Introduction

When it comes to setting policy, making water use plans, designing infrastructure, and making operational water management decisions, it is important to be informed concerning the frequency, severity, and duration of droughts (and floods) in the region. Questions posed by water planners, managers, and decision-makers that should be answered include the following: “How rare is the current drought?”, “How likely is it that the drought will end in X months?”, “How large a drought should we plan for?”, and “How rare is the drought of record?”. These questions can be answered by performing a frequency analysis of rainfall and using the results to compute statistics for the various sites in question.

Most tools that have been developed for this purpose require a license to use or are made specifically for a particular group’s interests. The International Center for Integrated Water Resources Management (ICIWaRM) is currently developing a non-proprietary and comprehensive Regional Frequency Analysis (RFA) tool (which has recently been renamed the ICIWaRM Regional Analysis of Frequency Tool or ICI-RAFT) that can be used in general to answer the questions above; the title screen is shown in Fig. 1. ICI-RAFT uses the method of L-moments, which provides the means for estimating the shape of the rainfall frequency distribution for an entire homogeneous region. L-moments are analogous to regular moments (mean, standard deviation, skewness and kurtosis) except that they are calculated using equations that represent linear combinations of the sample or population data elements, hence the name Linear-moments or L-moments. The resulting statistics are referred to as L-location (or mean), L-scale, L-CV (Coefficient of Variability), L-skewness, and L-kurtosis. The advantage of L-moments over regular moments is that using linear statistics makes the procedure more robust with respect to outliers. L-moments are also able to characterize a wider range of frequency distributions and are less subject to bias in their calculations than regular moments.

Another advantage of ICI-RAFT is that, as the name implies, it estimates site L-moments by first estimating the L-moments and the associated frequency analysis curves for multiple homogeneous regions that each contain multiple sites. Advantages of
performing a regional frequency analysis instead of using each individual site’s data separately is that it allows the inclusion of sites with short record lengths, few rain gage stations, a substantial amount of missing data, or station data with a higher variability due to the presence of mountain ranges and/or multiple precipitation sources. In addition, it is difficult to find frequency distributions that fit an individual site’s rainfall data well. A regional analysis that combines the data from all sites can obtain a closer fit. These advantages especially make this tool ideal for use in arid and semi-arid regions where other methods fall short.

Input Files

The first step when using ICI-RAFT is to enter quality-controlled data into the input file, which ICI-RAFT requires to be an Excel worksheet. The input file must contain two sheets: moxyrData (Fig. 2a) contains the monthly total precipitation values to be analyzed, while OtherData (Fig. 2b) holds the site characteristics, e.g. mean annual precipitation, elevation, longitude and latitude; this information will assist the user when creating homogenous regions. In moxyrData (Fig. 2a) Column A contains each Site ID, whereas Column B contains each year for each site. As the user moves along a single row for a particular site and year, monthly values of total precipitation are entered. If a particular monthly total is not known or is known to be incorrect, then “-999” is entered.

In OtherData (Fig. 2b), each column represents a particular site characteristic that can be used when creating homogenous regions if such data are available. Only variables that are known for all sites and years will be used during regionalization; therefore not all columns are required to be filled in. The user should remember that the more data that are available, the more efficient the regionalization process will be. In addition to the site characteristics shown in Fig. 2b, the user is welcomed to add up to three of their own site characteristics in the Other columns. Within the Other columns, the name of the site characteristic is required in Row 2, and the units are required in Row 3. An alternative to

Distribution Statement A -- Approved for public release; distribution is unlimited.
Figure 2: Example input file, showing the (a) rainfall data and (b) site characteristics worksheets.
regionalization is to use Eco-regions, as defined by the World Wildlife Federation, the U.S. Environmental Protection Agency, or other groups, or to choose to manually set your own region for each site; this can be done using Column P. The final column will be discussed later in this document.

In terms of climate variability, a database of indices from various types of multi-decadal meteorological events such as the El Nino/Southern Oscillation, the North Atlantic Oscillation, the Pacific Decadal Oscillation, and others has been included. Using this database, only rainfall data from a particular phase of any one of these phenomena can be analyzed and compared to the opposite phase. This should give an indication of the effects of that phenomenon on climate variability in terms of the frequency of certain rainfall intensities. This feature will be ready in a future version of ICI-RAFT.

Data Screening

Once the input file is completed and loaded into ICI-RAFT, the next step is to choose the period for analysis. Upon selection of Screen Data within ICI-RAFT, a window appears (Fig. 3) that allows the user to determine the time period over which analysis is to be performed. A Beginning Month and Duration in Months must be entered to define this time frame; individual monthly values within the period chosen are totaled in order to obtain one value for each period; the length of the period analyzed can be from 1 month to 60 months or 5 years. If there are any missing or bad data during a particular year (represented by a -999 in the input file), that year is not used in the analysis for that
particular data site. Fewer years of data available for a particular site means that that particular site will contribute less to the overall regionalization process. A threshold for the number of valid years for a particular site can be set in the Screen Data window next to the caption Minimum number of records. Any sites having fewer than the specified number of annual records with good data will not be used in the regionalization process. This threshold must be more than four so that all of the program’s statistics are calculable for all sites. Finally, the user enters the Data Units of the monthly precipitation data.

Site Analysis

Upon completion of the Initial Screening form, the Site Details form appears (Fig. 4); this window can also be accessed through the Sites menu on the title screen. The Site Details form contains two lists of sites, one each for accepted and rejected sites based on the number of years of data they contained with no missing or bad data (-999’s) compared to the minimum numbers of years specified by the user in the Initial Screening window. Sites listed in the Sites Accepted column may be selected, which causes their L-moment statistics to be displayed in the left column of the form along with the Site ID and the Discordance of that site. In this case, the discordance is a measure of how a particular site’s L-moment ratios compare to the site L-moment ratios of the entire data set. High discordance could indicate that a site’s data may contain some type of measurement error, which may include, for example, an error related to data entry.

The right side of the Site Details window in Fig. 4 contains a graph of Latitude vs. Longitude. Discordant sites are shown on the graph as red dots, whereas the blue dots represent non-discordant sites. Discordant sites are also identified with an asterisk after their site-id in the Sites Accepted column. The site that has been selected in the Sites Accepted column to the left is identified in the graph by a large dot, which is either light blue or red depending whether it is discordant or non-discordant. Drop-down menus allow the user to display \( L_1, t_1, t_3, t_4, t_5 \), or any of the at-site information from the OtherData input sheet as the x- or y-variables on the graph. Graphing the various L-moments and site characteristics against each other may help to identify why some sites are discordant compared to others. For example, as can be seen in Fig. 4, many of the discordant sites seem to be located on the West Coast of South America on the windward side of the Andes Mountains, which reveals a different precipitation regime than the remainder of South America.

Any site in the Accepted list may be eliminated from the analysis by clicking the Delete Site button. A confirmation box (not shown) will appear; clicking OK will remove the site from your analysis. The site can only be added back to the analysis by repeating the initial screening procedure from the previous step.

Plots

Choosing Plots in the Sites menu within the Title Screen opens a window allowing the user to view graphs of the data for any individual site. Any site can be chosen using the drop-down box at the top of the window. Radio buttons below the chart window allow the user to view Raw Data, Rescaled Data, or a \( t_4/t_3 \) (L-Kurtosis vs. L-Skewness) plot.
Selecting **Raw Data** will allow the user to see a graph of the raw data for the selected site (Fig. 5a), whereas selecting the **Rescaled Data** option produces a graph of the raw data divided by the mean of that dataset (not shown). The value of rescaling the data will be seen in the regionalization portion of the analysis.

The graph of L-Kurtosis vs. L-Skewness (Fig. 5b) illustrates how each individual site’s L-moments compare to the other sites (represented by the gray circles) and to the possible L-skew and L-kurtosis values representative of various frequency distributions. The black dot represents the L-Kurtosis and L-Skew for the site selected in the drop-down box. The colored stars represent the L-Kurtosis and L-Skew for the two-parameter frequency distributions that are available within this program, while the curves represent the range of possible values of L-Kurtosis vs. L-Skew for each of the three-parameter distributions, as indicated in the legend on the graph. This graph provides an initial indication of which frequency distribution may best represent the data for a particular site based on the proximity of the site’s value of $t_4/t_3$ compared to each distribution. It should be noted that this provides only one clue as to which distribution to use and should never be used alone when making a final decision.

**Frequency Analysis**

The final option in the **Sites** menu on the **Title Screen** is **Frequency Distribution**. Once clicked, a window opens that displays the first site’s data probability density function (PDF) as vertical blue bars (Fig. 6a). Any of the accepted sites for analysis can be viewed using the drop-down box below the graph. In addition, fits of any of the thirteen different standard frequency distribution functions (see Table 1) to the L-moments of the site’s data can be displayed on the graph by clicking the box next to the

**Distribution Statement A -- Approved for public release; distribution is unlimited.**
Figure 5: Site Plots window, showing (a) the raw data for a particular site and (b) the $t_4/t_3$ graph, which includes all sites as gray dots, the selected site as a black dot, two-parameter distributions as colored stars, and the three-parameter distributions as colored curves. All distributions are identified in the legend.
Table 1: Standard frequency distribution abbreviations.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>EXP</td>
</tr>
<tr>
<td>Gumbel</td>
<td>GUM</td>
</tr>
<tr>
<td>Uniform</td>
<td>UNI</td>
</tr>
<tr>
<td>Normal</td>
<td>NOR</td>
</tr>
<tr>
<td>Generalized Extreme Value</td>
<td>GEV</td>
</tr>
<tr>
<td>Generalized Logistic</td>
<td>GLO</td>
</tr>
<tr>
<td>Generalized Pareto</td>
<td>GPA</td>
</tr>
<tr>
<td>Log-normal</td>
<td>LN3</td>
</tr>
<tr>
<td>Pearson Type III</td>
<td>PE3</td>
</tr>
<tr>
<td>Log-Pearson Type III</td>
<td>LP3</td>
</tr>
<tr>
<td>Gamma</td>
<td>GAM</td>
</tr>
<tr>
<td>Kappa (4-parameter)</td>
<td>KAP</td>
</tr>
<tr>
<td>Wakeby</td>
<td>WAK</td>
</tr>
</tbody>
</table>

desired distribution. Distributions are divided based on the number of parameters required. This graph provides a way to visually determine how well each standard frequency distribution fits the site’s sample data and offers another clue in determining which distribution may be best for a particular site. In addition, a site’s cumulative distribution function or CDF (not shown) and Exceedance curve (Fig. 6b) can also be displayed. The exceedance graph is shown on a logarithmic scale and provides the user with information on the intensity of a storm for a particular frequency. The estimated exceedance curve of the sample data is represented by the blue dots, and the distributions are represented as colored lines. Each of these graphs can be used to make a visual comparison between how well each standard distribution fits the site data. As in the previous screen, these graphs should not be used alone in finally determining which distribution to use to characterize a particular site’s distribution curve.

At the bottom of the At-site Distribution Plots window there are buttons that allow one to view the parameters for a particular standard frequency distribution estimated from the L-moments of the site selected. In addition, there is also a button titled Analysis, and once clicked, the screen shown in Fig. 7 appears. All calculations performed within this window relate to the site selected in the At-Site Distribution Plots window shown in Fig. 6. The Analysis window is used to make a quantitative estimate of the intensity of an event for a particular frequency. This can be accomplished using any of the thirteen standard frequency distributions mentioned earlier. Once the desired distribution has been selected and the desired frequency entered, clicking the Update button will give the associated intensity. Also shown in this window are the R-squared values and z-scores for each distribution, which give an indication of the closeness of fit of each standard distribution to the sample data of the site being analyzed. The R-squared value is determined by computing the square root of the means of the difference squared between the sample data and the standard distribution. The z-score is computed based on the distance between a site’s $t_4/t_3$ value on the graph shown in Fig. 5b and that of each 3-
Figure 6: **At-Site Distribution Plots** screen, showing (a) the probability density function (PDF) and (b) the exceedance curve for the sample data of the selected site as blue bars (PDF) and blue dots (exceedance curve). Also included are the estimations for each curve using three different standard frequency distributions: **EXP** (blue), **GEV** (black), and **GLO** (violet).
Figure 7: This screen estimates the storm intensity for an event of a particular frequency using one of 13 standard frequency distributions. Closeness-of-fit statistics (R-Squared and z-score) are shown for each distribution compared to the sample data of the site selected in the At-Site Distribution Plots window.

An R-Squared value close to 1 and a z-score close to 0 indicate a good fit. These statistics, along with the graph in Fig. 5b and the visual fits of the PDF, CDF, and exceedance curves should be used together to determine the preferable standard frequency distribution to use for a particular site.

**Regional Analysis**

Once the user is finished with site analysis, they can move on to creating homogenous regions and regional analysis. Creating regions that are as homogenous as possible can be a difficult and time-consuming process; therefore, ICI-RAFT attempts to facilitate this process as much as possible through its visual features. In order to create regions, the Regionalization option should be selected under the Regions menu on the Title Screen; the window that appears is shown in Fig. 8. Regions may be created using the site characteristics in the input file, by using standard eco-regions, or by entering the region numbers manually. Prior to choosing the Eco-regions/Own regions option, the user must enter the eco-region or specify their own region number for each site in the final column of the OtherData worksheet of the input file. If the final column of the OtherData worksheet is not complete for all sites, the user will be unable to use this option and will be required to use site characteristics to create regions.

Distribution Statement A -- Approved for public release; distribution is unlimited.
When using site characteristics to create regions, only those site characteristics that are completely filled in the OtherData worksheet of the input file are enabled in the Regionalization window. The user will enter the number of regions desired and the weights for each site characteristic as numbers between 0 and 1, making sure that all weights sum to 1. If the user entered data in the Other columns of the OtherData worksheet of the input file, these will also be enabled in the Regionalization window along with the name and units of the added site characteristic as specified in the input file; the name and units can be edited in the Regionalization window if desired. It is also possible to set all weights back to zero by clicking the Reset Weights button. The user can also make all weights for enabled site characteristics equal by clicking the box next to Equal Weights.

Details

After regionalization is completed, the Regional Details window (Fig. 9) will appear. The regional L-moments and homogeneity for a particular region are shown, while the L-moment statistics and discordance of a selected site within that region are also shown. In this case, the site discordance is a measure of how a particular site’s L-moment ratios compare to the site L-moment ratios of the other sites within the selected region. The homogeneity statistic gives an idea of how close the frequency distributions of all sites
within the region are to each other. Also displayed in this window is a graph similar to that shown in Fig. 4 during the site analysis. All sites not included in the selected region are shown as gray dots, whereas the blue and red dots are the same as before. The drop-down menus also allow the user to graph the same variables against each other as in the site analysis. Graphing the various L-moments and site characteristics against each other may help to identify why some sites are discordant compared to others within the selected region, which may lead to moving a particular site to a different region. When the Move to Region box is clicked, a list of the other region numbers is shown, from which a new region can be selected. The selected region appears in light green as illustrated in Fig. 9. This is a prime example of how ICI-RAFT visually facilitates the difficult process of regionalization.

**Plots**

Choosing the option to view Regional Plots is very similar to the corresponding option in the site analysis except that you can view data only from one particular region at a time. One interesting point that should be made here involves the resulting regional $t_4/t_3$ (L-kurtosis vs. L-skewness) plot (Fig. 10). Similar to what was stated before, this graph provides an initial indication of which frequency distribution may best represent the data for a particular region based on the proximity of the sites’ values of $t_4/t_3$ compared to each distribution. Compared to the site analysis, it is much easier to identify the sites of a region with a particular standard frequency distribution. As shown in Fig. 10, some of the sites within the region tend to follow the $t_4/t_3$ curve of the Generalized...
Figure 10: **Regional Plots** window displaying the $t_4/t_3$ graph, which includes all sites within the selected region as gray dots, two-parameter distributions as colored stars, and the three-parameter distributions as colored curves. All standard distributions are shown in the legend.

**Pareto** distribution, while the other sites seem to follow one of the other distributions. This observation may lead the user to separate these two sets of sites into different regions.

**Frequency Analysis**

The final product or objective of ICI-RAFT is to be able to estimate the intensity of a storm of a particular duration and frequency for a particular site using regional frequency analysis. This is finally accomplished using the final option under the **Regions** menu on the **Title Screen**. The window that opens displays all of the first region’s site data as a probability density function (PDF), which is represented by the vertical blue bars. The resulting PDF for Region #3 is shown in Fig. 11a. As during site analysis, fits of any of the thirteen different standard frequency distribution functions (see Table 1) to the L-moments of the region’s data can be displayed on the graph. This graph provides a way to visually determine how well each standard frequency distribution fits the region’s sample data and offers another clue in determining which distribution may be best for a particular region. In addition, a region’s cumulative distribution function or CDF (not shown) and exceedance curve (Fig. 11b) can also be displayed. Each of these graphs can
be used to make a visual comparison between how well each standard distribution fits the data of a region.

There are buttons at the bottom of the Regional Distribution Plots window that allow one to view the parameters for a particular standard frequency distribution estimated from the L-moments of the region selected. In addition, there is also a button titled Analysis, and once clicked, the screen shown in Fig. 12 appears. The window in Fig. 12 is divided into two tabs titled Onsite Analysis and Offsite Analysis. Onsite analysis (Fig. 12a) refers to estimating the magnitude of a storm event of a particular frequency for a specific site where data is available. This is similar to what was done in the Site Frequency Analysis window described earlier. A desired region and a site within that region can be selected from the two drop-down boxes at the top of the window. As was also done in the Site Frequency Analysis window, any of the thirteen standard frequency distributions mentioned earlier can be used to estimate the event intensity for a particular site. Additionally, the user needs to choose which site averaging method to use in order to convert the result from a regionally scaled or normalized regional quantile to a site-specific quantile. Two options are available: using the L-Mean (L1) computed from the selected site’s rainfall data provided in the input file or using an Index Flood value for that site. The Index Flood method uses the Mean Precipitation for the period entered in the Initial Screening window; the mean precipitation for this period needs to be entered by the user in the final column of the OtherData worksheet at the data entry phase, if available. An external source, such as a Mean Annual Precipitation map or database will be needed in order to use this option. If this information is not available, then the user must use the L1 option. Once the site averaging method has been chosen, the associated intensity can be found in the same units as the sample data. This method of computing the at-site estimate of the x-year storm event should be an improvement over direct at-site estimates as was performed earlier.

Also shown in this window are the R-squared values and z-scores for each distribution, and as stated before, these statistics, along with the graph in Fig. 10 and the visual fits of the PDF, CDF, and exceedance curves should be used together to determine the preferable standard frequency distribution to use for a particular region.

When the user clicks the Offsite Analysis tab of the Regional Frequency Analysis window, the screen in Fig. 12b appears. Offsite analysis is performed when attempting to estimate the intensity of storms of a particular frequency at sites where there are no data. This is accomplished using the sites that do contain data and some sort of interpolation scheme to estimate the values between these sites. The first step in this process is to identify the region over which the user would like to estimate storm intensities. This is accomplished by entering the range of latitudes (northern and southern limits) and longitudes (eastern and western limits) that the region is to cover. In addition the user needs to enter a resolution in units of degrees; the resolution entered must be less than the lesser of the latitude and longitude ranges. As before, the desired standard frequency distribution, site averaging method and rainfall frequency are required prior to completing the analysis. After all of the required information is provided, the user can
Figure 11: Regional Distribution Plots screen, showing (a) the probability density function (PDF) and (b) the exceedance curve for the sample data of all sites within the selected region as blue bars (PDF) and blue dots (exceedance curve). Also included are the estimations for each curve using three different standard frequency distributions: EXP, GEV, and GLO (see Table 1 for explanations of the distribution function abbreviations).
(a)
Figure 12: Shown are the input screens for regional (a) Onsite Analysis and (b) Offsite Analysis as described in the text.
export the results to a text file that can be imported into a GIS program for further analysis. Figure 13 shows an example of the text file along with information that is included.

**Deliverables & Training**

Using the process described above, the end products of ICI-RAFT include estimations of the intensity of rainfall events for individual sites and rainfall probably maps for a particular region for any rain event frequency, duration, and starting month. When looking at multiple maps of extreme low-rainfall events, the collection of maps is called a Drought Atlas.

The fact that this software will be free facilitates capacity building through training experts from other nations in its use. Initially this training will be focused in South and Central America as a step towards creating one drought atlas for the entire Latin American continental region. Other future applications include a regional analysis of precipitation frequency in the country of Rwanda in Africa and as a mechanism to compare the effects of climate variability and climate change. ICI-RAFT can facilitate a simple climate change induced analysis by looking at rainfall data over a region for a particular range of years using different starting years. Trends over the entire period can then be identified. Being able to differentiate between the effects of climate change and climate variability on regional rainfall using a simple, non-proprietary program is what will make ICI-RAFT a novel and useful tool for multiple user-types, including universities, water managers, and government agencies.

![Figure 13: Resulting output file when performing offsite analysis.](image)

*Distribution Statement A -- Approved for public release; distribution is unlimited.*