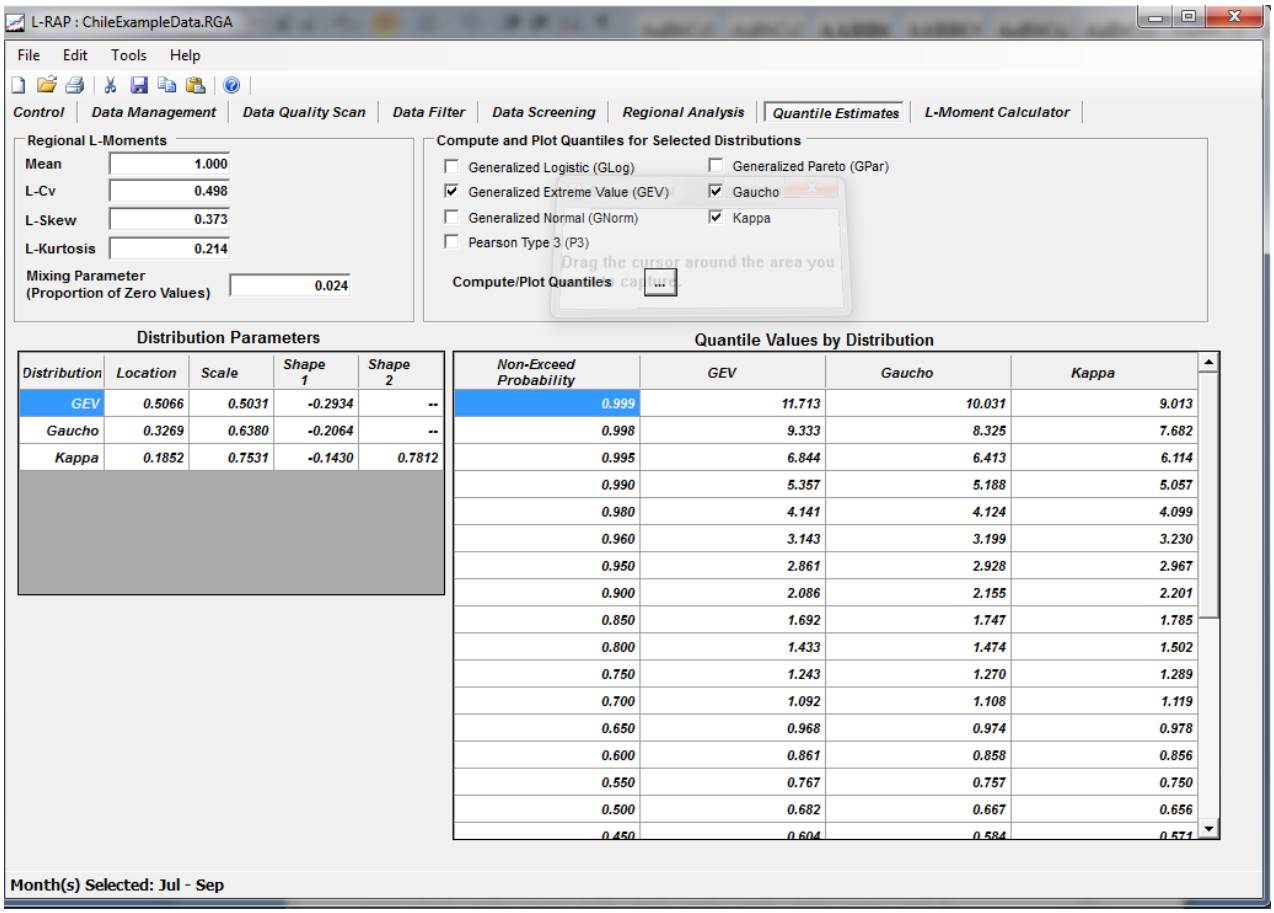


Screen Shot 6-12 – Plots of L-Cv versus L-Skewness and L-Moment Ratio Diagram

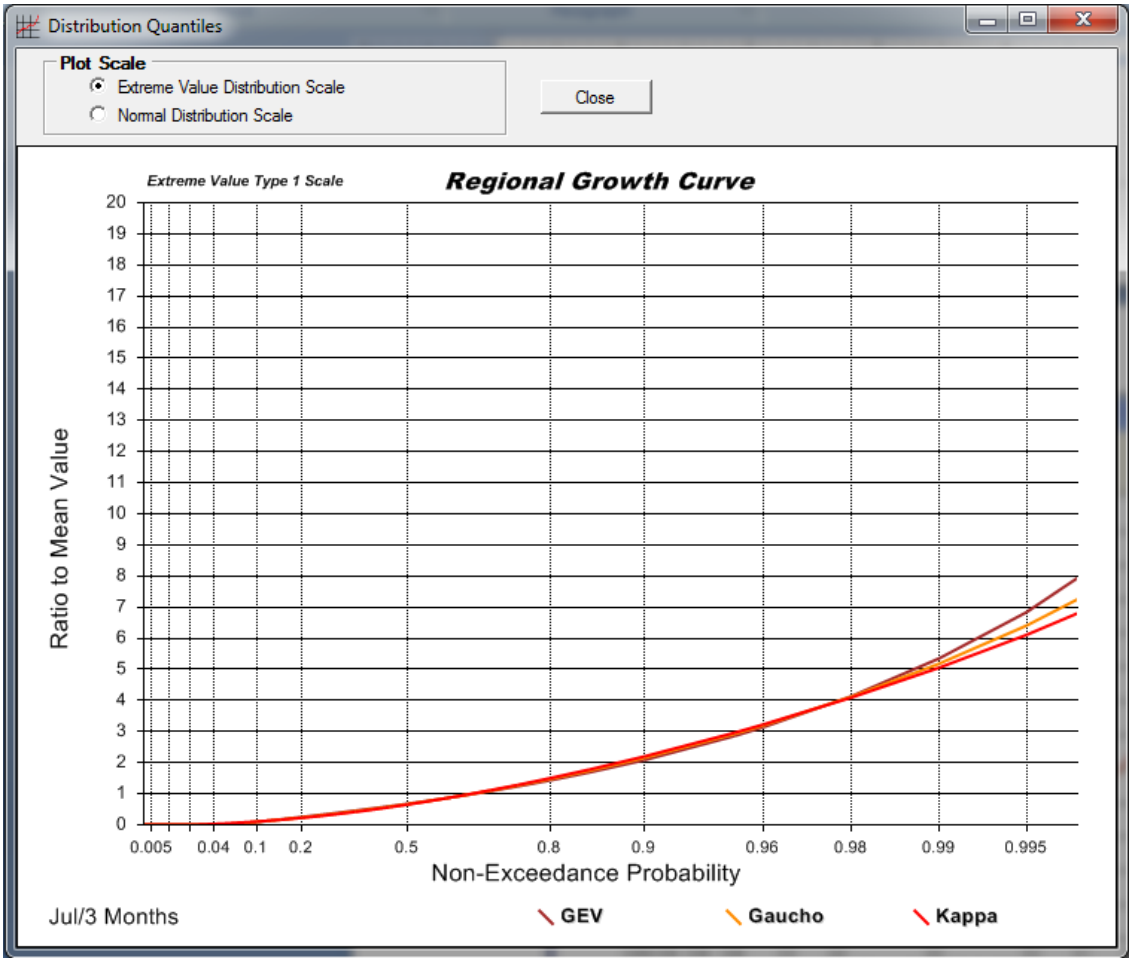
6-10 QUANTILE ESTIMATES

The ***Quantile Estimates*** screen is used for computing distribution parameters and regional growth curves for seven probability distributions (Screen Shot 6-13a). The L-moments input boxes are populated with the results from the ***Regional Analyses*** tab. The ***Quantile Estimates*** screen is operated as follows:

- Select the probability distribution(s) for computing quantile estimates for developing regional growth curves.
- Click on the *Compute Quantile Estimates* button to compute distribution parameters and quantile estimates for the selected probability distributions. A graphic window will appear with plots of the regional growth curves (Screen Shot 6-13b).
- If interested in site-specific quantile estimates, change the mean value from 1.000 to the at-site mean of interest and click on the *Compute Quantile Estimates* button to recomputed quantile estimates.



Screen Shot 6-13a – Layout of Quantile Estimates Screen

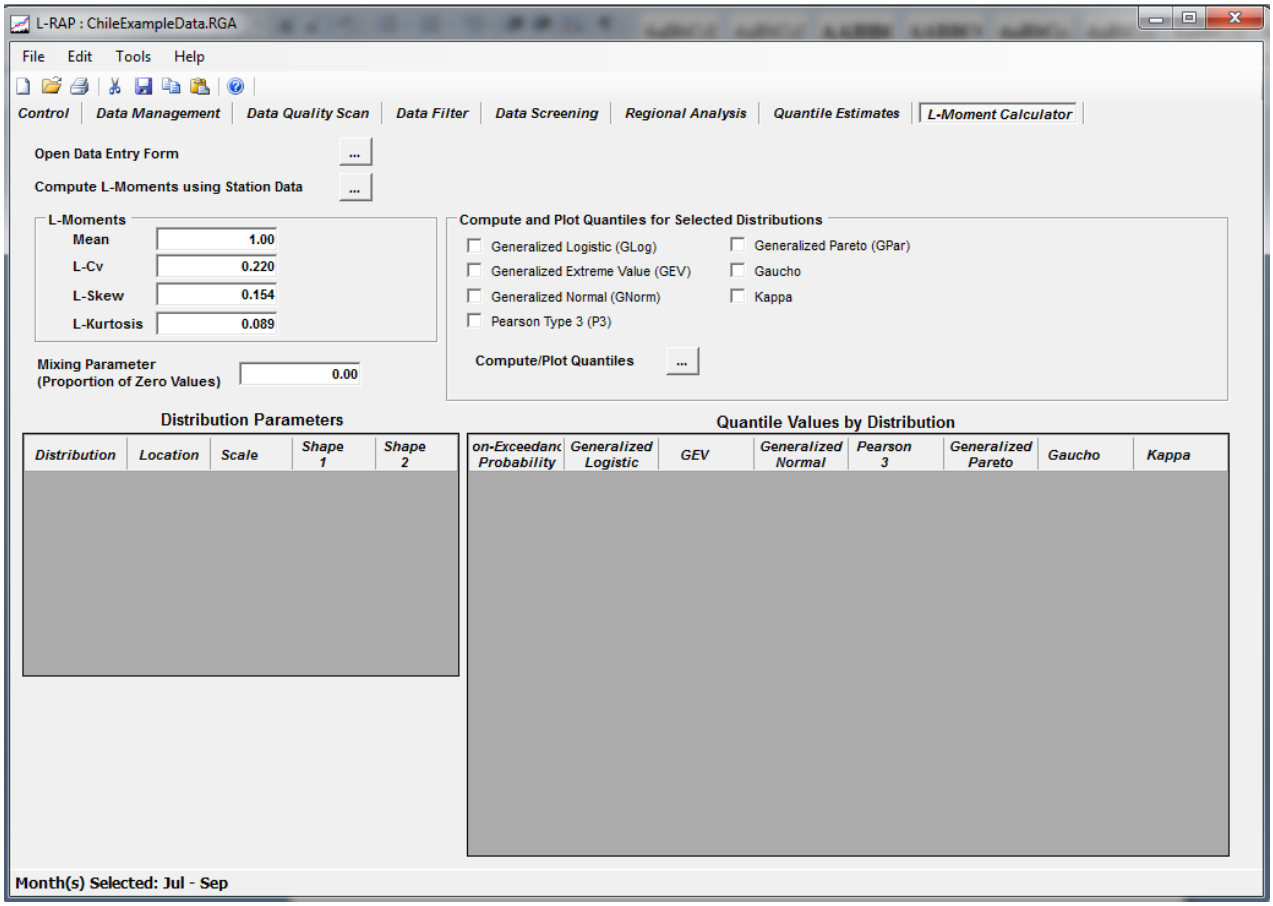


Screen Shot 6-13b – Regional Growth Curves for Selected Probability Distributions

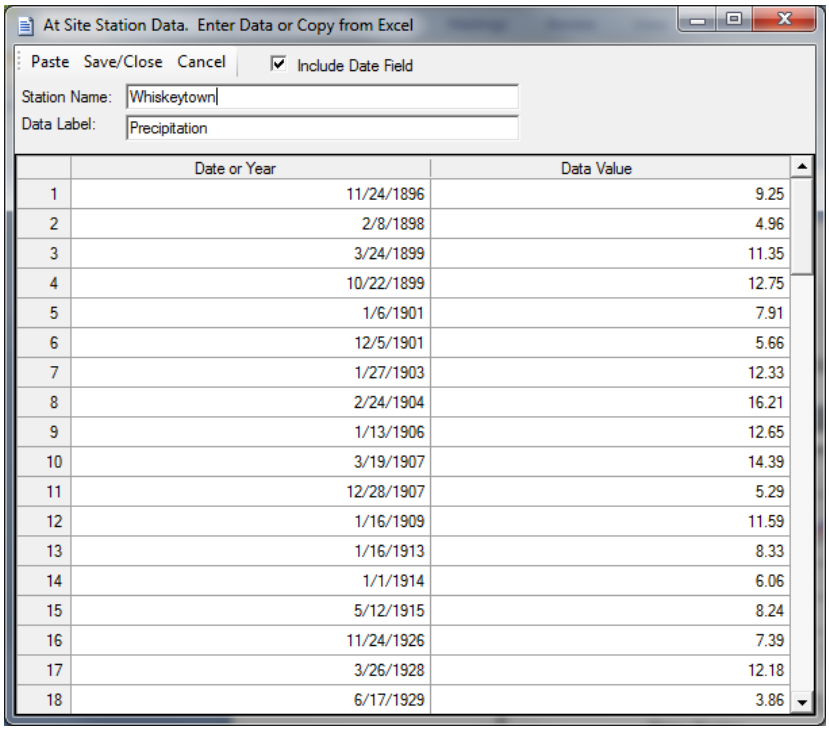
6-11 L-MOMENT CALCULATOR SCREEN

The ***L-Moment Calculator*** screen provides an easy method of computing L-moments for a single data series and viewing the resultant graphics for the data series (Screen Shot 6-14a). The calculator functions independently of the data and analyses performed on other tabs of the program. The supported operations are:

- The data series input screen is opened by clicking the *Open Data Entry Form* button. Data series may be entered by hand or by cut-and-past from another application such as Excel (Screen Shot 6-14b). Text associated with the Station Name and Data Label will be displayed on frequency and time series graphs described below. If the data to be analyzed is not a time series, i.e., does not have a corresponding date for each data point, uncheck the *Include Date Field* box.
- Once the data has been entered, click the *Save/Close* button.
- Clicking on the *Compute L-Moments* button returns the at-site sample L-moments to the text boxes on the L-Moment Tab.
- Use a checkmark to select the probability distribution(s) for computing quantile estimates for the data set.
- Clicking Compute/Plot Quantiles provides a probability-plot, time-series plot, L-moment diagram, and seasonality histogram for the data series in a separate graphics window (Screen Shot 6-14c).

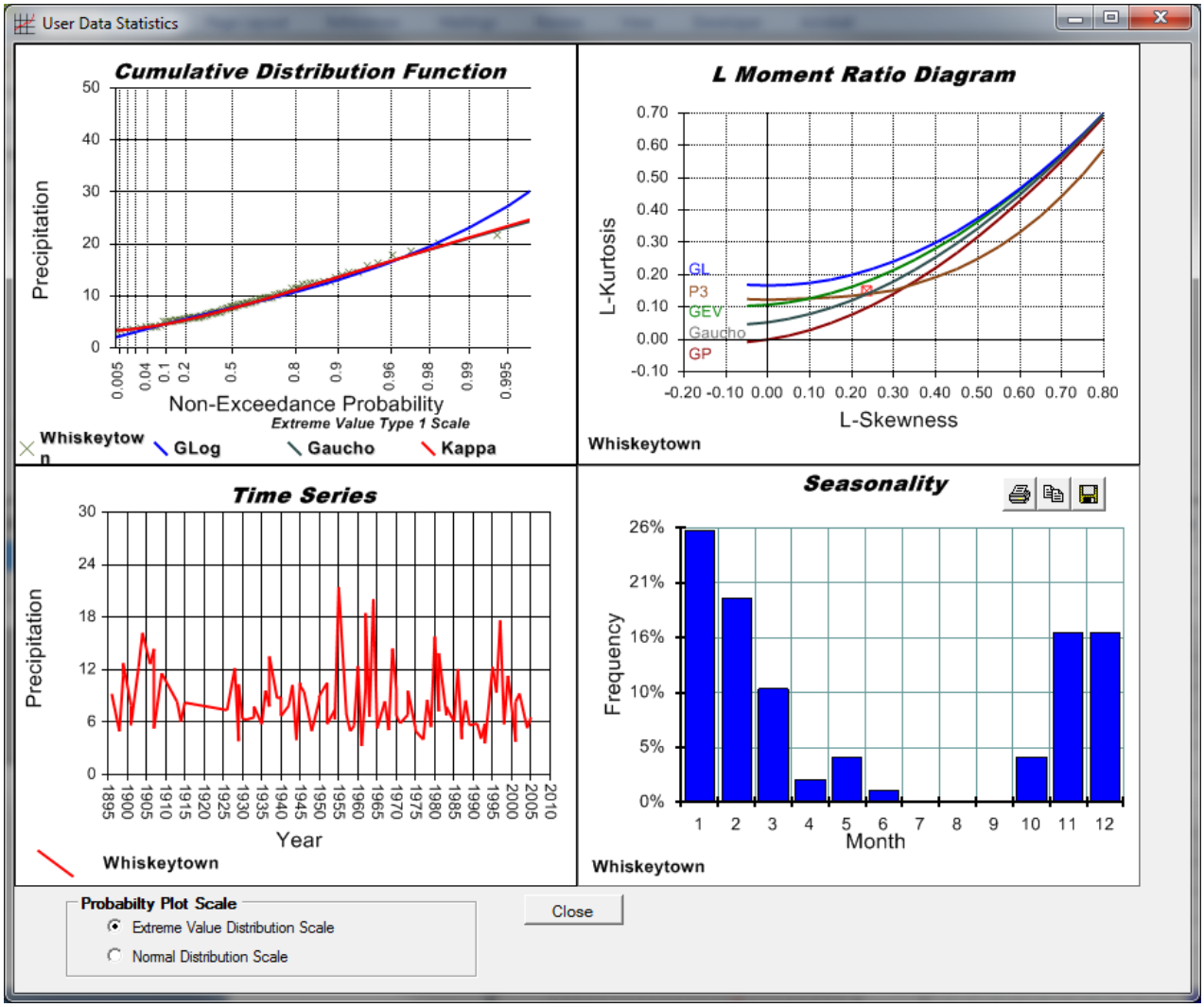


Screen Shot 6-14a – Layout of At-Site Computations Screen



	Date or Year	Data Value
1	11/24/1896	9.25
2	2/8/1898	4.96
3	3/24/1899	11.35
4	10/22/1899	12.75
5	1/6/1901	7.91
6	12/5/1901	5.66
7	1/27/1903	12.33
8	2/24/1904	16.21
9	1/13/1906	12.65
10	3/19/1907	14.39
11	12/28/1907	5.29
12	1/16/1909	11.59
13	1/16/1913	8.33
14	1/1/1914	6.06
15	5/12/1915	8.24
16	11/24/1926	7.39
17	3/26/1928	12.18
18	6/17/1929	3.86

Screen Shot 6-14b – Station Data Input Screen for L-Moment Calculator



Screen Shot 6-14c – Graphics Window for At-Site Computation of L-Moments for Site-Specific Data Series